p-ISSN 2521-4209
e-ISSN 2521-4217

جامعة التنمية البشرية
UNIVERSITY OF HUMAN DEVELOPMENT

UHD Journal of Science and Technology

A Scientific periodical issued by University of Human Developement

Vol.9 No.(2) December 2025

2025 2725

www.jst.uhd.edu.iq

Kurdistan Regional Government
University of Human Development

UHD Journal of Science and Technology

A periodic scientific journal issued by University of Human Development

Editorial Board

Professor Dr. Mariwan Ahmed Rasheed……Executive publisher

Assistant Professor Dr. Aso Mohammad Darwesh……Editor-in-Chief

Professor Dr. Muzhir Shaban Al-ni……Member

Professor Dr. Salih Ahmed Hama……Member

Professor Dr. Khalid Al-Quradaghi ……Member

Assistant Professor Dr. Tara Mahmood Hassan……Member

Assistant Professor Dr. Raed Ibraheem Hamed……Member

Dr. Nurouldeen Nasih Qader……Member

Technical

Mr. Hawkar Omar Majeed…… Head of Technical

Advisory Board

Professor Dr. Sufyan Taih Faraj Aljanabi…… Iraq

Professor Dr. Salah Ismaeel Yahya…… Kurdistan

Professor Dr. Sattar B. Sadkhan…… Iraq

Professor Dr. Amir Masoud Rahmani …… Kurdistan

Professor Dr. Muhammad Abulaish…… India

Professor Dr. Parham Moradi …… Iran

Introduction

UHD Journal of Science and Technology (UHDJST) is a semi-annual journal published by the University of Human Development, Sulaymaniyah, Kurdistan Region, Iraq. UHDJST member of ROAD, e-ISSN: 2521-4217, p-ISSN: 2521-4209 and a member of Crossref, DOI: 10.21928/issn.2521-4217. UHDJST publishes original research in all areas of Science, Engineering, and Technology. UHDJST is a Peer-Reviewed Open Access journal with Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0). UHDJST provides immediate, worldwide, barrier-free access to the full text of research articles without requiring a subscription to the journal, and has article processing charge (APC). UHDJST applies the highest standards to everything it does and adopts APA citation/referencing style. UHDJST Section Policy includes three types of publications: Articles, Review Articles, and Letters.

By publishing with us, your research will get the coverage and attention it deserves. Open access and continuous online publication mean your work will be published swiftly, ready to be accessed by anyone, anywhere, at any time. Article Level Metrics allow you to follow the conversations your work has started.

UHDJST publishes works from extensive fields including, but not limited to:

- Pure Science
- Applied Science
- Medicine
- Engineering
- Technology

Scope and Focus

UHD Journal of Science and Technology (UHDJST) publishes original research in all areas of Science and Engineering. UHDJST is a semi-annual journal published by the University of Human Development, Sulaymaniyah, Kurdistan Region, Iraq. We believe that if your research is scientifically valid and technically sound then it deserves to be published and made accessible to the research community. UHDJST aims to provide a service to the international scientific community enhancing swap space to share, promote and disseminate the academic scientific production from research applied to Science, Engineering, and Technology.

SEARCHING FOR PLAGIARISM

Plagiarism Policy: The UHD Journal of Science and Technology is committed to upholding the highest standards of academic integrity and originality. As such, we do not tolerate plagiarism in any form. Authors submitting manuscripts to the journal must ensure that their work is entirely original and properly cited. Plagiarism, including self-plagiarism and the use of verbatim text from other sources without appropriate attribution, will result in immediate rejection of the submission. To maintain the integrity of scholarly research, all submissions undergo rigorous plagiarism checks using industry-standard software. Authors found to have engaged in plagiarism, at any stage before publication of the manuscript - before or after acceptance, during editing or at page proof

stage, will be barred from submitting to the journal in the future. By submitting their work to the UHD Journal of Science and Technology, authors affirm that their manuscript is their own original work and that any borrowed content is properly cited.

Section Policies

No.	Title	Peer Reviewed	Indexed	Open Submission
1	Articles: This is the main type of publication that UHDJST will produce	☑	☑	☑
2	Review Articles: Critical, constructive analysis of the literature in a specific field through summary, classification, analysis, comparison.	☑	☑	☑
3	Letters: Short reports of original research focused on an outstanding finding whose importance means that it will be of interest to scientists in other fields.	☑	☑	☑

PEER REVIEW POLICIES

At UHDJST we are committed to prompt quality scientific work with local and global impacts. To maintain a high-quality publication, all submissions undergo a rigorous review process. Characteristics of the peer review process are as follows:

- The journal peer review process is a "double-blind peer review".
- Simultaneous submissions of the same manuscript to different journals will not be tolerated.
- Manuscripts with contents outside the scope will not be considered for review.
- Papers will be refereed by at least 2 experts as suggested by the editorial board.
- In addition, Editors will have the option of seeking additional reviews when needed. Authors will be informed when Editors decide further review is required.
- All publication decisions are made by the journal's Editors-in-Chief on the basis of the referees' reports. Authors of papers that are not accepted are notified promptly.
- All submitted manuscripts are treated as confidential documents. We expect our Board of Reviewing Editors, Associate Editors and reviewers to treat manuscripts as confidential material as well.
- Editors, Associate Editors, and reviewers involved in the review process should disclose conflicts of interest resulting from direct competitive, collaborative, or other relationships with any of the authors, and remove oneself from cases in which such conflicts preclude an objective evaluation. Privileged information or ideas that are obtained through peer review must not be used for competitive gain.
- Our peer review process is confidential and the identities of reviewers cannot be revealed.

Note: UHDJST is a member of CrossRef and CrossRef services, e.g., CrossCheck. All manuscripts submitted will be checked for plagiarism (copying text or results from other sources) and self-plagiarism (duplicating substantial parts of authors' own published work without giving the appropriate references) using the CrossCheck database. Plagiarism is not tolerated.

For more information about CrossCheck/iThenticate, please visit http://www.crossref.org/crosscheck.html.

OPEN ACCESS POLICY

This journal provides immediate open access to its content on the principle that making research freely available to the public supports a greater global exchange of knowledge. Open Access (OA) stands for unrestricted access and unrestricted reuse which means making research publications freely available online. It access ensures that your work reaches the widest possible audience and that your fellow researchers can use and share it easily. The mission of the UHDJST is to improve the culture of scientific publications by supporting bright minds in science and public engagement.

UHDJST's open access articles are published under a Creative Commons Attribution CC-BY-NC-ND 4.0 license. This license lets you retain copyright and others may not use the material for commercial purposes. Commercial use is one primarily intended for commercial advantage or monetary compensation. If others remix, transform or build upon the material, they may not distribute the modified material. The main output of research, in general, is new ideas and knowledge, which the UHDJST peer-review policy allows publishing as high-quality, peer-reviewed research articles. The UHDJST believes that maximizing the distribution of these publications - by providing free, online access - is the most effective way of ensuring that the research we fund can be accessed, read and built upon. In turn, this will foster a richer research culture and cultivate good research ethics as well. The UHDJST, therefore, supports unrestricted access to the published materials on its main website as a fundamental part of its mission and a global academic community benefit to be encouraged wherever possible.

Specifically:

- The University of Human Development supports the principles and objectives of Open Access and Open Science
- UHDJST expects authors of research papers, and manuscripts to maximize the opportunities to make their results available for free access on its final peer-reviewed paper
- All manuscript will be made open access online soon after final stage peer-review finalized.
- This policy will be effective from 17th May 2017 and will be reviewed during the first year of operation.
- Open Access route is available at http://journals.uhd.edu.iq/index.php/uhdjst for publishing and archiving all accepted papers,
- Specific details of how authors of research articles are required to comply with this policy can be found in the Guide to Authors.

ARCHIVING

This journal utilizes the LOCKSS and CLOCKSS systems to create a distributed archiving system among participating libraries and permits those libraries to create permanent archives of the journal for purposes of preservation and restoration.

LOCKSS: Open Journal Systems supports the LOCKSS (Lots of Copies Keep Stuff Safe) system to ensure a secure and permanent archive for the journal. LOCKSS is open source software developed at Stanford University Library that enables libraries to preserve selected web journals by regularly polling registered journal websites for newly published content and archiving it. Each archive is continually validated against other library caches, and if the content is found to be corrupted or lost, the other caches or the journal is used to restore it.

CLOCKSS: Open Journal Systems also supports the CLOCKSS (Controlled Lots of Copies Keep Stuff Safe) system to ensure a secure and permanent archive for the journal. CLOCKSS is based upon the open-source LOCKSS software developed at Stanford University Library that enables libraries to preserve selected web journals by regularly polling registered journal websites for newly published content and archiving it. Each archive is continually validated against other library caches, and if the content is found to be corrupted or lost, the other caches or the journal is used to restore it.

PUBLICATION ETHICS

Publication Ethics and Publication Malpractice Statement

The publication of an article in the peer-reviewed journal UHDJST is to support the standard and respected knowledge transfer network. Our publication ethics and publication malpractice statement is mainly based on the Code of Conduct and Best-Practice Guidelines for Journal Editors (Committee on Publication Ethics, 2011) that includes;

- General duties and responsibilities of editors.
- Relations with readers.
- Relations with the authors.
- Relations with editors.
- Relations with editorial board members.
- Relations with journal owners and publishers.
- Editorial and peer review processes.
- Protecting individual data.
- Encouraging ethical research (e.g. research involving humans or animals).
- Dealing with possible misconduct.
- Ensuring the integrity of the academic record.
- Intellectual property.
- Encouraging debate.
- Complaints.
- Conflicts of interest.

ANIMAL RESEARCHES

- For research conducted on regulated animals (which includes all live vertebrates and/or higher invertebrates), appropriate approval must have been obtained according to either international or local laws and regulations. Before conducting the research, approval must have been obtained from the relevant body (in most cases an Institutional Review Board, or Ethics Committee). The authors must provide an ethics statement as part of their Methods section detailing full information as to their approval (including the name of the granting organization, and the approval reference numbers). If an approval reference number is not provided, written approval must be provided as a confidential supplemental information file. Research on non-human primates is subject to specific guidelines from the Weather all (2006) report (The Use of Non-Human Primates in Research).
- For research conducted on non-regulated animals, a statement should be made as to why ethical approval was not required.
- Experimental animals should have been handled according to the highest standards dictated by the author's institution.
- We strongly encourage all authors to comply with the '*Animal Research: Reporting In Vivo Experiments*' (ARRIVE) guidelines, developed by NC3Rs.
- Articles should be specific in descriptions of the organism(s) used in the study. The description should indicate strain names when known.

ARTICLE PROCESSING CHARGES

UHDJST is an Open Access Journal (OAJ) and has article processing charges (APCs). The published articles can be downloaded freely without a barrier of admission.

Address

University of Human Development, Sulaymaniyah-Kurdistan Region/Iraq
PO Box: Sulaymaniyah 6/0778

Contact

Principal Contact

Dr. Aso Darwesh

Editor-in-Chief

University of Human Development – Sulaymaniyah, Iraq

Email: jst@uhd.edu.iq

Support Contact

UHD Technical Support

Phone: +964 773 393 5959

Email: jst@uhd.edu.iq

Contents

No.	Author Name	Title	Pages
1	Nza Hamid Al-Barznjy Katan Sabir Ali	Prevalence of Hepatitis B Core Antibodies and Occult Hepatitis B Infection among Blood Donors in Erbil Governorate, Iraq	1 - 6
2	Sura Mahdi Muhammad Yaseen Ahmad Hamaamin Nihad Bahaaldeen Salih	Environmental Impact Assessment of Sulaymaniyah Solid Waste Dumpsite Using Leachate and Soil Pollution Indices	7 -18
3	Amani Fadhil Abbas Peshwaz Abdulrahman Ahmed Nazera Salam Mena Qadir Khanda Gharib Aziz	Assessment of Nurses' Knowledge and Competence in Managing Preeclampsia at Maternity Teaching Hospital in Sulaimani city	19 - 29
4	Nawshirwan Mahmood Rahim	The Effectiveness of Geographic Information Systems in Sustainable Urban Planning in Iraq: An Analytical Study of Experts' Opinions	30 -38
5	Shena Ali Faris Pakiza Aziz Saied	Climatic Impacts on Drug Therapy Usage: A Comparative Study of Kurdish Populations in Sulaimani, Iraq, and Stockholm, Sweden	39 - 60
6	Ismael Abdulkareem Ali Sozan Abdulla Mahmood	Enhancing Clinical Decision Support: A Deep Learning Approach for Automated Diagnosis of Eye Diseases from Fundus Images	61 - 76
7	Abdulrahman Tawfeeq Jalal Mohammed Anwar Mohammed	Enhanced Integer-Based Homomorphic Encryption Scheme with Windowing Mechanism for Securing Electronic Health Records	77 - 91
8	Darun B. Mohammed Ayad H. Hasan	Molecular Identification and Antibiotic Resistance Profile of Some Pseudomonas aeruginosa Clinical Isolates	92 - 100
9	Rebin Abdulkareem Hamaamin Omar Mohammed Amin Ali	Enhancing Pelican Optimization Algorithm with Differential Evolution: A Novel Hybrid Metaheuristic Approach	101 - 114
10	Hero Abdullah Mohammed Salih Omer Haji Raghad Zuhair Yousif	Intelligent System for Screening Epileptic Seizures in the Erbil Electroencephalogram Epilepsy Dataset Images Utilizing Cascaded Histogram of Oriented Gradients-Gray Level Co-occurrence Matrix Features	115 - 128
11	Awat Jamal Hasan	Demographic and Clinical Profile of Patients Undergoing Colonoscopy in the Teaching Hospital for Gastroenterology and Hepatology in Sulaymaniyah City, Kurdistan/Iraq	129 - 135
12	Aras Jalal Mhamad Karim	Statistical Analysis of Realized Volatility of Bitcoin Price using Heterogeneous	136 - 147

		Autoregressive and Generalized Autoregressive Conditional Heteroskedasticity Models	
13	Chrakhan Rahim Rashid Aryan Hamid Fatah Jamila Jalal Tahir	Medical Waste Management and Treatment Techniques: Insights from Sulaymaniyah Governorate, Kurdistan Region- Iraq	148 - 155
14	Abdalmajeed Mohammed Rahman Nawbahar Faraj Mustafa	Regression Analysis of Soil Properties for Small Dam Bodies samples in Chamchamal and Qaradagh Districts, Sulaymaniyah Governorate	156 - 165
15	Rega Jalal Hawraz N. Jabbar	A Novel Class of Hybrid Conjugate Gradient Methods for Unconstrained Optimization with Applications to Image Denoising	166 - 183
16	Aram Saleem Hakem Beitollahi	An Ensemble-based Machine Learning Framework for Advanced Distributed Denial of Service Attack Detection in Software Defined Networks	184 - 197
17	Hamreen Ahmad Aree Mohammed	Adapting F5-TTS Model to Kurdish Sorani: Diffusion-based Speech Synthesis with a Specialized Dataset	198 - 207
18	Muna M. Najeeb Sevan O. Majed	Antimicrobial Susceptibility in Acute Myeloid Leukemia Patients in Erbil	208 - 215
19	Ari A. Aziz Aree A. Mohammed	LectBench-95: Preparing a University- Lecture Corpus for A/B Evaluation of Lecture- Processing Techniques	216 - 230
20	Dlshad Abdalrahman Mahmood Muhammad Amin Muhammadali	Hybrid U-Net Architectures with ResNet50 and VGG19 for Accurate CT-Based Kidney Disease and Stone Segmentation	231 - 250
21	Fryad Hama Ibrahim Kawa Zaidan Abdulrahman	Cascading Overtopping-Induced Dam Failures in a Transboundary Basin: Hydrodynamic Modeling from Gawshan to Darbandikhan Dam	251 - 265
22	Hataw Jalal Mohammed Fawzi Abdul Azeez Salih Tofiq Ahmed Tofiq Shaniar Tahir Mohammed	An Efficient Hybrid Framework U-Net-based deep learning Technique of Early Detection of Pulmonary Nodules using LIDC-IDRI Dataset	266 - 275
23	Rozhgar A. Mustafa Arol M. Anwar lkbal M. Albarzinji	Effects of Salicylic Acid on Barley Growth and Productivity under Cadmium Stress Conditions in Rania, Kurdistan Region, Iraq	276 - 286
24	Hawkar K. Hama Aram M. Ahmed Bryar A. Hassan Abdulhady Abas Abdullah	Hybrid Encryption and Steganography for Secure Image Data Hiding	287 - 296

25	Zamdar H Rasul Karim Hawnaz Jawhar Hameed Kazia Rizgar Latif Lania Sardar Mahmood	Exploring the Association between Screen Time Exposure and Autism Spectrum Disorder in Preschool-aged Children in Sulaymaniyah	297 - 304
26	Banw Azad Mohammed Ali Awat Hassan Azeez Soran Husen Mohamad	Infant Mortality in Iraq and Iran: A Comparative and Predictive Study	305 - 312
27	Niyan Hakim Ismael Delan Jamal Qader Gona Othman Faris Awayi Ghazy Abdulkareem Chnur Talib Mohamad	Awareness of Infection Control and Barriers among the Healthcare Workers in Sulaymaniyah City, Iraq	313 - 318
28	Alla Ahmad Hassan Hemin Sardar Abdulla Tara Yousif Mawlood Rebwar Khalid Muhammed Aso M. Aladdin Tarik A. Rashid	A Multi-Account Statistical Evaluation of ChatGPT Proficiency in the Kurdish Sorani Language	319 - 334
29	Renua Hiwa Ismael Jaza Mahmood Abdullah	Address Bar Spoofing in Contemporary Web Browsers: A Taxonomy, Exploitation Study, and Mitigation Guidelines	335 - 345
30	Khanda Mustafa Ahmed	Knowledge and Attitude toward Human Papillomavirus among Students at the University of Sulaimani/City Campus	346 - 356
31	Awayi Ghazy Abdulkareem Lana Abdul Hamed Muhamed Nury Soran Husen Mohamad	Forecasting of the Infant Mortality Rate in Iraq	357 - 362

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Prevalence of Hepatitis B Core Antibodies and Occult Hepatitis B Infection among Blood Donors in Erbil Governorate, Iraq

Nza Hamid Al-Barznjy and Katan Sabir Ali

Department of Medical Microbiology, College of Health Sciences, Hawler Medical University, Erbil, Iraq

ABSTRACT

The Hepatitis B Virus (HBV) remains a considerable risk to blood transfusion safety, especially through occult hepatitis B infection (OBI), defined by undetectable Hepatitis B surface antigen (HBsAg) yet the presence of HBV DNA in the bloodstream. Identifying and investigating the prevalence of OBI is essential as these infections can get past normal screening tests, which can lead to accidental transmission through transfusion. This study aimed to evaluate the prevalence of total hepatitis B core antibody (HBcAb) and identify OBI among blood donors in Erbil Governorate, Iraq. A total of 31,631 blood donors were tested for total HBcAb between September 2024 and January 2025, using the Liaison XL chemiluminescence immunoassay machine. Out of these 31,631 blood donors, 388 (1.23%) showed positive results for the total HBcAb. Among the positive cases, 65 samples were randomly chosen to detect OBI by viral load detection using quantitative real-time polymerase chain reaction. All samples were negative for HBsAg during routine screenings. Occult OBI was detected within 17 (26.15%) of the HBcAb-positive, HBsAg-negative blood donors. Despite the application of HBcAb screening, the absence of molecular testing may continue to provide an opportunity for HBV transmission. Incorporating HBV DNA testing for positive cases may enhance the safety of blood transfusions.

Index Terms: Blood Donors, Hepatitis B Core Antibody, Occult Hepatitis B, Quantitative Polymerase Chain Reaction, Erbil

1. INTRODUCTION

The Hepatitis B virus (HBV) infection is a significant worldwide health concern. In 2022, around 254 million individuals were chronically infected with HBV, accounting for 83% of the 1.3 million viral hepatitis deaths in that year [1]. Blood and blood derivatives are important elements used to preserve life. The safety of blood and its components is crucial, as individuals are frequently transfused when in a compromised health condition. The HBV, as a transfusion-transmissible disease, continues to pose a risk in donated blood despite advancements in blood donation precautions [2]. HBV transmission through transfusion poses a challenge to guaranteeing transfusion safety, particularly in regions with high HBV prevalence [3].

The extensive use of sensitive tests for Hepatitis B surface antigen (HBsAg) screening has significantly decreased the occurrence of transfusion-transmitted HBV. However, it has been established that persons lacking detectable HBsAg could transmit HBV. Nevertheless, these diagnostic methodologies are susceptible to overlooking occult hepatitis B infection (OBI), characterized by the detection of replication-competent HBV DNA in liver tissue or blood of patients who have tested negative for HBsAg employing chemiluminescent

Access this article online

DOI: 10.21928/uhdjst.v9n2y2025.pp1-6

E-ISSN: 2521-4217
P-ISSN: 2521-4209

Copyright © 2025 Al-Barznjy and Ali. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Nza Hamid Al-Barznjy, Department of Medical Microbiology, College of Health Sciences, Hawler Medical University, Erbil, Iraq. E-mail: nza.med@cihanuniversity.edu.iq

Received: 18-05-2025 Accepted: 12-07-2025 Published: 27-07-2025

immunoassay or enzyme-linked immunosorbent assay [4], with the existence of other viral markers, such as Hepatitis B core antibody (HBcAb) [5].

The OBI was identified in blood donors with a frequency that ranges from >1% to 16%, based on the endemicity of HBV infection, the screening nucleic acid testing (NAT) employed, and the confirmatory methods utilized [6]. Despite the considerable risk of transfusion-transmitted infection with HBV associated with OBI donations, screening for OBI is challenging because of the occasional presence of extremely small viral loads and mutations within the HBV genome [7]. If detected, virus loads in the blood of persons with OBI are typically low (below 200 IU/mL) [8], while the lowest infectious dosage of OBI donor plasma through transfusion was detected as 3 IU [9]. Although the increasing quantity of research on occult HBV infections, the global incidence patterns of such infections remain unclear, and their clinical significance has been contentious [10].

Considering the established HBcAb screening standards in Iraq, particularly in Erbil Governorate, the true frequency of OBI within blood donors remains poorly examined because of the absence of routine molecular analysis. The present research seeks to address this deficiency by evaluating the incidence of HBcAb and the detection of HBV DNA within blood donors in the area.

2. METHODOLOGY

2.1. Study Population

This cross-sectional study was conducted between September 2024 and January 2025, during which a total of 31,631 blood donor candidates visited the Blood Bank in Erbil Governorate, Kurdistan Region, Iraq. Blood samples were collected from all blood donors in gel tubes (with silica-based clot activators), centrifuged at 4000 rpm for 15 min, and sent for serological analysis. A whole blood sample was put into a citrate phosphate dextrose adenine-1 blood bag and spun at 4000 rpm for 10 min to separate the plasma, which was used for molecular analysis. A questionnaire was used to collect the blood donors' demographics.

2.2. Serological Testing

The screening for detection of HBcAb was done using the Liaison XL, which is a high throughput, fully automated immunoassay analyzer using chemiluminescence technology.

For the results, the cutoff value discriminating between the presence and absence of anti-HBc has an index value of 1. Sample results were interpreted as samples with anti-HBc levels equal to or above an index value of 1 being considered non-reactive in the assay, and samples with anti-HBc levels below an index value of 1 being considered reactive in the assay.

2.3. Molecular Detection

2.3.1. DNA extraction

The DNA extraction from samples was carried out using a DNA-sorb-B nucleic acid extraction kit (AmpliSens Russia). The extraction steps provided by the manufacturer were followed after the optimization. In summary, samples were initially settled at room temperature, followed by the addition of 300 µL of Lysis Solution and 100 µL of sample to microcentrifuge tubes, which were then vortexed and incubated at 65°C for 5 min. Subsequently, 25 µL of Universal Sorbent was introduced, followed by vortexing, 5 min of incubation at room temperature, and centrifugation at 5,000 rpm for 30 s. The samples were subjected to a series of three washing procedures using Washing Solutions 1 (1 time) and 2 (2 times), which involved vortexing, centrifugation at 5,000 rpm for 30 s, and supernatant removal, with repetition of the second washing step for comprehensive cleansing. The pellet was dried out at 65°C, and DNA was eluted using 50 µL of Tris-EDTA (TE) buffer, followed by incubation at 65°C for 5 min and vortexing. Following the final centrifugation at 12,000 rpm for 1 min, the supernatant, which contains pure DNA, was taken and preserved at −80°C for further analysis.

2.3.2. Viral DNA detection and quantification

For the extracted samples, an HBV-Monitor-L polymerase chain reaction (PCR) Kit (AmpliSens®, Russia) was used to quantify viral DNA. Using specialized primers and fluorescent probes, the Conformité Européenne - *in vitro* Diagnostic certified, lyophilized real-time PCR Kit quantifies HBV DNA in human blood plasma by targeting the Precore/Core region of the HBV genome. 50 µL of each of the extracted DNA, TE buffer as a negative reaction control, Positive and Negative Extraction Controls, and Calibrators (C1L and C2L) were added to separate reaction tubes containing lyophilised PCR-mix HBV-Lyo. The tubes were closed and then transferred to the machine.

Amplification was conducted using the Rotor-Gene Q system. The PCR reaction setup includes a starter two-step hold phase, beginning at 50°C for 15 min, proceeding by 95°C for an additional 15 min, for beginning denaturation. This is followed by two cyclic phases: Cycling 1 (5 cycles) comprises denaturation at 95°C for 5 s, annealing at 60°C for 20 s, and extension at 72°C for 15 s, excluding fluorescent

signal detection. Cycling 2 (40 cycles) maintains the exact temperature and timing protocol, although it incorporates fluorescent signal detection during the annealing phase, utilizing Fluorescein Amidite for amplified HBV DNA and 6-Carboxy-4',5'-Dichloro-2',7'-Dimethoxyfluorescein for internal control DNA dyes, facilitating real-time amplification monitoring.

2.3.3. Data interpretation

The Rotor-Gene program was used to analyze data. The device automatically determined Ct values when fluorescence exceeded a threshold. A calibration curve was created using Ct values from C1L and C2L calibrators to quantify HBV DNA content in test samples (IU/mL).

2.4. Statistical Analysis

Statistical analysis was conducted utilizing the Statistical Package for Social Sciences SPSS and (R software version 4.3.2.). Descriptive statistics were employed to present the demographic information. Chi-square tests were utilized to investigate correlations among categorical variables, while Pearson and Spearman correlation tests have been used to assess connections among continuous variables. A $P < 0.05$ was considered statistically significant.

3. RESULTS

3.1. Sociodemographic Characteristics of Blood Donors

Between September 2024 and January 2025, a total of 31,631 blood donors visited the Erbil Blood Bank in the Kurdistan region of Iraq. The distribution of the blood donors on a monthly basis showed variation with a decrease in donor numbers from 6,553 in September to 5,639 in November, followed by an increase and stability in the next months, with values increasing to 6,573 in December and 6,641 in January 2025, as shown in Fig. 1.

Of the total blood donors, 31,370 (99.18%) were male, whereas 261 (0.82%) were female, demonstrating a predominantly male donor demographic.

3.2. The Prevalence of Total HBcAb among Blood Donors

Among the whole-blood donor population, 388 individuals (1.23%) showed positive for total HBcAb, comprising 382 males and six females. The probability of blood donors being male was 98.5%, with the Chi-square test indicating statistical significance regarding the sex ratio ($P < 0.05$). Chi-square examination of gender distribution over months ($P = 0.9$) revealed no significant temporal change.

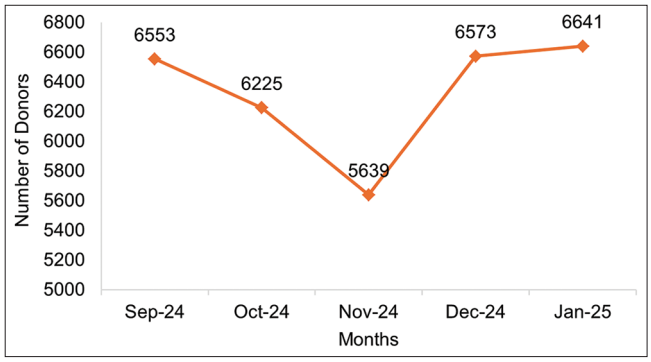


Fig. 1. Monthly distribution of blood donor numbers. The line figure reveals a decrease in donor numbers between September and November, followed by an uptrend and stability in December and January 2025.

Among the randomly selected 65 samples, the prevalence of cases according to their recency showed that the majority of cases were from Erbil city. Followed by Anbar, Ari, Bardarash, and Baghdad, with lower-case numbers from other areas.

The occupation distribution of the 65 cases showed that most HBcAb-positive cases were freelancers, followed by Peshmarga, employees, and police, as shown in Fig. 2.

The respective ages of the 65 HBcAb-positive cases varied from 23 to 59 years, having a mean age of 39.55 years, a standard deviation of 9.46, and a standard error of the mean of 1.17. A large percentage of HBcAb-positive blood donors was between 35 and 45 years of age. The highest rate was detected at the age of 40 years, as shown in Fig. 3.

3.3. Molecular Testing and Occult Hepatitis B (OBI) Detection

Among all positive cases, 65 random samples were subjected to DNA extraction and quantitative PCR (qPCR) analysis. Of them, 48 samples (73.85%) demonstrated undetectable viral loads, showing a negative OBI. Meanwhile, 17 samples exhibited detectable HBV DNA, indicating a total OBI prevalence of 26.15%. The viral load among the OBI-positive cases varied from 0.14 IU/mL (0.784 Copies/mL) to 69.64 IU/mL (570.9 Copies/mL). As shown in Figure 4.

The Spearman correlation coefficient between age and viral load showed $P = 0.282$.

4. DISCUSSION

The current study aimed to evaluate the overall prevalence of HBcAb among the blood donors attending Erbil Blood

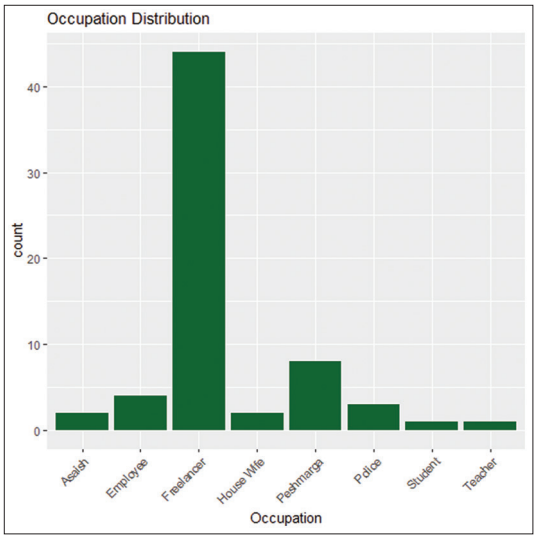


Fig. 2. The distribution of total hepatitis B core antibody-positive cases according to occupation. The figure shows that the highest prevalence was among the freelancers, followed by the Peshmarga.

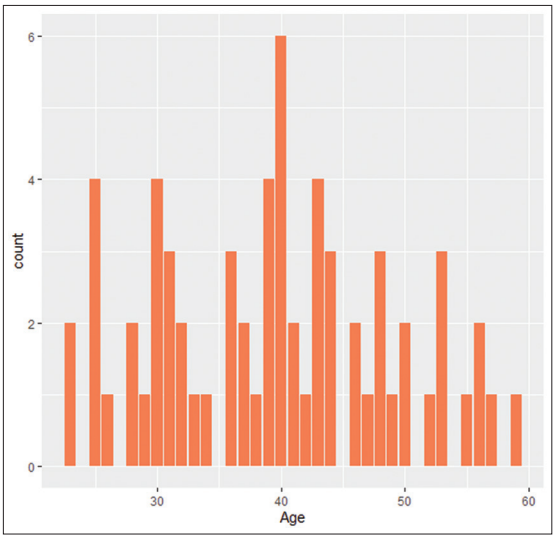


Fig. 3. The distribution of total hepatitis B core antibody-positive cases according to age. The figure shows that among the randomly selected 65 samples, ages varied from 23 to 59 years.

Bank in the Kurdistan Region of Iraq. It also assessed the possibility of OBI infection among this population through molecular analysis.

The results indicated that the prevalence of total HBcAb among the study population was 1.23% (388/31,631), including 382 males with only six females, indicating an imbalanced gender distribution. A study conducted among 438 blood donors in Duhok city, Kurdistan Region, Iraq [11], showed that the total prevalence of HBcAb was 8.2%

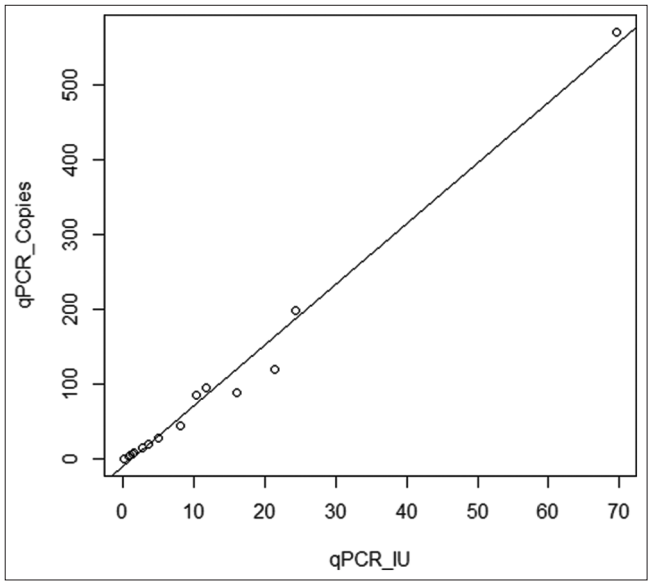


Fig. 4. The hepatitis B virus (HBV) viral load variation among occult hepatitis B infection (OBI) cases. The viral load in quantitative polymerase chain reaction-positive individuals varied from 0.14 IU/mL (0.784 copies/mL) to 69.64 IU/mL (570.9 copies/mL), showing significant diversity in HBV DNA quantities among OBI blood donors.

(36/438) in the study group. Another study in Yemen, including a total of 16,367 blood donors, revealed that the prevalence of HBcAb was 10.8% [12]. During a descriptive study in Khartoum, Sudan, blood samples were collected from 100 random blood donors to assess the prevalence of HBcAb, indicating a prevalence of 51% among the selected blood donors [13]. All the mentioned studies contrast with the findings of our study, showing a much higher prevalence among the same selected population. This can be due to the different exclusion criteria used by different blood banks and the difference in geographical distribution of HBV, as well as differences in the testing methods used.

In our study among the positive HBcAb cases, further molecular analysis on 65 random samples revealed that the percentage of detectable HBV DNA and the total prevalence of OBI among the samples was 26.15% (17/65). A study on 450 blood donors from Basrah Province in Iraq indicated that of the selected blood donors, 21.6% (97/450) were OBI positive [14]. Although some variation exists, our study's results align with this study's findings.

In a study among blood donors in the Aseer Province of Saudi Arabia, 28232 samples were screened for HBV DNA and anti-HBc to assess the prevalence and annual trends of OBI. Among the screened samples, Anti HBc reactive

samples were 3.21% (908/28232), while after matching the markers of OBI from this population, only 16 cases (18%) proved to be OBI positive [15]. In Cameroon, a study conducted on 288 HBsAg-negative blood donors revealed that 58% (167/288) of the blood donors showed positive for anti-HBc, and OBI was confirmed in 4.5% of the blood donors [16]. Furthermore, in a study conducted on 250 blood samples obtained from blood donors in Guinea, the OBI cases showed a prevalence of 15.6% [17]. While all the previous studies reveal a lower OBI prevalence, the higher detected values from our study can be due to a smaller sample size, different geographic areas, or variation in the sensitivity of the used methods.

Furthermore, the viral load detected in the OBI-positive cases in our study varied from 0.14 IU/mL (0.784 Copies/mL) to 69.64 IU/mL (570.9 Copies/mL). A study conducted in Nigeria involving 200 blood donors tested for serological markers, DNA detection, and viral load quantification revealed that among the HBcAb-positive blood donors, the HBV DNA was detected in only three cases (1.5%), showing viral loads of 753.1, 2.193 × 104, and 4.538 × 104 IU/mL [18]. In another study in Ethiopia, 973 HBsAg-negative samples were analyzed for anti-HBc and viral DNA. A total of 144 blood samples (14.8%) tested positive for anti-HBc antibodies, from which four samples (2.8% of anti-HBc-positive samples) were proven to have OBI using DNA testing. The average viral load in the verified OBI samples was 31 IU/mL [19]. The mentioned studies reported higher viral loads in comparison to our results, which can be due to differences in the immune status, genotype of the virus, or sensitivity of the utilized tests.

5. CONCLUSION

This study demonstrates the existence of occult HBV infection between HBsAg-negative blood donors in Erbil, with 26.15% of HBcAb-positive blood donors exhibiting measurable HBV DNA. Majority of the infected people were middle-aged males, and virus loads were predominantly low. These findings emphasize the inadequacies of HBsAg-only screening and recommend the integration of NAT to improve blood safety.

ACKNOWLEDGMENT

We would like to offer our deepest appreciation to the blood bank personnel for their invaluable help in collecting samples and data support during the trial. Their collaboration and commitment were crucial to the successful accomplishment of this project.

REFERENCES

[1] World Health Organization. "*Global Hepatitis Report 2024: Action for Access in Low-and Middle-Income Countries*". World Health Organization, Switzerland, 2024.

[2] World Health Organization. "*Global Hepatitis Report 2017*". World Health Organization, Switzerland, Switzerland, 2017.

[3] M. Abebe, B. Alemnew and S. Biset. "Prevalence of hepatitis B virus and hepatitis C virus among blood donors in Nekemte blood bank, Western Oromia, Ethiopia: Retrospective 5 years study". *Journal of Blood Medicine*, vol. 11, pp. 543-550, 2020.

[4] R. Raimondo, S. Locarnini, T. Pollicino, M. Levrero, F. Zoulim and A. S. Lok. "Update of the statements on biology and clinical impact of occult hepatitis B virus infection". *Journal of Hepatology*, vol. 71, no. 2, pp. 397-408, 2019.

[5] R. Nwalozie, L. S. Danagogo, J. A. Kareem and B. A. Nnokam. "The course of HBcAb as a surrogate marker for occult hepatitis B infection: A population based survey in Rivers State Nigeria". *Journal of Applied Health Sciences and Medicine*, vol. 4, no. 10, pp. 7-12, 2024.

[6] D. Fopa, D. Candotti, C. T. Tagny, C. Doux, D. Mbanya, E. L. Murphy, H. I. Kenawy, F. El Chenawi and S. Laperche. "Occult hepatitis B infection among blood donors from Yaoundé, Cameroon". *Blood Transfusion*, vol. 17, no. 6, pp. 403-408, 2019.

[7] M. Spreafico, A. Berzuini, B. Foglieni, D. Candotti, L. Raffaele, I. Guarnori, A. Colli, F. F. Maldini, J. P. Allain and D. Prati. "Poor efficacy of nucleic acid testing in identifying occult HBV infection and consequences for safety of blood supply in Italy". *Journal of Hepatology*, vol. 63, no. 5, pp. 1068-1076, 2015.

[8] M. X. Fu, A. Elsharkawy, B. Healy, C. Jackson, D. Bradshaw, E. Watkins, I. Ushiro-Lumb, J. Griffithsh, J. Neubergerb, K. Maguirei, M. Desaij, N. McDougallk, N. Priddeel, S. T. Barclaym, S. Blackmoren, P. Simmondsa, W. L. Irvingo and H. Harvala. "Occult hepatitis B virus infection: Risk for a blood supply, but how about individuals' health?" *EClinicalMedicine*, vol. 81, p. 103095, 2025.

[9] D. Candotti, S. M. Assennato, S. Laperche, J. P. Allain and S. Levicnik-Stezinar. "Multiple HBV transfusion transmissions from undetected occult infections: Revising the minimal infectious dose". *Gut*, vol. 68, no. 2, pp. 313-321, 2019.

[10] J. P. Allain. "Global epidemiology of occult HBV infection". *Annals of Blood*, vol. 2, no. 5, p. 7, 2017.

[11] N. R. Hussein. "Risk factors of hepatitis B virus infection among blood donors in Duhok city, Kurdistan Region, Iraq". *Caspian Journal of Internal Medicine*, vol. 9, no. 1, pp. 22-26, 2018.

[12] T. K. Alzubiery, T. Alhazari, J. C. Alcantara, S. A. Majed, A. S. Bazaid and A. Aldarhami. "Updated seroprevalence of hepatitis B surface antigen and anti-hepatitis core antibody among blood donors in Yemen". *Infection and Drug Resistance*, vol. 15, pp. 2787-2796, 2022.

[13] M. H. O. Ebar, H. A. M. Abdalla and M. A. A. O. Albara. "Prevalence of hepatitis B core antibody in Sudanese random blood donors at Khartoum State". *Journal of Drug Delivery and Therapeutics*, vol. 12, no. 6, pp. 44-47, 2022.

[14] Y. AlRashdan, K. Al-Jaff and M. Najdawi. "Occult hepatitis B in blood donation centers". *Journal of Medicine and Life*, vol. 16, no. 4, p. 571, 2023.

[15] A. Khanum, A. Alamri, A. Alshehr, A. Alsabrah, M. Alshehri, F. S. Alshehri and G. Albahja. "Identification of occult hepatitis B infection (OBI) among blood donors at "Aseer" Southern Region in Saudi Arabia". *Journal of Blood Disorders and Transfusion*, vol. 15, p. 572, 2024.

[16] M. N. Mbencho, N. Hafza, L. C. Cao, V. N. Mingo, E. A. Achidi, S. M. Ghogomu and T. P. Velavan. "Incidence of occult hepatitis B infection (OBI) and hepatitis B genotype characterization among blood donors in Cameroon". *PLoS One*, vol. 19, no. 10, p. e0312126, 2024.

[17] S. Boumbaly, T. A. L. Balde, A. V. Semenov, Y. V. Ostankova, E. N. Serikova, E. V. Naidenova, D. E. Valutite, A. N. Shchemelev, E. B. Zueva, E. V. Esaulenko, A. A. Totolian. "Prevalence of viral hepatitis B markers among blood donors in the Republic of Guinea". *Problems of Virology*, vol. 67, no. 1, pp. 59-68, 2022.

[18] A. S. Alshrari, S. A. Hudu, S. H. Shinkafi, A. Tahir, H. Y. Raji and A. O. Jimoh. "Prevalence and transfusion risks of occult hepatitis B infection among HBcAb-positive blood donors in a high-endemic region". *Diagnostics*, vol. 15, no. 4, p. 486, 2025.

[19] G. Gemechu, W. E. Abagez, D. H. Alemayehu, A. Tesfaye, D. Tadesse, A. Kinfu, A. Mihret and A. Mulu. "Occult hepatitis B virus infection among blood donors in the capital city of Addis Ababa, Ethiopia: Implications for blood transfusion safety". *Frontiers in Gastroenterology*, vol. 1, p. 887260, 2022.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Environmental Impact Assessment of Sulaymaniyah Solid Waste Dumpsite Using Leachate and Soil Pollution Indices

Sura Mahdi Muhammad[1], Yaseen Ahmad Hamaamin[2], Nihad Bahaaldeen Salih[1]

[1]Department of Water Resources Engineering, College of Engineering, University of Sulaimani, KRG, Iraq, [2]Department of Civil Engineering, College of Engineering, University of Sulaimani, KRG, Iraq

ABSTRACT

Global urban population is rising that resulting more waste production. Globally, municipal solid waste (MSW) generation considered as a serious threat on the global environment and human wellbeing. Leachate from solid waste dumps poses significant environmental and health risks, particularly due to contamination in soil and water caused by heavy metals. In this study, environmental impacts of MSW are assessed and estimated for Sulaymaniyah city, KRG, Iraq, which is located at 10 km south of the city in the Tanjaro dumpsite. Soil and leachate samples were collected and analyzed for various expected pollutant, to assess the environmental contamination through using pollution indices. For assessing the leachate pollution index (LPI), some parameters were determined, such as potential of hydrogen (pH), total dissolved solid, biochemical oxygen demand (BOD_5), and chemical oxygen demand (COD), and chloride (Cl). LPI value (20.1377) is much higher than the related standards. High concentrations of metals, such as cadmium (Cd), iron (Fe), manganese (Mn), copper (Cu), chromium (Cr), nickel (Ni), and zinc (Zn), found in the soil near the site, however, the contamination level is not serious based on the checked pollution indices, such as pollution index (PI) and nemerow PI ($PI_{nemerow}$). PI for Cd, Fe, Mn, Cu, Cr, Ni, and Zn were 0.158, 0.024, 0.088, 0.176, 0.613, 0.786, and 0.225, respectively, whereas, $PI_{nemerow}$ value was 0.606, which classified the soil as a non-contaminated soil. Results of this study reveals that the Tanjaro dumpsite needs an engineered landfill and decent leachate treatment right away; since present conditions far over safe limits and threaten soil and water quality.

Index Terms: Solid Waste, Leachate, Soil Pollution, Pollution Index, Sulaymaniyah City

1. INTRODUCTION

Population growth and socioeconomic development has elevated solid waste generation. The world urban population, from 1950 to 2014 has increased rapidly, going from 746 million in 1950 to 3.9 billion in 2014 [1]. The population estimated by 2025 expected to be 9.7 billion inhabitants, in Asia and Africa's urban area increase with approximately 90% [2]. Global municipal solid waste (MSW) generation was 1.3 billion tons/year in 2016 and estimated to rise to around 2.2 billion tons/year by 2025 [3]. MSW generation rates influence economic development, industrialization, local climate and public habits. Solid waste is more generated due to the high urbanization and economic development rate. When income level and urbanization growing the consumption of products elevated and then more wastes will be generated. The waste generated by city residents is double that generated in rural areas [4], [5]. Solid waste generation is rising with urbanization and richness. In high income countries, MSW is produced at 2.13 kg/capita/day whereas, it is produced

Access this article online

DOI: 10.21928/uhdjst.v9n2y2025.pp7-18

E-ISSN: 2521-4217

P-ISSN: 2521-4209

Copyright © 2025 Muhammad, *et al*. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Sura Mahdi Muhammad, Department of Water Resources Engineering, College of Engineering, University of Sulaimani, KRG, Iraq. E-mail: sura.muhammad@univsul.edu.iq

Received: 16-04-2025 Accepted: 15-07-2025 Published: 01-08-2025

at rates as low as 0.84 kg/capita/day in lower and lower middle income [1].

MSW (trash and garbage) is the everyday discarded items by public from households, businesses and other places within a municipality. The principal constituents of MSW includes food leftovers, paper, such as office paper or newspaper and packaging paper, glass, plastics, textile, metals, horticultural waste, as well as bulky refuse. There exists a direct relation between the socioeconomic status of an area and the amount of waste produced [5]–[7].

Leachate is developed rainfall infiltrating into a MSW dumpsite or landfills which allows the pollutant draining [8]. In addition, the leachate from landfill dissolution has a considerable number of compounds due to this complex composition of waste, some of them expect to be a danger to nature and health if released into ecology [9]. Leachate mainly consists of; dissolved organic matter, inorganic micro components (calcium (Ca), magnesium (Mg), potassium (K), Fe, etc.), heavy metals and xenobiotic organic constituents transferring from households or industrial chemicals (phenols, aromatic hydrocarbons, chlorinated aliphatic hydrocarbons, pesticides, etc.) [10], [11]. Leachate from the solid waste dump has a significant positive influence on the both geotechnical and chemical properties of soil which is going to be changed because of leachate effect [12]–[16].

Heavy metals can be considered as one of the hazardous contaminations in leachate, they are toxic, disrupt the natural biological balance, and prevent self-purification mechanisms [17]. Researchers conducted on the effects of MSW leachate on the land environment. Awokunmi *et al*. 2010 found the highest concentration of heavy metals was recorded within the center of each dumpsite and there was a significant decrease in concentration as the distance from dumpsite center increased but leaching did not reach 200 m and more [18]. In the study of Kanmani and Gandhimathi 2013, leachate analysis shows high value of chemical oxygen demand (COD) and a low biochemical oxygen demand (BOD_5)/COD ratio, indicating methanogenic condition, which is the fourth stage of landfill life cycle. Heavy metals, such lead Pb, Cu, Mn, and Cd, were found in the soil, signaling contamination from leachate migration [19]. Hredoy *et al*. 2022 evaluated the impact of the Amin Bazar landfill, Bangladish on surrounding environments, focusing on the contamination of leachate and its effects on surface water, groundwater, soil, and plants [20]. The study results of Mouhoun-Chouaki *et al*. 2019 showed that MSW significantly altered soil characteristics, increasing organic matter content (4.35%) and heavy metal concentrations (Cu, Zn, Cd, Pb, Ni, Cr), though they did not exceed pollution thresholds [21]. Gebre and Debelie 2015 found that soil in the area was polluted by heavy metals, with Cd being the most significant pollutant, followed by Cr, cobalt (Co), Pb, Mn, and Ni. This pollution was caused by improper open landfill practices and lack of pollution control [22]. For instance, a leachate pollution index (LPI) is defined and measured for various landfills by Kumar and Alappat 2003 [23], and the sub-leachate pollution indices are determined: LPI organic (LPI_{or}), LPI inorganic (LPI_{in}) and LPI heavy metals (LPI_{hm}).

The main objective of the present study is to evaluate the public health and environmental impacts related with the MSW dumpsite located in the Tanjaro district of Sulaymaniyah city, KRG, Iraq. In addition, the study is aiming to provide critical insights to aid decision-makers and urban planners in identifying suitable areas for waste disposal and developing more effective waste management strategies.

2. MATERIALS AND METHODS

Soil and leachate samples from the dumpsite were also systematically collected for analysis to fulfil the aim of the study. The samples were then analyzed through a series of chemical and physical tests to find out the extent of pollution. Pollution index (PI) calculated by measuring the concentration of the heavy metals in leachate and soil. This index is an important indicator of environmental pollution caused by the improper disposal of solid waste and provides a measurable index of the degree of pollution at the site.

2.1. Study Area

Tanjaro dumpsite which is existing 10 km south of the Sulaymaniyah city, KRG, Iraq and is established in 1998 [24] is shown in Fig. 1. The population of city in 2024 is 782,500 inhabitants (Sulaymaniyah statistics directorate, 2025) and the population of Sulaymaniyah was 679563 inhabitants in 2008, it's increased by 15%. The estimated solid waste generation for the city of Sulaymaniyah was approximately 1000 tons/day in 2008 [24], and according to the Presidency of Sulaymaniyah Municipality around 1000–1100 tons/day of solid waste generated by 2024, currently about half the city solid waste treated properly according to specification in (Asia Al-Taqa Al-Mutajadida) solid waste treatment plant, and the remained transfer to the Tanjaro dumpsite without treatment. MSW in this site mainly domestic waste consisting of kitchen waste, paper, glass, plastic, metal, cardboard, and fabrics. Solid waste also includes waste from the city fish market,

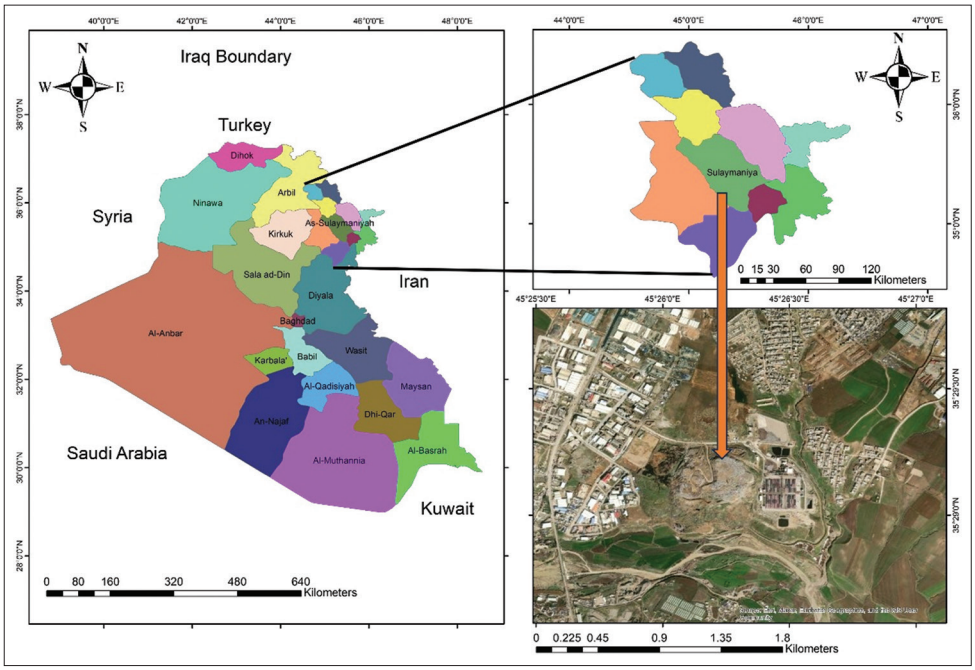


Fig. 1. Sulaymaniyah city dumpsite (Tanjaro district).

slaughterhouse, poultry market, dairy farm, non-infectious hospital waste, and car tires. The site is an unlicensed low-lying open dump where large mound of more than 12–20 m high waste existing.

2.2. Sampling and Testing of Leachate

The leachate samples of the Tanjaro dumpsite were collected in winter season when the leachate was stands as a pond in several locations from the dumpsite. The leachate samples were collected at the dumpsite randomly from two different locations, first sample was collected in December 2024 and second sample was collected in January 2025, the leachate samples locations are represented in Fig. 2. The samples were directly transferred through a polytetrafluoroethylene container to the laboratory for various physicochemical analyses. The leachate samples were analyzed for potential of hydrogen (pH), electrical conductivity (EC), total dissolved solid (TDS), COD, BOD_5, chloride (Cl), K, sodium (Na), nitrate (NO_3), phosphate (PO_4), and heavy metals (Ni, Cr, Cd, Cu, Pb, Fe, Zn).

2.3. LPI

2.3.1. LPI and sub-indices

The LPI is a method created when a group of 80 experts, who are experienced environmental engineers and had worked in the area of waste management and were distributed around the globe was selected to conduct the present survey. Based on two individual surveys, 18 leachate pollution parameters were included in the LPI. Table 1 list of 18 leachate pollutants and significance levels.

The LPI can be calculated by using (1) [25].

$$LPI = \sum_{i=1}^{n} w_i p_i \quad (1)$$

Where LPI= weighted additive LPI, wi= weight for the ith pollutant variable which is represented in Table 1; pi= sub-index score of the ith leachate pollutant variable and determined from the sub-index rating curves along with 90% confidence limits (represented as *x*-error bars) for the pollutants included in this subgroup [25]; n is the number of leachate pollutant variables applied in finding LPI; and $\sum_{i=1}^{n} w_i = 1$.

In cases of a lack of data for all leachate pollutant variables in LPI, the LPI can be determined using the available leachate pollutants concentration. If so, the LPI can be obtained by using (2).

$$LPI = \frac{\sum_{i=1}^{m} w_i p_i}{\sum_{i=1}^{m} w_i} \quad (2)$$

Where m is the quantity of leachate pollutant parameters present for the sample but in this case m <18 and $\sum_{i=1}^{m} w_i < 1$. Hence, not all the 18 leachate pollutant parameters are present

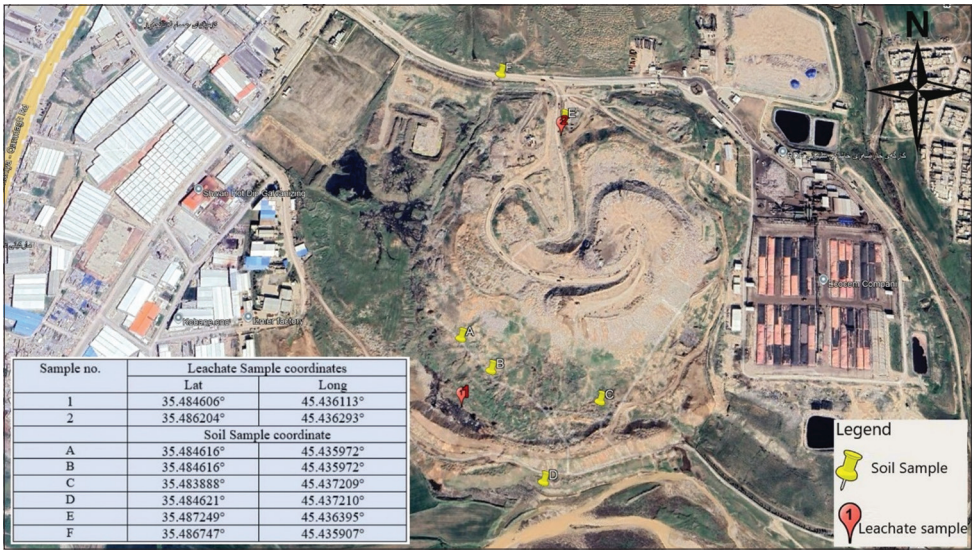


Fig. 2. Leachate and soil samples locations at Tanjaro dumpsite.

TABLE 1: The significance and weights of the 18 pollutants [25]

No.	Pollutant	Significance	Pollutant weight
1	COD	3.963	0.062
2	BOD_5	3.902	0.061
3	Total coliform bacteria	3.289	0.052
4	Phenolic compounds	3.627	0.057
5	pH	3.509	0.055
6	Chloride	3.078	0.048
7	TKN	3.367	0.053
8	Ammonia nitrogen	3.250	0.051
9	Total dissolved solid	3.196	0.050
10	Lead	4.019	0.063
11	Mercury	3.923	0.062
12	Zinc	3.585	0.056
13	Arsenic	3.885	0.061
14	Total chromium	4.057	0.064
15	Iron	2.830	0.045
16	Copper	3.170	0.050
17	Nickel	3.321	0.052
18	Cyanide	3.694	0.058
	Total	63.165	1.000

in the Tanjaro leachate samples, so the LPI will be calculated by using (2).

2.3.2. Sub-indices of LPI

To make the LPI a more informative and applicable index for the scientific community and practitioners, the LPI is broken down into three sub-indices. These three component sub-indices are derived from the leachate quality and they represent the high frequency pollutants in the landfill leachate. Leachate can be described as having physical, chemical, and bio compositions. The physical characteristics of leachate are color, odor, solids, temperature, etc., while the chemical components can be classified into organic and inorganic matter. The organisms involved are some of numerous viruses and pathogens which have been found in leachate. Other contaminants of concern include biodegradable organics, heavy metals and other dissolved inorganics.

The eighteen leachate pollutant variables used in the LPI are divided by the following three components groups to generate three sub-LPIs, as indicated in Table 2 [25].

However, because only 11 of those variables are available in the Tanjaro leachate samples, the sub-LPIs are divided as follow; LPI_{or} is consists of COD and BOD_5; LPI_{in} is consists of pH, TDS, and Cl; LPI_{hm} is consists of Cr, Pb, Zn, Ni, Cu, and Fe. The wi used in calculating the sub-indices is represented in Table 3 [25].

2.4. Sampling and Testing of Soil

The dumpsite is illegal and not according to specification. Solid waste dumped without proper engineered landfill. Soil samples collected after removal of the dumped solid waste. We dug down about 60 cm with a hand auger to gather the samples. From six different spots within the dump, we collected around 300–500 g of soil each and placed them in clean containers. The soil samples loactions are represented in Fig. 2. We then sent the soil samples to a lab to test for different chemical elements. We specifically looked for heavy metals, such as Fe, Mn, Zn, Ni, Pb, Cr, Cu, and Cd to see how polluted the soil samples were.

2.5. Soil PI

2.5.1. Single PI

PI exhibits the contamination degree of each heavy metal, which is calculated by using (3).

$$PI = \frac{C_i}{S_i} \quad (3)$$

Ci and Si refer to the heavy metal concentration in a soil sample and its allowable limit, respectively [26]. Based on this, we can divide the soil into five categories based on pollution levels. If the PI is <1, the soil is considered clean. On the other hand, if the PI is >5, the soil is highly polluted. PI classification is shown in Table 4 [27].

TABLE 2: Sub-LPIs of leachate pollutant variables

Index	Variables
LPI_{or}	COD
	BOD_5
	Phenolic compounds
	TCB
LPI_{in}	pH
	TKN
	Ammonia nitrogen
	TDS
	Cl
LPI_{hm}	Chromium
	Lead
	Mercury
	Arsenic
	Cyanide
	Zinc
	Nickel
	Copper
	Iron

COD: Chemical oxygen demand, BOD_5: Biochemical oxygen demand, TDS: Total dissolved solid, LPI: Leachate pollution index

TABLE 3: The sub-index weight factors of the LPI parameters

Index	Parameters	Weight factor
LPI_{or}	BOD_5	0.263
	COD	0.267
	Phenolic compounds	0.246
	TCB	0.224
LPI_{in}	pH	0.214
	TKN	0.206
	Ammonia nitrogen	0.198
	TDS	0.195
	Cl	0.187
LPI_{hm}	Chromium	0.125
	Lead	0.123
	Mercury	0.121
	Arsenic	0.119
	Cyanide	0.114
	Zinc	0.110
	Nickel	0.102
	Copper	0.098
	Iron	0.088

COD: Chemical oxygen demand, BOD_5: Biochemical oxygen demand, TDS: Total dissolved solid, LPI: Leachate pollution index

2.5.2. $PI_{nemerow}$

$PI_{nemerow}$ is used for evaluating the overall soil contamination level that contains all the heavy metals that have been analyzed, the $PI_{nemerow}$ must be found by using (4) [28]:

$$PI_{nemerow} = \sqrt{\frac{\left(\frac{1}{n}\sum_{i=1}^{n} PI\right)^2 + (PI_{max})^2}{2}} \quad (4)$$

Where PI has been defined in the previous section, PI_{max} refers to the upper limit of all intensive pollution indicators, whereas, n indicates the quantity of heavy metals. Based on $PI_{nemerow}$ factor, five groups of soil contamination are given in the Table 5.

3. RESULTS AND DISCUSSION

3.1. Tanjaro Dumpsite Solid Waste Composition

The MSW of Sulaymaniyah city is sorted in all zones separately before being transported to the Tanjaro dumpsite, this MSW is separated according to the four seasons as shown in the Table 6. Fig. 3. Shows the waste components for the year 2024.

3.2. Leachate Sample Characteristics

Characteristics of leachate that collected from the Tanjaro dumpsite in Sulaymaniyah city are presented in Table 7, chemical, physical, and biological characteristics.

pH of the Tanjaro dumpsite leachate was 7.83, as a mean value of the two samples. pH value of landfill leachate of this study exceeds the recommended standard values according to Lee and Jones-Lee 1993 [29], the typical concentration values of (pH) ranged (5–7.5) for municipal landfill leachate. The leachate samples were determined to be alkaline. The EC was measured at a mean value of 40120.0 µS/cm. Major ions,

TABLE 4: Based on PI soil pollution index classification [27]

PI value	Soil pollution
PI < 1	-
1 < PI < 2	Low
2 < PI < 3	Medium
3 < PI < 5	Strong
PI > 5	Very strong

TABLE 5: Classification of soil contamination according to $PI_{nemerow}$ [27]

Class	A	B	C	D	E
$PI_{nemerow}$	≤0.7	0.7–1	1–2	2–3	≥3
Soil quality	No contamination	Very low contamination	Low contamination	Medium contamination	Extreme contamination

TABLE 6: Municipal solid waste composition based on the four seasons of 2024 year

Components	January, February, March	April, May, June	July, August, September	October, November, December	Total
Food	67.11	64.75	64.62	56.32	63.2
Paper	3.16	3.47	3.07	3.98	3.42
Plastic	7.21	7.27	9.37	9.36	8.30
Metal	3.62	4.14	3.55	5.46	4.20
Glass	2.90	2.10	2.35	2.02	2.30
Fabrics	2.88	2.93	2.93	3.33	3.0
Garden waste	1.91	1.93	2.98	4.22	2.8
Miscellaneous	11.21	13.41	11.13	15.30	12.8

All values are in tons

TABLE 7: Leachate characteristics of the Tanjaro dumpsite

Parameters	First location	Second location	Mean value
pH	8.13	7.54	7.83
EC	34000.0	46240.0	40120.0
TDS	20740.0	23800.0	22270.0
Cl	4670.0	4400.0	4535.0
Na	1560.0	1200.0	1380.0
K	4000.0	4101.0	4050.5
PO_4	92.0	62.4	77.2
COD	1200.0	2420.0	1810.0
BOD_5	910.1	1510.1	1210.1
NO_3	728.0	720.0	724.0
Ni	1.229	3.066	2.1475
Fe	10.12	39.26	24.69
Pb	0.9955	1.334	1.16475
Cu	0.158	0.226	0.192
Cd	0.046	0.014	0.03
Zn	0.431	1.87	1.1505
Cr	0.885	1.142	1.0135

Except pH and EC, all parameters are in mg/L. TDS: Total dissolved solid

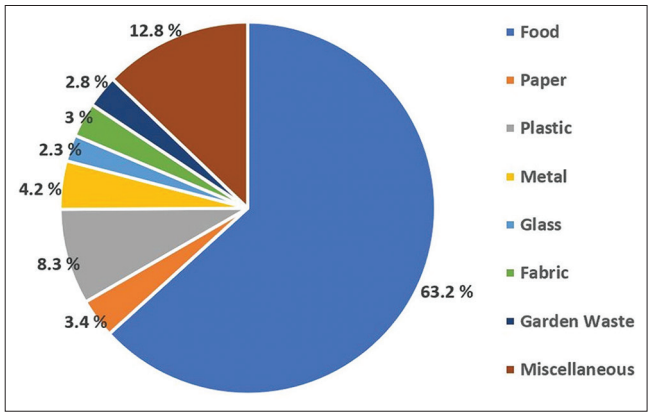


Fig. 3. Composition of MSW of Sulaymaniyah city in the year 2024.

such as Ca, Mg, Na, and K are the reasons for the presence of EC in leachate samples [30]. TDS of the Tanjaro leachate is 22270.0 mg/L, respectively, this result is approximately similar to that result of Trichy dumpsite in India [19]. The reason responsible for the high values of TDS is leaching of ions.

Cl concentration was 4535.0 mg/L, which was higher than the value as observed by De *et al.* 2016 [31]. The complexation, sorption and precipitation reactions are negligible in Cl because it is one of the inorganic macro components. Therefore, it can react as a reasonable contaminant [11]. Na can be a common contaminant since it doesn't easily form precipitates or complexes [32]. Similarly, K can also act as a contaminant and show how much the leachate is diluted [33]. This is why we found high levels of Na and K at the dumpsite, measuring 1380.0 mg/L and 4050.5 mg/L, respectively.

PO_4- P is mainly released into the leachate by the biological degradation of the organic matter containing phospholipids and phosphoproteins [34]. The concentration values of phosphate (PO_4) in the Tanjaro dumpsite first and second leachate sample were 62.4 mg/L and 92.0 mg/L, respectively, and the average mean value of 77.2 mg/L. The results showed that these concentrations are greater than the permissible limits which were recommended by Lee and Jones-Lee 1993 [29]. The mean value for concentration of COD is 1810.0 mg/L, and the concentration of BOD_5 was 1210.1 mg/L. The concentration of BOD_5 and COD was low and it showed that it was in its young stage as the ratio of BOD_5/COD was between 0.5 and 1.0 [35]. High concentration of NO_3

was observed in the Tanjaro dumpsite at a mean value of 724 mg/L. The reason that was in charge of for the reveal of NO_3 from the MSW is the proteins decomposition. The significant outcomes of nitrogen (N) pollution by ammoniacal nitrogen are the eutrophication of the aquatic resources and the increasing in oxygen demand [36].

A slightly alkaline pH, a low biodegradability ratio and a high NO_3 level appear. The Tanjaro dumpsite leachate now produces methane. Chemical reactions with sulfides, carbonates, hydroxides limited heavy metal amounts [37], this fact explains the methane production stage [11].

Thus, high concentration of Pb was found in the Tanjaro dumpsite leachate with the mean value of 1.16475 mg/L. The high Pb concentration in the Tanjaro dumpsite leachate is associated with human activities, such as the use of Pb batteries, the use of Pb in paints and pipes [38]. Leachate samples showed values of Zn (1.1505 mg/L) and Cr (1.0135 mg/L). The major sources of Zn are pesticides and fertilizers as an agro-chemicals [39]. As for Cr, it is mostly from Pb-Cr batteries, colored plastic bags, empty paint cans, and other discarded plastic stuff lying around at the dumpsite [39]. Electroplating waste and leather tanning also represent a significant chromium origin [40]. The dark blackish-brown color of the Tanjaro dumpsite leachate samples resulted from a shift in the oxidation state. Ferrous ions (Fe^{+2}) changed to ferric ions (Fe^{+3}), which caused this color change. This blackish brown color of leachates is because of the forming of ferric hydroxide colloids and fulvic complexes by Fe^{+3} [41]. Mean value of Fe was 24.69 mg/L in the dumpsite, which probably the steel-based scraps are the cause [38]. Cu concentration of the Tanjaro leachate samples was 0.192 mg/L which is within the permissible limits of recommended by [29] for components of municipal landfill leachate "typical" concentration range (0.02–1 mg/L), but Ni is not in the range whereas, it's value was 2.147 mg/L. Heavy metals in leachate might come from mixing different types of waste together at the dump [31].

3.3. Leachate Pollution Indices

For calculating LPI for the first and second location of the dumpsite, the concentrations of 18 leachate characteristics are summarized in 3rd and 6th columns, respectively, in the Table 8. The sub-index curves mentioned by [42] help us figure out the pollution ratings for different leachate pollutants. For example, the value for NO_3 was 60, which show they have the highest pollution potential at the Tanjaro dumpsite locations. Cl, BOD_5 and COD, had the medium value for pi; 57, 27 and 42 for first location; 53, 32.5 and 37 for the second location, respectively. pi for TDS was 55.5 and for total Cr was 6 for the Tanjaro dumpsite leachate.

The calculated mean LPI value for the Tanjaro leachates is 20.13, and the LPI standard value for disposal limit of treated leachate for inland surface water was 7.378 [43], that is mean

TABLE 8: LPI for the first and second leachate samples and their mean value for the Tanjaro dumpsite, Sulaymaniyah city

Parameters (1)	Weights, *wi* (2)	First location	*pi* (4)	*wipi* (5)	Second location	*pi* (7)	*wipi* (8)	Mean value
		Concentration values, mg/L (3)			Concentration values, mg/L (6)			LPI (9)
pH	0.055	8.13	2.5	0.1375	7.54	5	0.275	
TDS	0.050	20740.0	63	3.15	23800.0	55.5	2.775	
BOD_5	0.061	910.1	27	1.647	1510.0	32.5	1.9825	
COD	0.062	1200.0	57	3.534	2420.0	53	3.286	
Fe	0.045	10.12	5	0.225	39.26	5	0.225	
Cu	0.050	0.158	5.5	0.275	0.226	5.5	0.275	
Ni	0.052	1.229	5.5	0.286	3.066	7	0.364	
Zn	0.056	0.431	5	0.28	1.87	5	0.28	
Pb	0.063	0.9955	7	0.441	1.334	7	0.441	
Cr	0.064	0.885	5.5	0.352	1.142	6	0.384	
Cl	0.048	4670	42	2.016	4400.0	37	1.776	
Sum	$\sum w_i=0.606$			$\sum w_i p_i=12.343$			$\sum w_i p_i=12.07$	
LPI				$LPI1=\frac{\sum w_i p_i}{\sum w_i}$			$LPI2=\frac{\sum w_i p_i}{\sum w_i}$	LPI=20.13
				20.36			19.91	

Except pH, all parameters are in mg/L

the LPI of this study is much greater than the standard LPI. The highly contaminated dumping ground is responsible for the high value of LPI, so the leachate before being discharged into the inland surface water also the soil around the site, it must be adequately treated.

Table 9 shows the sub-LPI scores for the first and second location of the dumping site, along with the limits for treated leachate disposal in inland waters. The average values of LPI_{or}, LPI_{in}, and LPI_{hm} in the leachate samples were 42.46, 33.12, and 5.79, respectively. These values for the Tanjaro dumpsite were significantly higher than the standards for treated leachate sub-indices (LPI_{or}, 7.03; LPI_{in}, 6.57, and LPI_{hm}, 7.89) [31]. The main value of COD in LPI_{or} participating 65.24% in the ground of the Tanjaro (according to w_ip_i), the major pollutant for the LPI_{in} was TDS, which contributed about 58.51% in the dumpsite ground, in LPI_{hm} all heavy metals were approximately have the same percentage, but Pb was the greatest which contributed about 22.99%, Fig. 4.

3.4. Soil Pollution Indices

Many studies have been conducted on soil contamination issues that polluted with heavy metals, and the findings of most of those studies are that soils near industrial areas and landfill sites are polluted [44], [45]. Leachate toxicity affects human cells; therefore, this study highlights the significance of MSW, which releases contaminants into the water sources and soil [46]. Dumpsite leachate can lead to heavy metal contamination in water and soil [19]. Table 10 represents the result values of the heavy metals found in the soil samples.

In the six locations studied, compared to other heavy metals, Fe had the highest levels, whereas, Cd had the lowest level.

TABLE 9: Sub-LPI for the first and second locations and mean value in the Tanjaro dumpsite, Sulaymaniyah city

Index	Parameter	*wi*	Pollutant conc. (first location)	*pi*	*wipi* (first location)	Pollutant conc. (second location)	*pi*	*wipi* (second location)	*wipi* (mean value)	Disposal standards for leachate (inland surface water) [25] Sub-LPI
LPI_{or}	COD	0.267	1200.0	57	15.219	2420.0	53	14.151	14.685	
	BOD_5	0.263	910.1	27	7.101	1510.0	32.5	8.5475	7.824	
	Summation	$\sum w_i$=0.53			$\sum w_i p_i$=22.32			$\sum w_i p_i$=22.69	22.509	
	$LPI_{or} = \frac{\sum w_i p_i}{\sum w_i}$				42.11			42.81	42.46	7.03
LPI_{in}	pH	0.214	8.13	2.5	0.535	7.54	5	1.07	0.8025	
	TDS	0.195	20740.0	63	12.285	23800	55.5	10.8225	11.55	
	Cl	0.187	4670.0	42	7.854	4400	37	6.919	7.386	
	Summation	$\sum w_i$=0.596			$\sum w_i p_i$=20.67			$\sum w_i p_i$=18.81	19.738	
	$LPI_{in} = \frac{\sum w_i p_i}{\sum w_i}$				34.68			31.56	33.12	6.57
LPI_{hm}	Cr	0.125	0.885	5.5	0.6875	1.142	6	0.75	0.718	
	Pb	0.123	0.9955	7	0.861	1.334	7	0.861	0.861	
	Zn	0.110	0.431	5	0.55	1.87	5	0.55	0.55	
	Ni	0.102	1.229	5.5	0.561	3.066	7	0.714	0.637	
	Cu	0.098	0.158	5.5	0.539	0.226	5.5	0.539	0.539	
	Fe	0.088	10.12	5	0.44	39.26	5	0.44	0.44	
	Summation	$\sum w_i$=0.646			$\sum w_i p_i$=3.638			$\sum w_i p_i$=3.854	3.745	
	$LPI_{hm} = \frac{\sum w_i p_i}{\sum w_i}$				5.63			5.96	5.79	7.89

All parameters are in mg/L, except pH. COD: Chemical oxygen demand, BOD_5: Biochemical oxygen demand, TDS: Total dissolved solid, LPI: Leachate pollution index

As shown in Fig. 5, the metal amounts in the soil varied across different sites. The high levels of Fe, which range from 69.0% to 83.23%, are likely due to the dumping of tin-based waste and metal scraps. On the other hand, Cd levels are below 1 mg/kg because the soil is good at absorbing metals. Sources, such as used batteries, paint, pipes, and electrical waste contribute to Cu emissions [38], [47]. Factors that influence the heavy metal levels in the soil and nearby water around landfills include the shape of the land, the soil's ability to absorb metals, and the proximity of water sources to the landfill [48]. Moving various pollutants, especially heavy metals, from the landfill's liquid waste to water sources nearby is a big problem in managing this waste. Many developing countries have been studying this issue [49].

Generally, heavy metal concentrations in soil relies on how concentrated the metals are in the liquid that seeps through it [38], the soil particles absorption rate for metals [19],

TABLE 10: Tanjaro dumpsite heavy metals concentrations in the specific soil samples (mg/kg)

Locations	Ni	Fe	Pb	Cu	Cd	Zn	Cr	Mn	Total
A	56.125	783.61	ND	9.75	0.0447	21.76	84.375	72.85	1028.52
B	69.5	775.33	ND	18.59	0.149	35.55	127.5	169.65	1196.27
C	10.875	623.55	ND	2.99	0.149	9.66	19.9875	81.94	749.15
D	24.125	744.67	ND	9.89	0	28.30	45.875	172.18	1025.04
E	42.875	757.98	ND	4.89	0.0596	26.64	113.25	152.09	1097.78
F	36	761.70	ND	7.84	0.1192	20.40	77.875	153.13	1057.06

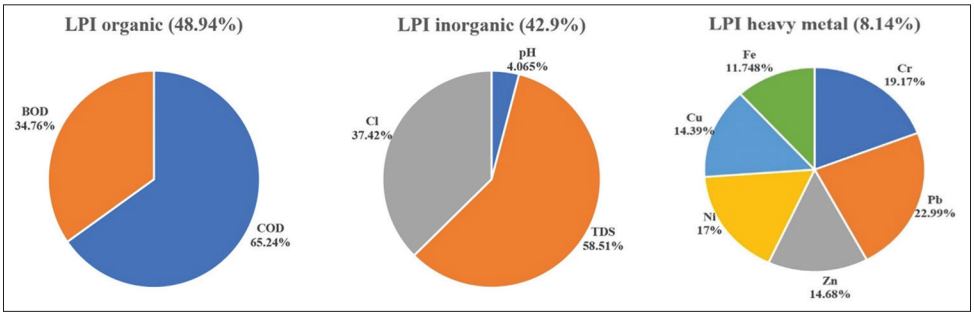


Fig. 4. Distribution of the mean value of the elements of: LPI_{or}, $LPI_{in,}$ and LPI_{hm} for the leachate of the Tanjaro dumpsite, Sulaymaniyah city.

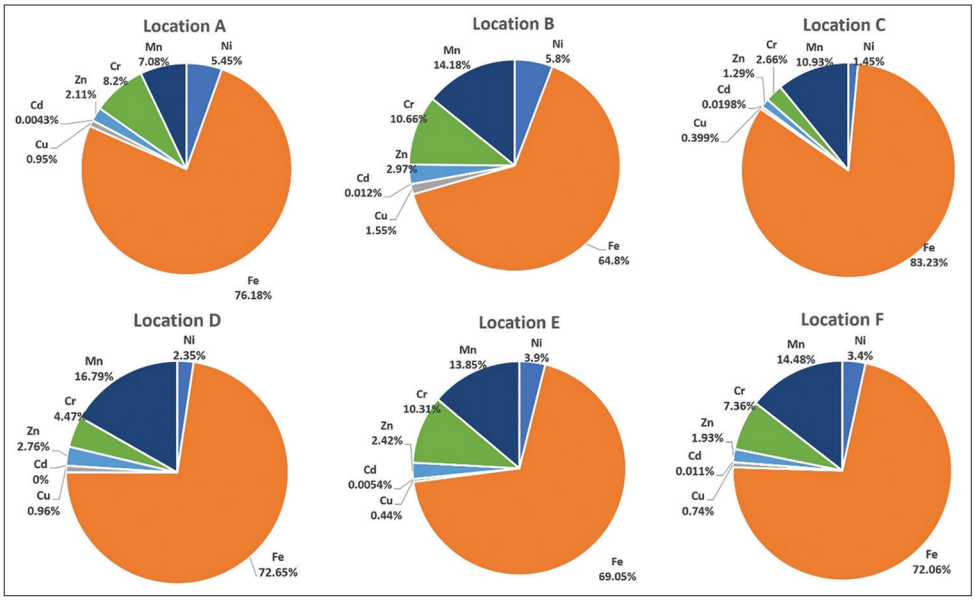


Fig. 5. Percentage of heavy metals in the soil of the six locations.

TABLE 11: The PI and $PI_{nemerow}$ calculated for the study area locations

Sample	$PI_{nemerow}$	PI						
		Cd	Fe	Mn	Cu	Cr	Ni	Zn
A	0.816	0.0812	0.0261	0.0482	0.1911	0.6617	1.1059	0.2072
B	1.032	0.2709	0.0258	0.1123	0.3645	1.0	1.3694	0.3380
C	0.2106	0.2709	0.0207	0.0542	0.0586	0.1567	0.2142	0.092
D	0.366	0	0.0248	0.1140	0.1939	0.3598	0.4753	0.2695
E	0.6702	0.1083	0.0252	0.1007	0.0958	0.8882	0.8448	0.2537
F	0.5411	0.2167	0.0253	0.1014	0.1537	0.6107	0.7093	0.1942
S_i (mg/kg) [52]		0.55	30000	1510	51	127.5	100.5	105

$PI = \frac{C_i}{S_i}$, C_i (from Table 10)

precipitation amount, and long-term accumulation of the metal in soil [50]. When there is more metal in the leachate, the levels of heavy metals in the soil arise, and the absorption ability of ions by the soil also increases. During the rainy season, the amount and concentration of heavy metals in the contaminated soil are lower [50].

The levels of metals in soil samples are as follows: Fe> Mn> Cr> Ni> Zn> Cu> Cd. PI and $PI_{nemerow}$ have been used to figure out the pollution levels at the Tanjaro dumpsite, which can be noted in Table 10. For most heavy metals in the soil samples, the PI showed they were considered non-contaminated, except for Ni in sample A, Cr and Ni in sample B, which fell into a low contamination category. As noted in Table 5, the $PI_{nemerow}$ for the soil at the Tanjaro dumpsite is also labeled as non-contaminated. In comparison, the PI and $PI_{nemerow}$ for soil samples from Tehran's landfill are also non-contaminated; this improvement is due to effective waste separation operations at the Aradkooh Complex, Iran [27]. The management of industrial and hazardous waste has stopped the inflow of domestic waste from the current waste type. That's why proper disposal of industrial waste is crucial for keeping the levels of heavy metals low. Another analysis suggested that the land around the dumpsite was polluted with heavy metallic elements, but the ground beneath the dumpsite was even more polluted [51]. The impact of leachate heavy metals on soil pollution can be identified by comparing their concentration with the allowable limits [27].

4. CONCLUSION

The following conclusions can be drawn from the conducted experimental work of this study:

- About half of the MSW from Sulaymaniyah city goes to the Tanjaro dumpsite, which has hazardous implications and degrading the environment. Results of leachate and soil samples from the dumpsite area showed diverse contaminations, which need urgent movement toward proper solid waste treatment.
- The penetration of available leachate in the Tanjaro dumpsite area yielded in diverse contamination levels. Urgent actions are proposed to be carried out toward proper solid waste treatment.
- The study outcomes yielded in an average heavy metal concentrations regarding the two studied locations in the Tanjaro dumpsite, which is 30.38 mg/L. Fe concentration has been found to as the greatest concentration, which comprised 78.33% of all metals, whereas, Cd got the lowest percentage and constituted 0.18% of the total heavy metals.
- The obtained LPI value shows that the area is very unbalanced and poses risks for leachate from the Tanjaro dumpsite. LPI value found to be 20.1377, meaning that the collected leachate from the Tanjaro exceeds the allowable limit (the highest acceptable LPI value for leachate in inland surface water is 7.378).Furthermore, the obtained sub-LPI values, LPI_{or} = 42.46; LPI_{in} = 33.12 and LPI_{hm} = 5.79 mean that these values are much higher than the leachate disposal standards of 7.03, 6.57 and 7.89 for LPI_{or}, $LPI_{in,}$ and LPI_{hm}, respectively, for inland surface water.
- The average value of total heavy metals in the checked soil samples is 1025.63 mg/kg. Mn and Fe found to be the greatest concentrations, 13.03% and 72.26%, respectively. Cu and Cd found to be the lowest concentrations, 0.87% and 0.0085%, respectively, among the metals accumulated in the soil samples. The average value of $PI_{nemerow}$ for checked soil samples found to be 0.606, which is classified within the non-contamination class.
- The uncontrolled changes that happen regularly in the dumpsite of the Tanjaro are increasing the risk level,

which requires a need to set up a proper landfill for the existing solid waste.

5. AKNOWLEDGMENT

Authors of this research would like to thank the Presidency of Sulaymaniyah Municipality for their help to conduct the study.

REFERENCES

[1] D. Hoornweg and P. Bhada-Tata. "*What a Waste: A Global Review of Solid Waste Management*", 2012. Available from: https://openknowledge.worldbank.org/entities/publication/1a464650-9d7a-58bb-b0ea-33ac4cd1f73c [Last accessed on 2025 Mar 02].

[2] UNEP. "*Beyond an Age of Waste*". UNEP, Kenya, 2024.

[3] S. Lee, J. Kim and W. K. Chong. "The causes of the municipal solid waste and the greenhouse gas emissions from the waste sector in the United States". *Waste Management*, vol. 56, pp. 593-599, 2016.

[4] A. Liu, F. Ren, W. Y. Lin and J. Y. Wang. "A review of municipal solid waste environmental standards with a focus on incinerator residues". *International Journal of Sustainable Built Environment*, vol. 4, no. 2, pp. 165-188, 2015.

[5] A. Pires, G. Martinho, S. Rodrigues and M. I. Gomes. "*Sustainable Solid Waste Collection and Management*". Springer International Publishing, Cham, 2019.

[6] B. T. Abur, E. E. Oguche and G. A. Duvuna. "Characterization of municipal solid waste in the federal capital Abuja, Nigeria". *Global Journal of Science Frontier Research: H Environment and Earth Science*, vol. 14, no. 2, pp. 1-7, 2014.

[7] A. Le Courtois. "Municipal solid waste: Turning a problem into resource". *Private Sector and Development*, vol. 15, pp. 1-28, 2012.

[8] A. Mojiri, J. L. Zhou, H. Ratnaweera, A. Ohashi, N. Ozaki, T. Kindaichi and H. Asakura. "Treatment of landfill leachate with different techniques: An overview". *Water Reuse*, vol. 11, no. 1, pp. 66-96, 2021.

[9] C. B. Öman and C. Junestedt. "Chemical characterization of landfill leachates-400 parameters and compounds". *Waste Management*, vol. 28, no. 10, pp. 1876-1891, 2008.

[10] A. Baun, A. Ledin, L. A. Reitzel, P. L. Bjerg and T. H. Christensen. "Xenobiotic organic compounds in leachates from ten Danish MSW landfills-chemical analysis and toxicity tests". *Water Research*, vol. 38, no. 18, pp. 3845-3858, 2004.

[11] P. Kjeldsen, M. A. Barlaz, A. P. Rooker, A. Baun, A. Ledin and T. H. Christensen. "Present and long-term composition of MSW landfill leachate: A review". *Critical Reviews in Environmental Science and Technology*, vol. 32, no. 4, pp. 297-336, 2002.

[12] F. M. Francisca and D. A. Glatstein. "Long term hydraulic conductivity of compacted soils permeated with landfill leachate". *Applied Clay Science*, vol. 49, no. 3, pp. 187-193, 2010.

[13] E. Panahpour, A. Gholami and A. H. Davami. "Influence of garbage leachate on soil reaction, salinity and soil organic matter in east of Isfahan". *Advances in Environmental Biology*, vol. 5, no. 8, pp. 2336-2342, 2011.

[14] S. Pillai, A. E. Peter, B. M. Sunil and S. Shrihari. "Soil Pollution Near a Municipal Solid Waste Disposal Site in India". In: *International Conference on Biological, Civil and Environmental Engineering (BCEE-2014)*, 2014. Available from: https://www.researchgate.net/profile/anju-peter-2/publication/324795096_soil_pollution_near_a_municipal_solid_waste_disposal_site_in_india/links/64c1007195bbbe0c6e6057e1/soil-pollution-near-a-municipal-solid-waste-disposal-site-in-india.pdf [Last accessed on 2025 Mar 08].

[15] B. M. Sunil, S. Shrihari and S. Nayak. "*Soil-Leachate Interaction and their Effects on Hydraulic Conductivity and Compaction Characteristics*", 2008. Available from: https://www.researchgate.net/profile/shrihari-surathkal/publication/266893766_soil-leachate_interaction_and_their_effects_on_hydraulic_conductivity_and_compaction_characteristics/links/5764faab08ae1658e2ee3bf1/soil-leachate-interaction-and-their-effects-on-hydraulic-conductivity-and-compaction-characteristics.pdf [Last accessed on 2025 Mar 08].

[16] E. C. Ukpong and J. C. Agunwamba. "Effect of open dumps on some engineering and chemical properties of soil". *Continental Journal Engineering Sciences*, vol. 6, no. 2, pp. 45-55, 2011.

[17] I. A. Talalaj. "Release of heavy metals from waste into leachate in active solid waste landfill". *Environment Protection Engineering*, vol. 41, no. 1, pp. 83-940, 2015.

[18] E. E. Awokunmi, S. S. Asaolu and K. O. Ipinmoroti. "Effect of leaching on heavy metals concentration of soil in some dumpsites". *African Journal of Environmental Science and Technology*, vol. 4, no. 8, pp. 495-499, 2010.

[19] S. Kanmani and R. Gandhimathi. "Assessment of heavy metal contamination in soil due to leachate migration from an open dumping site". *Applied Water Science*, vol. 3, no. 1, pp. 193-205, 2013.

[20] R. H. Hredoy, M. A. B. Siddique, M. A. Akbor, M. A. A. Shaikh and M. M. Rahman. "Impacts of landfill leachate on the surrounding environment: A case study on Amin Bazar Landfill, Dhaka (Bangladesh)". *Soil Systems*, vol. 6, no. 4, p. 90, 2022.

[21] S. Mouhoun-Chouaki, A. Derridj, D. Tazdaït and R. Salah-Tazdaït. "A study of the impact of municipal solid waste on some soil physicochemical properties: The case of the landfill of Ain-El-Hammam municipality, Algeria". *Applied and Environmental Soil Science*, vol. 2019, pp. 1-8, 2019.

[22] G. D. Gebre and H. D. Debelie. "Heavy metal pollution of soil around solid waste dumping sites and its impact on adjacent community: The case of Shashemane open landfill, Ethiopia". *Journal of Environment and Earth Science*, vol. 5, no. 15, pp. 169-178, 2015.

[23] D. Kumar and B. J. Alappat. "A Technique to Quantify Landfill Leachate Pollution". In: *Ninth International Landfill Symposium*, pp. 243-244, 2003.

[24] K. Abdullah. "*Environmental Implications of Tanjaro Waste Disposal Site in the City of Sulaimani*". Sulaimani, Iraq, 2010.

[25] D. Kumar and B. J. Alappat. "Analysis of leachate pollution index and formulation of sub-leachate pollution indices". *Waste Management and Research Journal*, vol. 23, no. 3, pp. 230-239, 2005.

[26] P. Zhou, D. Zeng, X. Wang, L. Tai, W. Zhou, Q. Zhuoma, F. Lin. "Pollution levels and risk assessment of heavy metals in the soil of a landfill site: A case study in Lhasa, Tibet". *International Journal of Environmental Research and Public Health*, vol. 19, no. 17, p. 10704, 2022.

[27] S. M. H. Beinabaj, H. Heydariyan, H. M. Aleii and A. Hosseinzadeh. "Concentration of heavy metals in leachate, soil, and plants in Tehran's landfill: Investigation of the effect of landfill age on the

intensity of pollution". *Heliyon*, vol. 9, no. 1, p. e13017, 2023.

[28] G. Qingjie, D. Jun, X. Yunchuan, W. Qingfei and Y. Liqiang. "Calculating pollution indices by heavy metals in ecological geochemistry assessment and a case study in parks of Beijing". *Journal of China University of Geosciences*, vol. 19, no. 3, pp. 230-241, 2008.

[29] G. F. Lee and A. Jones-Lee. "Groundwater pollution by municipal landfills: Leachate composition, detection and water quality significance". In: *Proceeding Sardinia '93 IV International Landfill Symposium, Sardinia, Italy*, 1993, pp. 1093-1103.

[30] O. J. Johansen and D. A. Carlson. "Characterization of sanitary landfill leachates". *Water Research*, vol. 10, no. 12, pp. 1129-1134, 1976.

[31] S. De, S. K. Maiti, T. Hazra, A. Debsarkar and A. Dutta. "Leachate characterization and identification of dominant pollutants using leachate pollution index for an uncontrolled landfill site". *Global Journal of Environmental Science and Management*, vol. 2, no. 2, p.177, 2016.

[32] A. S. Erses, T. T. Onay and O. Yenigun. "Comparison of aerobic and anaerobic degradation of municipal solid waste in bioreactor landfills". *Bioresource Technology*, vol. 99, no. 13, pp. 5418-5426, 2008.

[33] D. Demirbilek, A. Öztüfekçi Önal, V. Demir, G. Uslu and H. Arslanoglu-Isık. "Characterization and pollution potential assessment of Tunceli, Turkey municipal solid waste open dumping site leachates". *Environmental Monitoring and Assessment*, vol. 185, no. 11, pp. 9435-9449, 2013.

[34] D. Fatta, A. Papadopoulos and M. Loizidou. "A study on the landfill leachate and its impact on the groundwater quality of the greater area". *Environmental Geochemistry and Health*, vol. 21, no. 2, pp. 175-190, 1999.

[35] K. Y. Foo and B. H. Hameed. "An overview of landfill leachate treatment via activated carbon adsorption process". *Journal of Hazardous Materials*, vol. 171, no. 1-3, pp. 54-60, 2009.

[36] V. K. Gupta, H. Sadegh, M. Yari, G. R. Shahryari, B. Maazinejad and M. Chahardori. "*Removal of Ammonium Ions from Wastewater a Short Review in Development of Efficient Methods*", 2015. Available from: https://www.sid.ir/en/vewssid/j_pdf/55000620150206.pdf [last accessed on 2025 Mar 08].

[37] I. M. C. Lo. "Characteristics and treatment of leachates from domestic landfills". *Environment International*, vol. 22, no. 4, pp. 433-442, 1996.

[38] S. Mor, K. Ravindra, R. P. Dahiya, and A. Chandra. "Leachate characterization and assessment of groundwater pollution near municipal solid waste landfill site". *Environmental Monitoring and Assessment*, vol. 118, no. 1-3, pp. 435-456, 2006.

[39] V. Parth, N. N. Murthy and P. R. Saxena. "Assessment of heavy metal contamination in soil around hazardous waste disposal sites in Hyderabad city (India): Natural and anthropogenic implications". *Journal of Environmental Research and Management*, vol. 2, no. 2, pp. 27-34, 2011.

[40] J. Aravind, G. Sudha, P. Kanmani, A. J. Devisri, S. Dhivyalakshmi and M. Raghavprasad. "*Equilibrium and Kinetic Study on Chromium (VI) Removal from Simulated Waste Water Using Gooseberry Seeds as a Novel Biosorbent*", 2015. Available from: https://www.sid.ir/en/vewssid/j_pdf/55000620150305.pdf [Last accessed on 2025 Mar 08].

[41] L. M. Chu, K. C. Cheung and M. H. Wong. "Variations in the chemical properties of landfill leachate". *Environmental Management*, vol. 18, no. 1, pp. 105-117, 1994.

[42] D. Kumar and B. J. Alappat. "Evaluating leachate contamination potential of landfill sites using leachate pollution index". *Clean Technologies and Environmental Policy*, vol. 7, no. 3, pp. 190-197, 2005.

[43] D. Kumar and B. J. Alappat. "Analysis of Leachate Contamination Potential of a Municipal Landfill using Leachate Pollution Index". In: *Workshop on Sustainable Landfill Management*, 2003, p. 153. Available: https://www.researchgate.net/profile/dinesh-kathpalia/publication/266353417_analysis_of_leachate_contamination_potential_of_a_municipal_landfill_using_leachate_pollution_index/links/5631a6fa08ae506cea679848/analysis-of-leachate-contamination-potential-of-a-municipal-landfill-using-leachate-pollution-index.pdf [Last accessed on 2025 Mar 08].

[44] A. Kasassi, P. Rakimbei, A. Karagiannidis, A. Zabaniotou, K. Tsiouvaras, A. Nastis and K. Tzafeiropoulou. "Soil contamination by heavy metals: Measurements from a closed unlined landfill". *Bioresource Technology*, vol. 99, no. 18, pp. 8578-8584, 2008.

[45] C. Liu, J. Cui, G. Jiang, X. Chen, L. Wang and C. Fang. "Soil heavy metal pollution assessment near the largest landfill of China". *Soil and Sediment Contamination: An International Journal*, vol. 22, no. 4, pp. 390-403, 2013.

[46] A. Jabłońska-Trypuć, U. Wydro, E. Wołejko, A. Pietryczuk, A. Cudowski, J. Leszczyński, J. Rodziewicz, W. Janczukowicz and A. Butarewicz. "Potential toxicity of leachate from the municipal landfill in view of the possibility of their migration to the environment through infiltration into groundwater". *Environmental Geochemistry and Health*, vol. 43, no. 9, pp. 3683-3698.

[47] M. C. Z. Moturi, M. Rawat and V. Subramanian. "Distribution and fractionation of heavy metals in solid waste from selected sites in the industrial belt of Delhi, India". *Environmental Monitoring and Assessment*, vol. 95, no. 1-3, pp. 183-199, 2004.

[48] R. Alam, Z. Ahmed and M. F. Howladar. "Evaluation of heavy metal contamination in water, soil and plant around the open landfill site Mogla Bazar in Sylhet, Bangladesh". *Groundwater for Sustainable Development*, vol. 10, p. 100311, 2020.

[49] M. E. El-Mathana, N. G. Mostafa, M. M. Galal and A. Elawwad. "Assessment and simulation of a solid waste dumpsite impact on the surrounding water resources: A case study in Abu Zaabal, Egypt". *Heliyon*, vol. 7, no. 11, p. e08421, 2021.

[50] N. Vongdala, H. D. Tran, T. D. Xuan, R. Teschke and T. D. Khanh. "Heavy metal accumulation in water, soil, and plants of municipal solid waste landfill in Vientiane, Laos". *International Journal of Environmental Research and Public Health*, vol. 16, no. 1, p. 22, 2019.

[51] O. O. Afolabi and S. O. Eludoyin. "Evaluation of heavy metals and contamination status of soil around abandoned and active Nigerian dumpsites". *Journal of Geography, Environment and Earth Science International*, vol. 25, no. 10, pp. 1-11, 2021.

[52] RAK. "*Environmental Standards and Allowable Limits*". RAK, United Arab Emirates, 2020.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Assessment of Nurses' Knowledge and Competence in Managing Preeclampsia at Maternity Teaching Hospital in Sulaimani city

Amani Fadhil Abbas[1], Peshwaz Abdulrahman Ahmed[1], Nazera Salam Mena Qadir[1], Khanda Gharib Aziz[2]

[1]Department of Maternal Neonate Nursing, College of Nursing, University of Sulaimani, Sulaimani City, Kurdistan Region, Iraq, [2]Department of International Trade, College of Law and Administrative, Halabja University, Halabja City, Kurdistan Region, Iraq

ABSTRACT

Background: Preeclampsia is one of the most commonly reported complications during pregnancy, affecting approximately 2–15% of all pregnancies. It is characterized by the onset of hypertension after 20 weeks of gestation, accompanied by proteinuria, generalized edema, or evidence of organ dysfunction. This condition poses a serious threat to both maternal and fetal health, significantly increasing the risk of morbidity and mortality. According to the Centers for Disease Control and Prevention, "the number of pregnant women with high blood pressure has been increasing, with cases doubling from 1.8% in 2008 to 3.7% in 2021" reported in the United States. Risk factors for preeclampsia included race, advanced maternal age, obesity, null parity, multifetal pregnancy, and co-existing medical disorders. Managing preeclampsia is important because it can lead to complications for both the mother and baby. **Aim:** This study aimed to assess nurses' knowledge and competence in managing preeclampsia at the Maternity Teaching Hospital in Sulaimani City. **Materials and Methods:** A cross-sectional descriptive study was conducted at Maternity Teaching Hospital in Sulaimani city, Iraq, from January 5, 2024, to June 13, 2024. The study included 25 nurses working in the emergency care unit and labor room of the Obstetrics and Gynecology Department. A questionnaire format was created according to the aim of the study and delivered by a team of five experts, consisting of two parts. Part one: The sociodemographic characteristics of the nurses, and the second part assessed their knowledge and practices regarding the management of preeclampsia data were collected through a structured face-to-face questionnaire and analyzed using the Statistical Package for the Social Sciences version 24. **Results:** The majority of nurses were over 35 years old, with a mean age of 36 (standard deviation = 12.86). In terms of education, most nurses held a diploma (48%), and 64% were married. Around 56% of nurses reported participating in workshops or programs related to preeclampsia in the obstetrics field. When asked to define eclampsia, 80% of the nurses answered correctly, and the same percentage correctly identified the best anticonvulsant for managing preeclampsia. Overall, the nurses demonstrated a fair level of knowledge in managing preeclampsia. No significant association was found between age and knowledge level, suggesting age did not influence knowledge or competency in managing preeclampsia. **Conclusion and Recommendations:** Based on the study's findings, the nurses demonstrated a fair level of knowledge in managing preeclampsia, especially in treatment and prevention. To improve patient care and reduce maternal and fetal risks, addressing the gaps in nurses' knowledge through regular training courses and educational programs is essential. These efforts will enhance their qualifications and ensure better care for women with preeclampsia.

Index Terms: Preeclampsia, Maternity Hospital, Nurses' Competence and Knowledge

Access this article online

DOI: 10.21928/uhdjst.v9n2y2025.pp19-29

E-ISSN: 2521-4217

P-ISSN: 2521-4209

Copyright © 2025 Abbas, *et al*. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

1.INTRODUCTION

Hypertensive disorders during pregnancy pose a substantial threat to both maternal and fetal health. Among these conditions, preeclampsia is one of the most prevalent and clinically significant, affecting approximately 2–15%

Corresponding author's e-mail: Amani Fadhil Abbas, Department of Maternal Neonate Nursing, College of Nursing, University of Sulaimani. amany.abas@univsul.edu.iq

Received: 17-03-2025 Accepted: 21-06-2025 Published: 10-08-2025

of pregnancies globally. Preeclampsia is classified as a hypertensive disorder that develops after 20 weeks of gestation and is typically accompanied by proteinuria or generalized edema. It may also involve hematologic abnormalities such as thrombocytopenia, or signs of end-organ dysfunction including renal impairment, elevated liver enzymes, pulmonary edema, and neurological symptoms such as cerebral or visual disturbances. If left untreated or if it progresses to a severe form, preeclampsia can lead to serious maternal and fetal complications [1]. Multisystem involvement and impaired uteroplacental perfusion can result in adverse outcomes such as intrauterine growth restriction (IUGR), preterm birth, and increased perinatal morbidity and mortality. In severe cases, preeclampsia may become life-threatening for both the mother and fetus [2].

The exact cause of preeclampsia remains unclear, but several risk factors increase its likelihood. Women at higher risk include those pregnant for the 1st time, aged 35 years or older, those with a pregnancy interval of more than 10 years, or those with a family history of preeclampsia. Other risk factors include multiple pregnancies, a body mass index (BMI) of 35 kg/m^2 or higher, diabetes, previous hypertensive disorders in pregnancy, vascular disease, and kidney disease [3].

Preeclampsia is often associated with poor placental development or reduced blood flow to the placenta, which can lead to irregular blood pressure regulation in the mother. Nurses play a crucial role in identifying risk factors by taking detailed medical and family histories during a pregnant woman's first visit. They also educate expectant mothers on the dangers of preeclampsia and the importance of antenatal care. Women who do not receive proper prenatal care are seven times more likely to experience fatal complications from preeclampsia than those who do [4].

The management of preeclampsia involves a multidisciplinary approach aimed at ensuring maternal and fetal safety while minimizing complications. The strategy depends on the severity of the condition, gestational age, and the clinical status of both the mother and fetus. In mild cases, management includes close monitoring of blood pressure, regular assessment of symptoms, and ongoing evaluation of fetal growth and well-being. Lifestyle modifications, rest, and frequent prenatal visits are often recommended. In more severe cases, hospitalization may be required to manage complications, stabilize blood pressure, and conduct continuous fetal monitoring. Pharmacologic interventions such as antihypertensive medications and magnesium sulfate may be used as part of the management plan to reduce the risk of seizures and other severe outcomes. In certain cases, corticosteroids are administered to enhance fetal lung maturity if early delivery is anticipated. The timing of delivery is carefully determined based on clinical findings, with the goal of optimizing outcomes for both mother and child [5].

Nurses play a key role in ensuring timely diagnosis, patient education, and comprehensive maternal care, ultimately improving pregnancy outcomes. Nurses have a critical role in identifying and managing preeclampsia. They monitor blood pressure, assess symptoms (e.g., headaches, vision changes, and swelling), conduct urinalysis for protein, and educate pregnant women on warning signs and preventive measures. Early detection can prevent severe complications, as untreated preeclampsia increases the risk of maternal and fetal mortality. Uninformed nurses may fail to recognize danger signs, delaying treatment and leading to preterm birth, fetal growth restriction, or maternal organ damage [4].

Recent studies have highlighted a global rise in pregnancy-related hypertension, including preeclampsia, due to factors such as advanced maternal age, increasing obesity rates, and lifestyle changes. Although advancements in maternal-fetal medicine have improved outcomes in some regions, many healthcare systems still face challenges related to early detection and management of preeclampsia. In particular, nurses play a crucial role in antenatal care, yet studies suggest gaps in their knowledge and preparedness to manage this condition effectively [6]. Despite the importance of nursing competencies in improving maternal and fetal outcomes, limited research has explored nurses' knowledge and practices regarding preeclampsia in maternity hospital in Sulaimani city of Iraq. This study aims to fill that gap by assessing the knowledge and competency of nurses in managing preeclampsia, thereby identifying areas for targeted training and improvement.

2. MATERIALS AND METHODS

2.1. Design

A cross-sectional descriptive study was conducted between January 5th and June 13th, 2024, to assess nurses' knowledge and competence of nurses in nurses in managing preeclampsia at Maternity Teaching Hospital in Sulaimani City.

2.2. Study Sample

A purposive (non-probability) sampling method was used to select 25 nurses working specifically in the intensive care unit (ICU) and labor room departments at the Maternity Teaching Hospital in Sulaimani City. These nurses were directly involved in the care and management of pregnant women,

particularly those with preeclampsia, making them the most relevant group for the purpose of this study. While this targeted approach was appropriate for the study's objectives, it limits the generalizability of the findings to other nursing departments or healthcare settings.

2.3. Inclusion and Exclusion Criteria

For the study, nurses had to fulfill the following criteria: Nurses working in departments that manage pregnant women (e.g., labor room and ICU) and who consented to participate. Exclusion criteria included nurses unwilling to participate or those working in units not directly involved with maternal care (e.g., Premature ICU).

2.4. Data Tools

A questionnaire was created based on the study's objectives and it is consisted of two sections: (1) Sociodemographic information and (2) items assessing knowledge and practices related to the management of preeclampsia and eclampsia. The tool was reviewed by five domain experts. Verbal consent was obtained from all participants before data collection. Data were collected through face-to-face interviews and analyzed using the Statistical Package for the Social Sciences (SPSS) version 24.

2.5. Validity and Reliability of the Study Instrument

The study tool used to assess nurses' knowledge and competence in managing preeclampsia at Maternity Teaching Hospital in Sulaimani City was reviewed for content and face validity by a panel of five experts from Maternity and Neonatal Care Nursing, Pediatric Nursing, and Community Nursing. Based on their feedback, some minor revisions were made. The tool's reliability was confirmed with a Pearson correlation coefficient of 0.88. A pilot study with 10 nurses was conducted and included in the final analysis, as no modifications were made to the questionnaire, and they met the study's inclusion criteria. Thus, the inclusion was considered methodologically appropriate and contributed to the overall validity and robustness of the study findings.

2.6. Data Collection

Data were collected by interviewing each participant individually using a specially designed questionnaire in February 2024. Each interview lasted about 15 minutes. Verbal consent was obtained, and nurses agreed to participate in the study.

2.7. Data Analysis

SPSS for Windows, version 24, was used to analyze the data. Frequency, percentage, and Chi-square tests were used to evaluate any significant association between variables. *P* value of 0.05 was used as the cutoff for statistical significance and 0.001 for high statistical significance.

3. RESULTS

Table 1 shows that the age distribution of nurses is fairly balanced nurses, with 24% under 25 years, 36% between 25 and 35 years, and 40% over 35 years. The mean age of the nurses was 35.56 years, with a standard deviation of 12.86, indicating a diverse range of ages. Regarding education, 48% hold a Diploma in Nursing, 24% have a Master's degree or higher, 16% are midwifery school graduates, and 12% hold a Bachelor's degree, reflecting diverse qualifications. Marital status data reveal that 64% of nurses are married, 36% are single, and none were widowed in this sample.

Table 2 shows that 48% of the nurses have over 10 years of experience, reflecting a highly experienced group, while 28% have <5 years, and 24% have between 5 and 10 years. The average professional experience of the nurses was approximately 12.72 years (Mean = 12.72, standard deviation [SD] = 11.39), suggesting a mix of new and experienced nurses. In the obstetric unit, 48% of nurses have <5 years of experience, 24% have 5–10 years, and 28% have over 10 years, with an average of around 9 years (Mean = 9.02, SD = 10.59). This indicates a relatively lower level of experience in this specialized area, indicating less experience in this area. In addition, 56% of nurses reported having attended workshops or training programs related to preeclampsia, while 44% had not participated in such programs.

TABLE 1: Distribution of study samples according to demographic characteristics

Socio demographic data	Frequency	Percentage
Age of nurse (years)		
<25	6	24.0
25–35	9	36.0
>35	10	40.0
Mean±SD	35.56≅36±12.86	
Level of education in nursing		
Midwifery school	4	16.0
Diploma in nursing	12	48.0
Bachelors in nursing	3	12.0
Master degree and above	6	24.0
Marital status		
Married	16	64.0
Single	9	36.0
Widowed	0	0.0
Total	25	100.0

SD: Standard deviation

TABLE 2: Distribution of study samples according to profession background

Items	Frequency	Percentage
Years in the nursing profession		
<5	7	28.0
5–10	6	24.0
>10	12	48.0
Mean±SD	12.72≅13±11.39	
Years working [experience] in the obstetric unit		
<5	12	48.0
5–10	6	24.0
>10	7	28.0
Mean±SD	9.02≅9 ± 10.59	
Attending workshops or programs related to preeclampsia		
Yes	14	56.0
No	11	44.0
Total	25	100.0

SD: Standard deviation

TABLE 3: Distribution of study samples according to their knowledge in recognizing the correct definition of preeclampsia

Items	Frequency	Percentage	Total score
Hypertension of at least 140/90 mmHg recorded on at least two separate occasions and at least 4 h apart and in the presence of at least 300 mg protein in a 24-h collection of urine	20	80.0	20
Hypertension of at least 140/90 mmHg recorded on at least two separate occasions and at least 4 h apart without protein in urine	5	20.0	Results
Normal blood pressure with the presence of at least 300 mg protein in a 24 h collection of urine	0	0.0	Good
Total	25	100.0	

Table 3 shows that 80% of the nurses correctly identified preeclampsia as hypertension of at least 140/90 mmHg on two separate occasions, along with at least 300 mg of protein in a 24-h urine collection. However, 20% incorrectly defined preeclampsia as hypertension without proteinuria. None of the nurses (0%) mistakenly identified preeclampsia as normal blood pressure with protein in the urine. This demonstrates a high level of accurate knowledge about the definition of preeclampsia among most of the nursing staff.

Table 4 assesses nurses' knowledge of preeclampsia care, including risk factors, symptoms, complications, magnesium toxicity signs, and treatments. Each participant was evaluated on a 0–1 scale. Among the 25 nurses, scores below 50% indicated poor knowledge, 50–74% signified fair knowledge, and 75% or higher reflected good knowledge.

Table 4 regarding the risk factors for preeclampsia: The majority of nurses (80%) correctly identified chronic hypertension as a risk factor for preeclampsia ("Good"), while awareness was lower for autoimmune diseases (36%"Poor") and first-time pregnancy (28%"Poor"). Knowledge of diabetes as a risk factor was also limited (44%, "Poor"). In contrast, many recognized blood clotting disorders (60% "Fair") and previous preeclampsia (72% "Fair"). Awareness was higher for family history (72% "Fair"), multiple pregnancies (64% "Fair"), and high BMI (84% "Good'). Age was well recognized by 88% of participants ("Good") and preeclampsia before 20 weeks was identified by 56% ("Fair").

Table 5 assesses nurses' knowledge of preeclampsia symptoms. Based on the predefined classification (Good: ≥75%, Fair: 50–74%, and Poor: <50%), nurses demonstrated strong knowledge of preeclampsia symptoms, with (84% "Good") correctly identifying persistent frontal headaches, recognizing sudden swelling (88% "Good"), and identifying seizures (84% "Good"). Blurry vision was recognized by (76% "Fair") also oliguria (68% "Fair"). However, knowledge of epigastric pain was significantly lower, with only (40% "Poor") identifying it correctly. This highlights a need for further education on less noticeable but important symptoms.

Table 6 assesses nurses' knowledge of preeclampsia complications based on the predefined classification (Good: ≥75%, Fair: 50–74%, and Poor: <50%). Nurses demonstrated strong knowledge of seizures (84% "Good."), renal failure (84%"Good."), intracranial hemorrhage (84% "Good."), intrauterine death (96% "Good."), and IUGR (92% "Good."), However, awareness was moderate for other complications, including pulmonary edema (60% "Fair"), placental abruption (68% "Fair"), and HELLP syndrome (52% "Fair"). This suggests that while nurses are highly knowledgeable about the most severe complications, there is an opportunity to enhance their understanding of other risks, particularly pulmonary edema and HELLP syndrome, which require prompt recognition and management to prevent maternal and fetal morbidity.

TABLE 4: Distribution of study samples according to their knowledge toward preeclamptic women care (Risk factors for preeclampsia)

Knowledge of nurses toward preeclamptic women care	Frequency	Percentage	Total score	Results
History of chronic high blood pressure				
No	5	20.0	20	Good
Yes	20	80.0	9	Poor
Autoimmune disease				
No	16	64.0	9	Poor
Yes	9	36.0		
Being pregnant for 1st-time				
No	18	72.0	7	Poor
Yes	7	28.0		
History of diabetes				
No	14	56.0	11	Poor
Yes	11	44.0		
Blood clotting disorder thrombophilia				
No	10	40.0	15	Fair
Yes	15	60.0		
Having preeclampsia during an earlier pregnancy				
No	7	28.0	18	Fair
Yes	18	72.0		
Having a family history of preeclampsia				
No	7	28.0	18	Fair
Yes	18	72.0		
Carrying twins, triplets, or more babies at one time				
No	9	36.0	16	Fair
Yes	16	64.0		
Body mass index of 35 or more				
No	4	16.0	21	Good
Yes	21	84.0		
A pregnant person's age				
No	3	12.0	22	Good
Yes	22	88.0		
Preeclampsia typically develops before 20th week of pregnancy				
No	11	44.0	14	Fair
Yes	14	56.0		

(Contd...)

TABLE 4: (Continued)

Knowledge of nurses toward preeclamptic women care	Frequency	Percentage	Total score	Results

TABLE 5: Distribution of study samples according to their knowledge toward preeclamptic women care (Other signs and symptoms of preeclampsia in addition to hypertension and proteinuria are)

Knowledge of nurses toward preeclamptic women care	Frequency	Percentage	Total score	Results
Frontal head ache that do not go away after taking pain medication				
No	4	16.0	21	Poor
Yes	21	84.0		
Epigastric pain (Right side abdominal pain)				
No	15	60.0	10	Poor
Yes	10	40.0		
Blurry vision, light sensitivity, or seeing spots				
No	6	24.0	19	Fair
Yes	19	76.0		
Sudden swelling in the hands and face (edema)				
No	3	12.0	22	Good
Yes	22	88.0		
Seizures				
No	4	16.0	21	Good
Yes	21	84.0		
Oliguria (<500 mL of urine in 24 h)				
No	8	32.0	17	Fair
Yes	17	68.0		

TABLE 6: Distribution of study samples according to their knowledge toward preeclamptic women care (Complications of preeclampsia include which of the following?)

Knowledge of nurses toward preeclamptic women care	Frequency	Percentage	Total score	Results
Seizures				
No	4	16.0	21	Good
Yes	21	84.0		
Renal failure				
No	4	16.0	21	Good
Yes	21	84.0		
Pulmonary edema				
No	10	40.0	15	Fair
Yes	15	60.0		
Intracranial hemorrhage				
No	4	16.0	21	Good
Yes	21	84.0		
Intrauterine death				
No	1	4.0	24	Good
Yes	24	96.0		
Intrauterine growth restriction				
No	2	8.0	23	Good
Yes	23	92.0		
Placental abruption				
No	8	32.0	17	Fair
Yes	17	68.0		
HELLP syndrome				
No	12	48.0	13	Fair
Yes	13	52.0		
Yes	22	88.0		

Table 7 assesses nurses' knowledge regarding magnesium toxicity symptoms based on the predefined classification (Good: ≥75%, Fair: 50–74%, and Poor: <50%). It shows that nurses demonstrated a good understanding of magnesium toxicity symptoms, with high recognition rates for difficulty speaking or moving (84% "Good"), weakness (76% "Good"), and respiratory compromise (88% "Good."). Blurry vision (72% "Fair") and unresponsiveness (72%"Fair") were classified as moderate awareness. Although (64% "Fair")of nurses identified loss of deep tendon reflexes as a symptom, highlighting the need for further education on less obvious yet clinically critical symptoms to ensure early detection and prevent severe complications associated with magnesium toxicity.

Table 8 assesses nurses' knowledge regarding treatments for preeclampsia. Based on the predefined classification (Good: ≥75%, Fair: 50–74%, and Poor: <50%), it indicates

TABLE 7: Distribution of study samples according to their knowledge toward preeclamptic women care (Which of these are signs and symptoms of magnesium toxicity)

Difficulty speaking or moving	Frequency	Percentage	Total score	Results
No	4	16.0	21	Good
Yes	21	84.0		
Blurry vision				
No	7	28.0	18	Fair
Yes	18	72.0		
Weakness				
No	6	24.0	19	Good
Yes	19	76.0		
Unresponsive				
No	7	28.0	18	Fair
Yes	18	72.0		
Loss of deep tendon reflexes				
No	9	36.0	16	Fair
Yes	16	64.0		
Respiratory compromise				
No	3	12.0	22	Good
Yes	22	88.0		

that nurses demonstrated strong knowledge regarding the treatment of preeclampsia. A high percentage (80% "Good") recognized bed rest as a main treatment. Moreover, (96% "Good") understood the importance of a multidisciplinary team. Furthermore, (76% "Fair") correctly identified delivery of the fetus as the definitive treatment. This is particularly important, as timely delivery remains the only definitive treatment to halt disease progression and reduce maternal and fetal risks.

Table 9 assesses nurses' knowledge regarding the immediate measures of magnesium toxicity. It outlines that the majority of nurses (92%) correctly identified stopping the magnesium sulfate infusion as the essential intervention, demonstrating good knowledge. However, only (8%) recognized the need for an indwelling urinary catheter, and none (0%) identified the administration of calcium gluconate. This is concerning, as calcium gluconate is the primary antidote for magnesium toxicity and is essential for reversing life-threatening complications.

Table 10 highlights nurses' knowledge of the drug of choice for managing eclamptic fits. The majority of nurses (88%) incorrectly identified Calcium-likely referring to calcium gluconate-as the drug of choice, which is in fact the antidote for magnesium toxicity, not for treating eclamptic seizures.

Magnesium sulfate is the internationally recommended first-line medication for preventing and treating eclamptic seizures. Only 4% correctly selected Magnesium, while another 4% chose potassium, and another 4% selected electrolyte. This suggests a significant misunderstanding among nurses regarding the correct medication for managing eclamptic fits.

Table 11 assesses nurses' knowledge regarding the immediate nursing actions after a convulsion. It shows that the majority of nurses (80%) correctly identified the importance of ensuring airway patency, administering oxygen, and checking vital signs and fetal heart rate, demonstrating "Good" knowledge. While a portion of the nurses selected "sending blood and urine for testing" (20%) as priority actions, and none (0%) of the participants selected "Document the Event of eclamptic fit." It is important to note that immediate post-convulsion nursing actions should first focus on maintaining airway patency, administering oxygen, and assessing vital and fetal signs. Failure to document the event or perform laboratory evaluations may hinder follow-up care, but these are secondary to initial stabilization.

Table 12 shows the distribution of study samples according to their knowledge about nursing actions after the administration of magnesium sulfate. The correct nursing action, which includes the assessment of vital signs, patellar reflex, and

TABLE 8: Distribution of study samples according to their knowledge toward preeclamptic women care (Treatments for preeclampsia)

Knowledge of nurses toward preeclamptic women care	Frequency	Percentage	Total score	Results
One of the main treatments for preeclampsia is bed rest				
No	5	20.0	20	Good
Yes	20	80.0		
The most successful outcomes are attained with a robust team of physicians, nurses, pharmacists, and healthcare aids, with goal of providing the best possible care and improving maternal and fetal outcomes				
No	1	4.0	24	Good
Yes	24	96.0		
The definitive treatment of preeclampsia is the delivery of the fetus				
No	6	24.0	19	Good
Yes	19	76.0		

TABLE 9: Distribution of study samples according to their knowledge about magnesium toxicity

Items What are the immediate measures in case of magnesium toxicity?	Frequency	Percentage	Total score
Stop magnesium sulfate infusion/administration	23	92.0	23
Insert an indwelling urinary catheter	2	8.0	Results
Give calcium gluconate	0	0.0	Good
Total	25	100.0	

TABLE 10: Distribution of study samples according to their knowledge about management of eclamptic fit

Items What is the drug of choice for the management of eclamptic fit?	Frequency	Percentage	Total score
Calcium	22	88.0	22
Potassium	1	4.0	
Magnesium	1	4.0	Results
Electrolyte	1	4.0	poor
Total	25	100.0	

TABLE 11: Distribution of study samples according to their knowledge about immediate nursing actions after a convulsion (eclamptic fit) in a preeclamptic patient

Items	Frequency	Percentage	Total score
Document the event of eclamptic fit	0	0.0	20
Send blood and urine for testing	5	20.0	Results
Ensure airway patency administer oxygen, check vital signs and fetal heart rate	20	80.0	Good
Total	25	100.0	

urinary output, was identified by only 28% of nurses, indicating "poor" knowledge. Meanwhile, 72% of nurses only identified the need to assess vital signs, which is incomplete. None of the nurses identified the need to send blood and urine for testing. Neglecting to monitor patellar reflexes and urinary output increases the risk of missing signs of magnesium toxicity, which can lead to serious complications such as respiratory depression or renal impairment.

Table 13 illustrates the result of the study which indicates that, that factors such as age and level of education did not significantly impact nurses' knowledge of preeclampsia, highlighting a universal need for targeted training.

Table 14 shows the results of the study. Indicating that, there is no statistically significant association between nurses' demographic/workshop-related variables and their knowledge and competence in managing preeclampsia, as all *P*-values were greater than the common alpha 0.05.

Fig. 1 illustrates the distribution of nurses' knowledge levels concerning the management of preeclampsia, based on predefined classifications (Good: ≥75%, Fair: 50–74%, Poor: <50%). Among the 25 participants, 64% demonstrated fair knowledge, while 36% exhibited good knowledge. Notably, none of the nurses were categorized as having poor knowledge, indicating that all participants possessed at least a basic level of competence in managing preeclampsia.

TABLE 12: Distribution of study samples according to their knowledge about nursing actions after administration of magnesium sulfate

Items	Frequency	Percentage	Total score
Assessment of vital signs	18	72.0	7
Assessment of vital signs, and patellar reflex, urinary output	7	28.0	Results
Send blood and urine for testing	0	0.0	Poor
Total	25	100.0	

4. DISCUSSION OF THE RESULTS

Nursing care plays a pivotal role in promoting positive outcomes for both women and fetuses affected by preeclampsia. This study highlights the importance of nurses, as the frontline caregivers responsible for managing preeclamptic patients. The socio-demographic characteristics of the study sample reveal that the majority of nurses were over 35 years old, consistent with findings from [7], which suggests that the experience level of the nurses could influence their approach to care. However, while older nurses may bring more experience, the relationship between age and knowledge was not statistically significant in this study ($P > 0.05$), suggesting that factors beyond age are important for knowledge acquisition.

Regarding educational attainment, 48% of the nurses in this study held a nursing diploma. This finding aligns with Mohamed *et al.* [8], who reported a similar trend among nurses working in obstetrics and gynecology departments. In contrast, Khatun *et al.*, [9] found that more than half of the nurses in their study held a bachelor's degree, reflecting a higher academic qualification among their sample. Although our study did not find a statistically significant association between education level and knowledge ($P > 0.05$), differences in formal education could influence theoretical

TABLE 13: The association between knowledge of nurses toward preeclamptic women care and socio-demographic characteristics

Socio demographic data	Fair		Good		*P*-value
	Frequency	Percentage	Frequency	Percentage	
Age of nurse (years)					
<25	4	25.0	2	22.2	0.797
25–35	5	31.3	4	44.4	
>35	7	43.8	3	33.3	
Level of education in nursing					
Midwifery school	3	18.8	1	11.1	0.939
Diploma in nursing	7	43.8	5	55.6	
Bachelors in nursing	2	12.5	1	11.1	
MA degree and above	4	25.0	2	22.2	
Marital status					
Married	11	68.8	5	55.6	0.509
Single	5	31.3	4	44.4	
Total	16	100.0	9	100.0	

TABLE 14: The association between knowledge of nurses toward preeclamptic women care and some information data

Items	Fair		Good		*P*-value
	Frequency	Percentage	Frequency	Percentage	
Years in the nursing profession					
<5	5	31.3	2	22.2	0.699
5–10	3	18.8	3	33.3	
>10	8	50.0	4	44.4	
Years working [experience] in the obstetric unit					
<5	8	50.0	4	44.4	0.669
5–10	3	18.8	3	33.3	
>10	5	31.3	2	22.2	
Attending workshops or programs related to preeclampsia					
Yes	8	50.0	6	66.7	0.42
No	8	50.0	3	33.3	
Number of workshops or courses attended					
Yes	6	37.5	3	33.3	0.835
No	10	62.5	6	66.7	

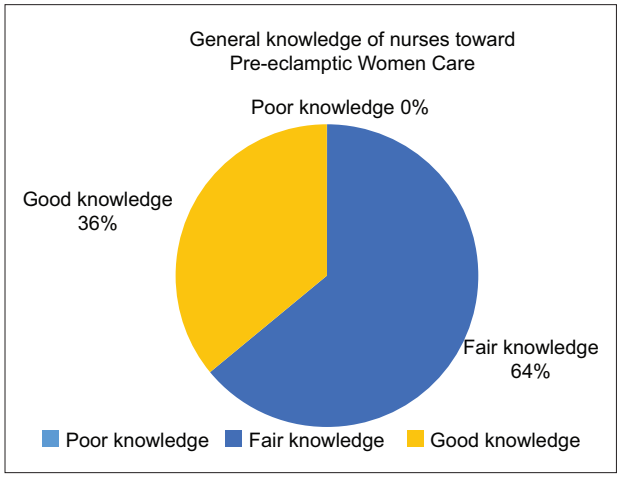


Fig. 1. Overall knowledge level of nurses regarding the management of preeclampsia.

understanding, critical thinking, and evidence-based decision-making factors that are crucial in managing preeclampsia. This highlights the need for further research into how education level affects the quality of maternal care.

Experience is another key factor influencing knowledge and care delivery. Over half of the nurses in this study had <5 years of experience, a finding consistent with Abd El Monem *et al.*[10], who observed similar trends among nurses in Khartoum. While this suggests that newer nurses may lack sufficient clinical exposure, it also offers an opportunity to improve training and mentorship programs, ensuring that less experienced nurses receive targeted support. However, it is also important to consider that variations in knowledge may be influenced by additional factors such as institutional policies, staffing levels, workload, and the availability of resources, all of which could affect nurses' ability to acquire and apply clinical knowledge effectively.

In terms of training, the majority of nurses (56%) had attended workshops or courses related to preeclampsia. However, this finding contrasts with Mkumbo and Moshi [11], who found that 77.7% of participants had not received any formal training. This discrepancy may be due to differences in the healthcare settings or the availability of professional development opportunities. Regardless, it underscores the need for consistent and continuous education programs for all nurses working with preeclamptic patients, regardless of prior training.

The study also found that 80% of nurses demonstrated good knowledge in identifying preeclampsia, aligning with ElazimElmenshewy *et al.* [12], who noted a similar level of awareness in their study. In terms of magnesium toxicity, 92% of nurses knew the correct immediate action to take, which indicates a strong understanding of this critical area of preeclampsia management. These findings are encouraging, as they reflect a solid foundation of knowledge in key aspects of care. However, statistical analysis indicated no significant correlation between knowledge levels and factors such as education level or years of experience ($P > 0.05$), suggesting that knowledge gaps may be influenced by other variables. Therefore, there is still room for improvement, particularly in broader care practices beyond immediate recognition and response.

Overall, 64% of nurses in this study had a fair knowledge of preeclampsia care, with the remaining 36% demonstrating good knowledge. This result contrasts with the study by

Faris and Faris [13], which reported poor knowledge in their sample. This variation suggests that there is no uniform level of understanding across different settings and highlights the need for targeted interventions to address these knowledge gaps. Fair knowledge may indicate that while nurses possess a basic or moderate understanding, there are gaps in critical areas such as early warning signs, complication management, or medication monitoring that could compromise patient safety and hinder timely interventions.

To improve nurses' knowledge and care for preeclamptic women, several targeted solutions can be implemented based on the study's findings. Given that 64% of nurses demonstrated only fair knowledge of preeclampsia care, standardized training programs are essential to reduce knowledge gaps and address inconsistencies in educational opportunities. Developing evidence-based educational programs focused on the clinical management of preeclampsia will ensure nurses remain up to date with current research and best practices [8]. Furthermore, since over half of the participants had <5 years of experience in obstetric care, mentorship and on-the-job training should be prioritized. Pairing experienced nurses with less experienced ones can provide hands-on guidance, build clinical confidence, and support effective knowledge transfer from senior staff [14]. These strategies are crucial for strengthening the competence of nursing staff and ultimately improving maternal and fetal outcomes.

Another key solution is the implementation of regular assessments and feedback, such as quizzes, simulations, and case-based evaluations, to measure nurses' knowledge and clinical judgment. Given that the majority of nurses in this study demonstrated only fair knowledge, such assessments are vital for identifying specific gaps and guiding targeted follow-up training sessions [15]. For example, if a significant number of nurses show limited understanding of magnesium sulfate toxicity management, focused retraining on that topic should be provided.

Moreover, promoting multidisciplinary collaboration – among nurses, obstetricians, midwives, and other healthcare providers – is essential for delivering holistic care to preeclamptic patients. This approach not only improves communication and decision-making but also reduces the risk of delayed interventions or mismanagement [13]. Given the absence of a significant association between knowledge levels and demographic factors in this study, it is evident that system-level strategies such as team-based care and institutional protocols may play a more influential role in ensuring safe outcomes. By implementing these strategies in response to the identified knowledge patterns, healthcare institutions can build stronger nursing competencies and enhance maternal and fetal outcomes in cases of preeclampsia.

5. CONCLUSION

This study shows that nurses' knowledge about managing preeclampsia needs improvement to ensure better outcomes for both mothers and babies. While some nurses demonstrated a good understanding of key areas such as the definition of preeclampsia and responding to magnesium toxicity, many showed only a fair knowledge overall. Notably, gaps were observed in broader aspects of care, including monitoring complications, interpreting clinical signs, and applying standardized care protocols. Factors such as age, education level, and years of experience appeared to influence knowledge, though not always significantly. To address these issues, it is essential to implement comprehensive training programs, provide mentorship opportunities, conduct regular knowledge assessments, and promote teamwork between nurses and other healthcare professionals. Strengthening these areas will help reduce complications and improve the quality of care provided to women with preeclampsia.

6. ACKNOWLEDGMENT

We express our sincere gratitude to all participants for their participation.

7. ETHICAL CONSIDERATIONS

The ethical committees of the College of Nursing and the College of Medicine at the University of Sulaimani approved the study. Formal authorization was then obtained from the Director of Health in Sulaimani and the administration of the Maternity Teaching Hospital. Before data collection, nurses were informed about the study's purpose, their right to confidentiality, and the voluntary nature of their participation. Verbal consent was obtained with assurances that their responses would remain anonymous and confidential. Participation was voluntary, and nurses had the right to withdraw at any time during the data collection period.

8. CONFLICTS OF INTEREST

The author affirms that they have no conflicts of interest.

9. FUNDING

The study was financially supported through author fees

REFERENCES

[1] W. Tadele, F. Debebe, A. Tadele and L. Tilahun. Assessment of knowledge and practice of nurses working in gynecology emergency rooms towards pregnancy-induced hypertension in selected government public hospitals in Addis Ababa, Ethiopia. *International Journal of Nursing and Health Science*, vol. 7, no. 3, pp. 1-10, 2020.

[2] K. J. Chang, K. M. Seow and K. H. Chen. Preeclampsia: Recent advances in predicting, preventing, and managing the maternal and fetal life-threatening condition. *International Journal of Environmental Research and Public Health*, vol. 20, no. 4, p. 2994, 2023.

[3] T. K. Beyuo, E. R, Lawrence, S. A. Oppong, E. K. Kobernik, M. Amoakoh-Coleman, D. E. Grobbee, J. L. Browne and K. W. M. Bloemenkamp. Impact of antenatal care on severe maternal and neonatal outcomes in pregnancies complicated by preeclampsia and eclampsia in Ghana. *Pregnancy Hypertension*, vol. 33, no. 3, pp. 46-51, 2023.

[4] K. Hong, S. H. Kim, D. H. Cha and H. J. Park. Defective uteroplacental vascular remodeling in preeclampsia: Key molecular factors leading to long term cardiovascular disease. *International Journal of Molecular Sciences,* vol. 22, no. 20, pp. 11202, 2021.

[5] Z. A. Colvin, M. Feng, A. Y. Pan and A. Palatnik. Duration of labor induction in nulliparous women with preeclampsia and maternal and neonatal outcomes. *American Journal of Obstetrics and Gynecology*, vol. 222, no. 1, p. S608, 2020.

[6] N. D. Ford, S. Cox and J. Y. Ko. Hypertensive disorders in pregnancy and mortality at delivery, *Morbidity and Mortality Weekly Report,* vol. 71, no. 17, pp. 585-591, 2022.

[7] S. A. Seif and S. A. Rashid. Knowledge and skills of pre-eclampsia management among healthcare providers working in antenatal clinics in Zanzibar. *BMC Health Services Research*, vol. 22, no. 1, pp .1512, 2022.

[8] S. L. Mohamed, A. S. Gonied and A. M. Elsayed. Knowledge and practices of nurses regarding nursing care for pregnant women with pre-eclampsia. *International Journal of Medical Research and Health Sciences*, vol. 11, no. 10, pp. 1-7, 2022.

[9] R. Khatun, F. Mohammad and A. A. Chowdhury. Nurse's knowledge regarding management of eclampsia at selected private hospital in Bangladesh. *International Research Journal of Natural and Applied Sciences*, vol. 4, no. 12, pp. 69-81, 2017.

[10] M. M. Abd El Monem, M. A. Shahin and H. M. Salim. Safety bundle application on nurses' knowledge and performance regarding management of hypertension during pregnancy and postpartum period. *Egyptian Nursing Journal*, vol. 18, no. 2, pp. 98-106, 2021.

[11] W. W. Mkumbo and F. V. Moshi. Nurses' level of knowledge on management of preeclampsia/eclampsia and the associated factors in Northern Tanzania: An analytical cross-sectional study. *The East African Health Research Journal*, vol. 7, no. 1, p. 49-57, 2023.

[12] S. S. A. ElazimElmenshewy, N. H. Khidr and S. O. Hassan. Investigate evidence based nursing practices in preeclampsia among pregnant woman. *IOSR Journal of Nursing and Health Science (IOSR-JNHS)*, vol. 5, no. 2, pp. 25-31, 2016.

[13] A. F. N. Faris and S. H. Faris. Nurses' knowledge regarding treatment and prevention of pregnancy induced hypertension. *International Journal of Health Sciences*, vol. 6, no. S1, pp. 8684-8690, 2022.

[14] E. M. Abdelhakm and A. R. Said. Developing nursing management protocol for maternity nurses regarding emergency obstetric care. *American Journal of Nursing Science*, vol. 6, no. 5, pp. 418-425, 2017.

[15] W. A. Alkhelaiwi, M. Traynor, K. Rogers and I. Wilson. Assessing the competence of nursing students in clinical practice: The clinical preceptors' perspective. *Healthcare (Basel)*, vol. 12, no. 10, p. 1030, 2024.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

The Effectiveness of Geographic Information Systems in Sustainable Urban Planning in Iraq: An Analytical Study of Experts' Opinions

Nawshirwan Mahmood Rahim

The Development Centre for Research and Training, University of Human Development, Sulaymaniyah, Iraq.

ABSTRACT

This descriptive analytical study assesses the effectiveness of geographic information systems (GIS) in supporting sustainable urban planning in Iraq. The study aims to gather and analyze the opinions of specialists and experts in urban planning, environmental management, and GIS across various cities in Iraq. A sample of 100 experts from Baghdad, Basra, Erbil, Mosul, and Najaf was selected based on their experience and competence in projects utilizing GIS. Data was collected through a structured questionnaire consisting of closed-ended questions and a Likert scale, focusing on four main dimensions: Experts' knowledge and use of GIS, the effectiveness of GIS in urban planning, its role in achieving sustainable development goals (SDGs), and the challenges faced in implementing these systems in Iraq. Data analysis was conducted using the Statistical Package for the Social Sciences statistical analysis software. The results showed that most experts had a good level of knowledge of GIS and considered it an effective tool in improving the accuracy of urban planning and facilitating decision-making. However, some challenges were identified, such as the lack of updated data and the technical capacity to use complex software. The study also indicated that GIS significantly contributes to achieving SDGs, especially in the areas of environmental sustainability and monitoring urban expansion. However, there is a need to improve institutional support and provide updated data and financial resources to implement GIS more effectively. Regarding statistical analysis, the results of the analysis of variance test showed significant differences in the effectiveness of GIS based on experience level, with more advanced experts showing greater effectiveness in using the system. The *t*-test revealed a significant difference between those who received formal GIS training and those who did not, with the trained group demonstrating higher knowledge levels. Finally, the correlation analysis results indicated a positive relationship between GIS knowledge and its effectiveness in urban planning. This study provides valuable insights into the effectiveness of GIS in enhancing sustainable urban planning in Iraq, highlighting the challenges faced in its implementation in this context.

Index Terms: Geographic Information Systems, Urban Planning, Urban Growth, Iraq, Environmental Sustainability, Urban Governance, Urban Infrastructure

1. INTRODUCTION

Urbanization is one of the most significant global trends in the 21st century, with cities becoming central hubs for economic, social, and cultural activities. However, the rapid expansion of urban areas presents several challenges, particularly in terms of sustainable development. Urban planners and decision-makers are increasingly seeking effective tools and methods to ensure that urban growth is balanced with environmental conservation, economic viability, and social inclusiveness. One of the most promising tools in this regard is geographic information systems (GIS and remote sensing [RS]). GIS offers a powerful framework for spatial data management, analysis,

Access this article online

DOI: 10.21928/uhdjst.v9n2y2025.pp30-38

E-ISSN: 2521-4217
P-ISSN: 2521-4209

Copyright © 2025 Rahim NM. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Nawshirwan Mahmood Rahim, The Development Centre for Research and Training, University of Human Development, Erbil, Iraq. E-mail: Nawshirwan.mahmud@uhd.edu.iq

Received: 14-05-2025 Accepted: 03-07-2025 Published: 12-08-2025

and decision-making, making it an essential tool for sustainable urban planning [1].

In Iraq, urban growth has been accelerated by factors such as population increase, rural-to-urban migration, and infrastructure development. While this growth has brought economic opportunities, it has also raised concerns about environmental degradation, resource management, and the adequacy of urban planning strategies. The integration of GIS into urban planning in Iraq is increasingly being viewed as a vital solution for addressing these challenges. Using GIS, urban planners can analyze spatial data, optimize land use, plan infrastructure development, and monitor the environmental more efficiently and comprehensively. This study seeks to explore the effectiveness of GIS in sustainable urban planning in Iraq. It focuses on gathering and analyzing expert opinions on the potential applications of GIS in the urban planning process, its challenges, and its role in fostering sustainable development. Through an analytical study, the research aims to highlight how GIS can contribute to better urban governance, improved resource management, and the promotion of environmental sustainability in Iraq's cities, also consider the data used in the studies and the methodologies adopted.

2. THEORETICAL FRAMEWORK AND PREVIOUS STUDIES

2.1. Theoretical Framework

GIS have become an essential tool in sustainable urban planning, enabling the efficient collection, analysis, and management of spatial data. GIS aids urban planners in visualizing, analyzing, and interpreting geographic information to make informed decisions regarding urban development. By integrating various data sources such as topography, land use, infrastructure, and environmental factors, GIS facilitates sustainable urban development through effective resource allocation, land use optimization, and environmental conservation [2].

Sustainable urban planning requires the consideration of multiple factors, including environmental sustainability, economic viability, and social equity. GIS GIS and RS help in the identification of suitable areas for development, efficient transport networks, green spaces, and the mitigation of environmental impacts. Furthermore, GIS and RS-based tools such as spatial decision support systems (SDSS) and satellite imagery assist in scenario modeling and impact assessment, providing decision-makers with the necessary data to prioritize projects that contribute to sustainable growth.

2.2. Previous Studies

Several studies have explored the integration of GIS in urban planning, emphasizing its role in promoting sustainable development. Lewis and Ogra (2010) highlighted the application of GIS in achieving good urban governance, suggesting that GIS facilitates transparent, efficient, and participatory planning processes, crucial for sustainable urban management (Lewis and Ogra, 2010) [1]. Maliene *et al.* (2011) also discussed how GIS technology enables urban planners to integrate old principles with new capabilities, offering greater precision in urban design (Maliene *et al.*, 2011) [2]. This capability is particularly beneficial in urban areas where rapid growth often leads to unplanned sprawl and environmental degradation.

Xhafa and Kosovrasti (2015) examined the use of GIS in urban planning, particularly in managing urban expansion and zoning, which is critical for ensuring sustainability in fast-growing cities (Xhafa and Kosovrasti, 2015) [3]. Similarly, Linh *et al.* (Khaleel and Bety, 2013) Urban Geomorphology of Sulaimani City, Using RS and GIS Techniques, Kurdistan Region, Iraq. using advanced technologies in RS and GIS to analyze the existing data enables the creation of maps depicting various types of lineaments, their distribution, and density within the Sulaimani city master plan area [13]. (2022) applied GIS and the analytic hierarchy process to select sustainable areas for urban green spaces in Hue City, Vietnam, emphasizing the importance of GIS in selecting ecologically viable sites for urban development (Linh *et al.*, 2022) [4].

Gharineiat and Khalfan (2011) explored GIS applications in sustainable transportation planning, showing how GIS can be used to design transportation systems that reduce environmental impacts and improve accessibility (Gharineiat and Khalfan, 2011) [5]. Longley (2002) pointed out the role of GIS in urban RS and its potential to improve urban geography by providing accurate spatial data for planning and policy formulation (Longley, 2002) [6]. Walsund (2013) further highlighted GIS as a tool in sustainable urban development, stressing its importance in managing land resources, monitoring urban growth, and ensuring environmental conservation (Walsund, 2013) [7].

Cabrera-Jara *et al.* (2017) utilized GIS to assess sustainable urban densification, contributing to the understanding of how GIS can help manage population growth while

minimizing environmental impacts (Cabrera-Jara *et al.*, 2017) [8]. In rural contexts, Kurowska *et al.* (2020) studied the role of GIS in promoting sustainable development in rural areas, demonstrating its versatility in both urban and rural planning scenarios (Kurowska *et al.*, 2020) [9]. Zuidgeest *et al.* (2015) emphasized GIS's contribution to sustainable urban transport systems, illustrating how spatial data can be used to optimize transportation networks for environmental and social sustainability (Zuidgeest *et al.*, 2015) [10].

Ouchra *et al.* (2022) conducted a comprehensive study on the use of RS and GIS for urban planning, focusing on the combination of satellite imagery and spatial data to improve urban development strategies (Ouchra *et al.*, 2022) [11]. Sultani *et al.* (2009) introduced GIS-based SDSS as a powerful tool for enhancing urban planning processes, particularly in terms of spatial analysis and decision-making (Sultani *et al.*, 2009) [12]. Manning (1990) provided early insights into GIS's role in sustainable development, arguing that GIS is indispensable for assessing and managing natural resources in urban planning (Manning, 1990) [14]. Younis and Ahmed (2022). Application of RS and GIS to mapping groundwater potential zones (GWPZs) of Khazir River Basin, Northern Iraq. GIS has been used to run the model to obtain GWPZs map [15]. Al-Janqli and Ismaeel (2025). Spatial Modeling of Groundwater Potential Existence Using GIS: Nineveh Governorate as a Model, using GIS, to assess groundwater quality and develop spatial models using inverse distance weighting (IDW), and Kriging interpolation techniques in the southwestern desert of Iraq [18]. Alsalmany *et al.*, (2023). Determining the impact of climatic elements on vegetation cover based on RS data, satellite imagery has been used to determine some of the climatic factors affecting the variation of vegetation cover in the district of Hamdaniya [16].

Finally, An *et al.* (2024) examined the application of GIS in urban forest ecological compensation and sustainable development, showing how GIS can be applied to environmental conservation efforts in urban planning (An *et al.*, 2024) [17]. This wide range of studies collectively illustrates the critical role GIS plays in shaping sustainable urban environments, ensuring that development is both effective and environmentally responsible.

3. METHODS

This study adopts a descriptive analytical approach to investigate the effectiveness of GIS in supporting sustainable urban planning in Iraq. The research focuses on collecting and analyzing the opinions of professionals and experts working in urban planning, environmental management, and GIS across various Iraqi cities.

3.1. Study Population and Sample

The study population consists of experts in the fields of urban planning, architecture, GIS, and environmental engineering, as well as officials from relevant government institutions (such as the Ministry of Construction, Housing, Municipalities, and Public Works) and urban development agencies in Iraq.

A purposive sample of 100 experts was selected from key Iraqi cities, including Baghdad, Basra, Erbil, Mosul, and Najaf. The sample was chosen based on the participants' experience, qualifications, and involvement in projects that utilize GIS for urban development or sustainability purposes.

3.2. Data Collection Tools

Data were collected using a structured questionnaire developed by the researcher, consisting of both closed-ended and Likert-scale questions. The questionnaire was designed to assess:

1. Experts' knowledge and use of GIS
2. Perceived effectiveness of GIS in urban planning decisions
3. GIS's role in promoting sustainable development goals (SDGs)
4. Challenges facing GIS implementation in Iraq.

The questionnaire was validated through expert review and a pilot test involving 10 participants, who were excluded from the final sample.

3.3. Data Analysis

Quantitative data obtained from the questionnaires were analyzed using the statistical package for the social sciences. Descriptive statistics (mean, standard deviation [SD]) were used to summarize responses, while inferential statistics, such as *t*-tests and analysis of variance (ANOVA), were used to determine differences in opinions based on demographic variables such as years of experience, place of work, and educational background.

4. RESULTS

This chapter presents the results of the study based on the responses of the 100 experts who participated in the questionnaire. The data are organized according to the

four main dimensions of the study. Each table displays the distribution of responses for five statements, as well as the mean and SD for each item.

The results in Table 1 indicate a generally high level of knowledge and familiarity with GIS among the respondents. The first statement, *"I have a good understanding of GIS principles,"* received strong agreement from a majority of participants, with 45% strongly agreeing and 35% agreeing, resulting in a high mean of 4.15. This suggests that most experts consider themselves well-informed about the fundamentals of GIS.

The second statement, *"I have used GIS in one or more projects,"* also showed a high level of practical engagement, with 70% of respondents either strongly agreeing or agreeing. However, a small percentage (15%) remained neutral, while 15% disagreed to varying extents. The mean score of 3.90 supports the conclusion that a significant number of participants have hands-on experience using GIS in real projects.

For the third statement, *"I keep up with the latest developments in GIS,"* the data reflect a continued professional interest, as 73% of the respondents agreed or strongly agreed. The relatively high mean of 3.96 and a low SD indicate consensus and awareness of advancements in the field.

The fourth item, *"I have received specialized training in GIS,"* showed slightly more variation in responses. While 62% agreed or strongly agreed, a notable 18% disagreed or strongly disagreed, and 20% remained neutral. This resulted in a lower mean of 3.68, pointing to a moderate level of formal training among participants.

Finally, the fifth statement, *"I can independently operate GIS software,"* yielded a mean of 3.73, reflecting moderate self-reported competence. The majority of respondents (64%) agreed or strongly agreed, while a small portion (14%) expressed disagreement, and 22% chose neutral. This suggests that while many participants are confident in their technical ability, others may rely on team-based or assisted use of GIS software.

This resulted in a lower mean of 3.68, pointing to a potential gap in formal or structured training among some experts, despite their general familiarity with GIS concepts.

The fifth statement, *"I can independently operate GIS software,"* received a somewhat balanced response, with 28% strongly agreeing and 36% agreeing, totaling 64% expressing confidence in their operational skills. Meanwhile, 22% remained neutral, and 14% disagreed to some extent. The mean score of 3.73 indicates a moderate to high level of self-reported technical competency, although the responses also reflect a variation in individual proficiency levels.

In summary, the findings from this dimension suggest that while the majority of the sample possesses substantial knowledge and familiarity with GIS, there is some variation in the depth of experience, particularly in terms of formal training and hands-on technical skills. This highlights the need for continuous professional development opportunities in GIS, especially in contexts where urban planning requires advanced technological integration.

The responses to this dimension reflect a strong belief among experts in the effectiveness of GIS as a tool in urban planning. For the first statement, *"GIS contributes significantly to improving urban planning processes,"* a combined 82% of respondents agreed or strongly agreed, with a high mean of 4.20, indicating widespread recognition of GIS's strategic importance.

The second statement, *"GIS helps in identifying and analyzing urban expansion patterns,"* also received substantial support, with 78% agreement and a mean of 4.12. This suggests that most professionals see GIS as essential in visualizing and managing urban growth.

TABLE 1: Knowledge of GIS usage

Statement	Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Mean	Standard deviation
1. I have a good understanding of GIS principles.	45	35	12	6	2	4.15	0.89
2. I have used GIS in one or more projects.	40	30	15	10	5	3.90	1.10
3. I keep up with the latest developments in GIS.	35	38	18	6	3	3.96	0.94
4. I have received specialized training in GIS.	32	30	20	10	8	3.68	1.17
5. I can independently operate GIS software.	28	36	22	9	5	3.73	1.03

GIS: Geographic information systems

The third statement, *"GIS improves decision-making in infrastructure development,"* garnered 76% agreement and a mean of 4.05, reinforcing the perception of GIS as a decision-support tool, particularly in long-term infrastructure planning.

The fourth item, *"GIS enhances coordination between different planning departments,"* had slightly lower agreement (68%), with a mean of 3.85. This indicates recognition of GIS's integrative role, though perhaps with some limitations in practical inter-agency collaboration.

The fifth statement, *"GIS data is always accurate and up-to-date,"* received more mixed responses, with only 45% in agreement and a higher neutral rate (30%) and some disagreement (25%), yielding a lower mean of 3.50. This points to concerns about data quality and update cycles in local applications of GIS.

In summary, experts largely view GIS as a valuable and effective tool for urban planning in Iraq, particularly in enhancing analysis and decision-making. However, there are concerns regarding the accuracy and timeliness of GIS data, which may affect the system's full potential in practice.

The third dimension explores perceptions of GIS in promoting sustainability. The first statement, *"GIS supports the integration of environmental considerations in planning,"* received 80% agreement and a mean of 4.18, indicating strong recognition of GIS's role in environmental planning.

For the second statement, *"GIS helps in monitoring urban pollution and environmental risks,"* 75% agreed or strongly agreed, resulting in a mean of 4.05. This confirms that many experts view GIS as a monitoring tool for sustainability indicators.

The third item, *"GIS facilitates equitable access to urban services,"* had 66% agreement, with a notable neutral response (22%) and a mean of 3.80, suggesting moderate confidence in GIS's role in promoting equity.

The fourth statement, *"GIS supports long-term sustainable planning strategies,"* was widely supported (78%) with a mean of 4.10, reflecting the tool's strategic relevance in sustainable urban design.

However, the fifth item, *"GIS alone is sufficient to ensure sustainable urban development,"* saw divided responses, with only 40% agreement and a significant 30% neutral and 30% disagreement, resulting in the lowest mean in this dimension (3.45). This shows that experts view GIS as a supportive, but not standalone, tool in achieving sustainability goals.

In general, the responses suggest that while GIS is widely seen as contributing positively to sustainable development, it must be integrated with other policies, tools, and governance strategies to achieve comprehensive sustainability outcomes [Table 3].

This dimension addresses perceived barriers. The first statement, *"There is a shortage of qualified personnel to operate GIS systems,"* received 70% agreement and a mean of 3.95, indicating a recognized skills gap in the local context.

The second item, *"High costs of GIS software and infrastructure hinder its adoption,"* received the highest agreement in this dimension (82%) and a mean of 4.25, showing cost as a major barrier.

For the third statement, *"There is limited awareness among decision-makers about GIS capabilities,"* 74% of participants agreed or strongly agreed, with a mean of 4.05, suggesting that leadership awareness remains a key challenge.

The fourth item, *"Data availability and integration across departments is poor,"* had 68% agreement and a mean of 3.85, pointing to infrastructural and institutional challenges in data sharing.

The fifth statement, *"There is insufficient governmental support for expanding GIS use,"* received 65% agreement, with 20% neutral and a mean of 3.78, reflecting some perceived lack of policy-level commitment.

Overall, this dimension reveals several barriers facing GIS implementation in Iraq, particularly related to financial constraints, institutional capacity, and inter-agency collaboration. Addressing these challenges is crucial to enhancing the strategic application of GIS in sustainable urban development.

4.1. Hypothesis 1

There is a significant difference in the effectiveness of GIS in urban planning based on the level of expertise of the respondents.

4.1.1. Analysis of variance

To test this hypothesis, we would conduct an analysis of variance to determine if there are significant differences in GIS effectiveness scores based on the level of expertise. The

respondents can be grouped into different levels: novice, intermediate, and expert.

The analysis of variance test reveals a significant difference in GIS effectiveness scores between different levels of expertise ($P = 0.012$). Experts reported higher effectiveness of GIS (mean score of 4.30), while novices reported the lowest (mean score of 3.65). This confirms that the effectiveness of GIS in urban planning is influenced by the level of expertise of the respondents.

4.2. Hypothesis 2

There is a significant difference in GIS knowledge between respondents with formal GIS training and those without.

4.2.1. T-test analysis

To test this hypothesis, we will conduct an independent *t*-test comparing the GIS knowledge mean scores between two groups: respondents who have received formal GIS training and those who have not.

The *t*-test results show a statistically significant difference between the two groups (t-value = 2.65, $P = 0.024$). Respondents who received formal GIS training reported a higher level of knowledge (mean score = 4.20) compared to those without training (mean score = 3.85), supporting the hypothesis that formal training enhances GIS knowledge.

4.3. Hypothesis 3

There is a significant positive correlation between GIS knowledge and the perceived effectiveness of GIS in urban planning.

4.3.1. Correlation analysis

We will test this hypothesis using Pearson's correlation coefficient to assess the relationship between respondents' GIS knowledge and their perceived effectiveness of GIS in urban planning.

The Pearson correlation coefficient of 0.78 indicates a strong positive relationship between GIS knowledge and the perceived effectiveness of GIS in urban planning. The $P = 0.005$ confirms that the correlation is statistically significant. This suggests that experts who possess higher GIS knowledge tend to perceive GIS as more effective in urban planning.

5. RESULTS AND DISCUSSION

This section presents and interprets the results of the study using statistical evidence derived from expert responses. Each dimension is discussed in relation to the corresponding statistical values to highlight the strength of the conclusions.

5.1. Knowledge and Use of GIS

As shown in Table 1, respondents exhibited a high level of knowledge of GIS. The highest mean (M = 4.15, SD = 0.89) was recorded for the statement "I have a good understanding of GIS principles," with 80% either agreeing or strongly agreeing. Similarly, the statement "I have used GIS in one or more projects" had a strong mean of 3.90 (SD = 1.10), with 70% in agreement. These values confirm a general trend of both conceptual understanding and practical exposure to GIS.

However, the relatively lower mean for receiving specialized training (M = 3.68, SD = 1.17) indicates that while many experts are familiar with GIS, formal education in the field may still be lacking. This is statistically supported by the independent *t*-test (Table 6), which shows a significant difference in GIS knowledge between trained and untrained respondents (t = 2.65, $P = 0.024$). Participants with training had a higher mean score of 4.20 compared to 3.85 for those without, underscoring the positive impact of structured training on technical competency.

5.2. Effectiveness of GIS in Urban Planning

Table 2 illustrates a strong consensus on the practical benefits of GIS in urban planning. The statement "GIS contributes to improving planning accuracy" recorded the highest mean (M = 4.28, SD = 0.81), with 85% agreement. Other high-scoring items include "GIS facilitates better visualization of urban data" (M = 4.17, SD = 0.93) and "GIS reduces time and effort in urban planning" (M = 4.16, SD = 0.89).

A statistically significant relationship was found between GIS effectiveness and the level of expertise (Table 5). The analysis of variance results (F = 4.52, p = 0.012) confirm that experts perceived GIS as more effective (M = 4.30) than novices (M = 3.65), suggesting that familiarity and competence enhance the perceived utility of GIS in planning contexts.

5.3. Role of GIS in Achieving SDGs

Respondents strongly acknowledged GIS's contribution to sustainable development. The statement "GIS supports environmental sustainability" achieved a mean of 4.22

TABLE 2: Effectiveness of GIS in urban planning

Statement	Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Mean	Standard deviation
1. GIS contributes to improving planning accuracy.	50	35	10	3	2	4.28	0.81
2. GIS facilitates better visualization of urban data.	48	30	15	5	2	4.17	0.93
3. GIS helps in identifying optimal land use.	42	33	17	6	2	4.07	0.95
4. GIS tools are reliable for decision-making.	36	40	15	6	3	4.00	0.98
5. GIS reduces time and effort in urban planning.	44	38	10	6	2	4.16	0.89

GIS: Geographic information systems

TABLE 3: Role of GIS in achieving sustainable development goals

Statement	Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Mean	Standard deviation
1. GIS supports environmental sustainability.	46	37	12	3	2	4.22	0.84
2. GIS improves monitoring of urban expansion.	40	35	18	5	2	4.06	0.94
3. GIS assists in disaster risk management.	39	36	16	6	3	4.02	0.97
4. GIS enables more efficient resource allocation.	34	38	20	6	2	3.96	0.92
5. GIS contributes to long-term urban sustainability.	42	40	10	6	2	4.14	0.90

GIS: Geographic information systems

TABLE 4: Challenges in implementing GIS in Iraq

Statement	strongly agree	Agree	Neutral	Disagree	Strongly disagree	Mean	Standard deviation
1. Lack of GIS-trained professionals is a barrier.	44	36	12	5	3	4.13	0.93
2. Limited access to updated spatial data hinders GIS use.	38	37	15	6	4	3.99	1.00
3. Budget constraints affect GIS implementation.	40	35	14	7	4	4.00	1.01
4. There is insufficient institutional support for GIS.	36	34	18	8	4	3.90	1.03
5. Technical infrastructure is inadequate in some regions.	32	37	19	8	4	3.85	1.02

GIS: Geographic information systems

TABLE 5: Analysis of variance results for geographic information systems effectiveness based on expertise level

Expertise level	Mean score	Standard deviation	F-value	*P*-value
Novice	3.65	0.50	4.52	0.012
Intermediate	4.10	0.40		
Expert	4.30	0.38		

TABLE 6: *t*-test results for geographic information systems knowledge based on formal training

Group	Mean score	Standard deviation	t-value	*P*-value
With training	4.20	0.45	2.65	0.024
Without training	3.85	0.55		

TABLE 7: Correlation between GIS knowledge and perceived effectiveness

Variable 1 (GIS knowledge)	Variable 2 (GIS effectiveness)	Correlation coefficient (r)	*P*-value
GIS knowledge	GIS effectiveness	0.78	0.005

GIS: Geographic information systems

(SD = 0.84), while "GIS contributes to long-term urban sustainability" recorded a mean of 4.14 (SD = 0.90), both with over 80% agreement. These values affirm GIS's role in integrating environmental considerations into planning.

A strong positive correlation was observed between GIS knowledge and perceived effectiveness (r = 0.78, $P = 0.005$; Table 7), indicating that those with greater GIS understanding were more likely to recognize its contributions to sustainability.

However, the statement "GIS alone is sufficient to ensure sustainable development" received the lowest mean (M = 3.45, SD = not provided), with a near-equal split between agreement, neutrality, and disagreement. This result

emphasizes the need for integrating GIS with broader policy and planning tools to achieve holistic sustainability goals.

5.4. Challenges in GIS Implementation

Table 4 reveals the key barriers to effective GIS use in Iraq. The most agreed-upon challenge was the lack of trained professionals (M = 4.13, SD = 0.93), followed by budget constraints (M = 4.00, SD = 1.01) and limited access to updated spatial data (M = 3.99, SD = 1.00). These values reflect systemic issues that hinder the full utilization of GIS technology.

Institutional support was also found lacking (M = 3.90, SD = 1.03), while inadequate infrastructure had a mean of 3.85 (SD = 1.02). These challenges correspond to findings from Walsund (2013) and Gharineiat and Khalfan (2011), who stressed the importance of institutional capacity and data reliability in effective GIS adoption.

6. CONCLUSION

This study aimed to analyze the effectiveness of GIS in urban planning in Iraq, with a focus on experts' opinions. The results demonstrate that GIS is widely regarded as an effective tool for urban planning, particularly in enhancing spatial analysis, decision-making, and sustainability. However, challenges such as data accuracy, high costs, and a lack of inter-departmental coordination were identified as barriers to its full potential in Iraq.

GIS effectiveness was positively correlated with the level of expertise, indicating that training and professional experience are crucial for maximizing the benefits of GIS in urban planning. Furthermore, the study found significant barriers to GIS adoption, including financial constraints and a lack of governmental support, which need to be addressed to improve GIS implementation in Iraq.

6.1. Recommendations

1. Increase GIS Training Programs: Given that the effectiveness of GIS is heavily influenced by the level of expertise, there is a need to implement more extensive training programs for urban planners, engineers, and decision-makers. These programs should focus on both the technical aspects of GIS and its application in urban planning.
2. Improve Data Collection and Management: One of the key challenges identified in this study is the accuracy and timeliness of GIS data. It is essential to establish robust data collection and management systems to ensure that GIS data is up-to-date and reliable for urban planning purposes. Regular updates to data should be prioritized.
3. Policy Support and Funding: The study found that high costs and insufficient governmental support were significant barriers to GIS adoption. Therefore, the government should allocate more resources and establish policies that incentivize GIS adoption, including subsidies for GIS software and infrastructure.
4. Foster Inter-Departmental Collaboration: GIS is most effective when it is used collaboratively across different urban planning departments. Efforts should be made to enhance communication and coordination between governmental agencies to ensure that GIS is effectively integrated into all aspects of urban planning.
5. Focus on Sustainable Urban Planning: As GIS is a valuable tool for achieving sustainable urban development, it is crucial to promote its use in planning green spaces, managing environmental risks, and enhancing the accessibility of urban services. Urban planners should be encouraged to incorporate sustainability principles in their GIS-based analyses.
6. Promote Public Awareness of GIS: In addition to improving technical capabilities, public awareness and understanding of GIS's role in urban development should be promoted. Public awareness campaigns could help create a supportive environment for the widespread adoption of GIS.

In conclusion, while GIS holds considerable promise for urban planning in Iraq, addressing the identified challenges and promoting its widespread adoption are crucial steps for leveraging its full potential in fostering sustainable urban development.

REFERENCES

[1] E. W. Manning. "Geographic information systems and sustainable development". *Government Information Quarterly*, vol. 7, no. 3, pp. 329-342, 1990.

[2] P. A. Longley. "Geographical information systems: Will developments in urban remote sensing and GIS lead to 'better'urban geography?the" *Progress in Human Geography*, vol. 26, no. 2, pp. 231–239., 2002.

[3] R. Sultani, A. Soliman and K. Al-Hagla. "The use of geographic information system (GIS) based spatial decision support system (SDSS) in developing the urban planning process". *Architecture and Planning Journal*, vol. 20, pp. 97-115, 2009.

[4] M. P. Lewis and A. Ogra. "An approach of geographic information system (GIS) for good urban governance". In: *2010 18th International Conference on Geoinformatics*. IEEE, United States, p. 1-6, 2010.

[5] V. Maliene, V. Grigonis, V. Palevičius and S. Griffiths. "Geographic

information system: Old principles with new capabilities". *Urban Design International*, vol. 16, pp. 1-6, 2011.

[6] Z. Gharineiat and M. Khalfan. "Using the geographic information system (GIS) in the sustainable transportation". *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, vol. 5, pp. 1425–1431, 2011.

[7] A. Khaleel and A. K. Bety. "*Urban Geomorphology of Sulaimani City, Using Remote Sensing and GIS Techniques, Kurdistan Region, Iraq, [Thesis]*". 2013.

[8] E. Walsund. "*Geographical Information Systems as a Tool in Sustainable Urban Development*". Malmö University Publications, Malmö, 2013.

[9] S. Xhafa and A. Kosovrasti. "Geographic information systems (GIS) in urban planning". *European Journal of Interdisciplinary Studies*, vol. 1, no. 1, pp. 74-81, 2015.

[10] M. H. Zuidgeest, M. J. Brussel and M. F. Van Maarseveen. "GIS for sustainable urban transport". *ISPRS International Journal of Geo-Information*, vol. 4, no. 4, p. 2583-2585, 2015.

[11] N. Cabrera-Jara, D. Orellana and M. A. Hermida. "Assessing sustainable urban densification using geographic information systems". *International Journal of Sustainable Building Technology and Urban Development*, vol. 8, pp. 237–243., 2017.

[12] K. Kurowska, R. Marks-Bielska, S. Bielski, A. Aleknavičius and C. Kowalczyk. "Geographic information systems and the sustainable development of rural areas". *Land*, vol. 10, no. 1, p. 6, 2020.

[13] H. Ouchra, A. Belangour and A. Erraissi. "A comprehensive study of using remote sensing and geographical information systems for urban planning". *Internetworking Indonesia Journal*, vol. 14, no. 1, pp. 15-20, 2022.

[14] N. H. K. Linh, P. G. Tung, H. V. Chuong, N. B. Ngoc and T. T. Phuong. "The application of geographical information systems and the analytic hierarchy process in selecting sustainable areas for urban green spaces: A case study in Hue City, Vietnam". *Climate*, vol. 10, no. 6, p. 82, 2022.

[15] J. Younis and C. I. Ahmed. "Application of remote sensing and GIS to mapping groundwater potential zones of Khazir River Basin, Northern Iraq". *Iraqi National Journal of Earth Science*, vol. 22, no. 2, pp. 209-230, 2022.

[16] M. Alsalmany, R. Thannoun and A. Shehab. "Determining the impact of climatic elements on vegetation cover based on remote sensing data - (Al-Hamdaniya district as a case study)". *Iraqi National Journal of Earth Science*, vol. 23, no. 1, pp. 13-25, 2023.

[17] L. An, G. Liu and M. Hou. "The application of geographic information system in urban forest ecological compensation and sustainable development evaluation". *Forests*, vol. 15, no. 2, p. 285, 2024.

[18] S. Elias Al-Janqli and O. Ismaeel. "Spatial modeling of groundwater potential existence using GIS: Nineveh governorate as a model". *Iraqi National Journal of Earth Science*, vol. 25, no. 2, pp. 38-56, 2025.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Climatic Impacts on Drug Therapy Usage: A Comparative Study of Kurdish Populations in Sulaimani, Iraq, and Stockholm, Sweden

Shena Ali Faris[1,2]*, Pakiza Aziz Saied[1]

[1]Department of Geography, College of Humanities, University of Sulaimani, Sulaymaniyah, Kurdistan Region, Iraq, [2]Ministry of Housing and Construction, Kurdistan Regional Government, General Directorate of Roads, Reconstruction, and Housing Sulaimani, Sulaimani, Iraq

ABSTRACT

The geographical area is influenced by climate impacts, which, in turn, affect the use of different drug therapies during seasonal weather fluctuations. Thus, this study investigates how geographical climate differences influence drug therapy usage by comparing two Kurdish populations residing in Sulaimani, Iraq, and Stockholm, Sweden. It also highlights significant variations in healthcare practices, demonstrating how environmental conditions shape medication patterns. Data collection was conducted through a structured online survey, covering sociodemographic factors, health behaviors, and medication practices, followed by statistical analysis using Python and SPSS. Geographic Information System (GIS) tools were applied to spatially analyze environmental variables across the two cities, enabling the validation of sampling locations and the statistical determination of optimal limitations for the sample collection dataset. In Stockholm, 73.33% of respondents reported that the cold and humid climate affected their health behavior, whereas in Sulaimani, 50.27% described the climate as moderate but highly variable. The study revealed that the key statistical values such as antibiotic usage were significantly higher in Sulaimani (38.03%) than Stockholm (14.00%, $P < 0.001$), indicating a more treatment-focused approach in Sulaimani versus a preventive focus in Stockholm. Similarly, painkiller usage was significantly higher in Sulaimani, correlating with climate-related seasonal illnesses. Meanwhile, multivitamin usage in Stockholm reached 44.67%, surpassing Sulaimani's 37.77%, reflecting a stronger emphasis on preventive healthcare strategies in colder climates. These findings emphasize that climate, more than cultural differences, significantly influences drug therapy patterns. The study determines that healthcare strategies should integrate climate variability, prioritizing preventive care in colder climates and infection control in warmer regions. Finally, the study concludes with key findings and outlines directions for future research, emphasizing the need for further investigation into climate-adaptive healthcare approaches.

Index Terms: Climate Influence, Drug Therapy, Geographical Area, Medication Usage, Seasonal Weather

1. INTRODUCTION

This geographical area plays a critical role in shaping various aspects of life. For example, climate conditions significantly influence physical and psychological well-being, affecting lifestyle choices [1]–[3]. In this case, the impact of climate on physical health and emotional well-being has been examined in relation to the environments where residents live, taking into account individuals from similar and diverse nationalities and cultural backgrounds [4], [5]. The extensive narrative reviews of current evidence have been conducted, and their implications have been discussed based on these findings.

Moreover, drug therapy usage is one of the modifiable factors influenced by climate and environmental conditions

Access this article online

DOI: 10.21928/uhdjst.v9n2y2025.pp39-60

E-ISSN: 2521-4217
P-ISSN: 2521-4209

Copyright © 2025 Fars and Saied. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Shena Ali Faris, Department of Geography, College of Humanities, University of Sulaimani, Sulaymaniyah, Kurdistan Region, Iraq. E-mail: Shenaali8@gmail.com

Received: 22-05-2025 Accepted: 19-07-2025 Published: 21-08-2025

across different regions or areas [6]. Climate impacts, such as temperature, humidity, and seasonal transitions, can significantly affect the pharmaceutical drugs of various medications, thereby altering their effectiveness [7], [8]. In addition, sociodemographic factors such as age, income, education level, and access to healthcare, along with cultural beliefs and practices, have a major impact on drug therapy usage during periods of illness [9], [10]. This impact becomes particularly evident during outbreaks of infectious diseases or periods of extreme or rapidly changing weather, especially at the beginning and end of limited seasonal periods. Understanding these complex interactions is essential for improving therapeutic outcomes and creating treatment strategies that are sensitive to both cultural and environmental contexts.

More important than that, a major point is the potential to make comparisons between residents of two or more different regions to illustrate the effects of climate on drug therapy usage in general. Such comparisons can help explain how residents adapt using protective measures, such as vaccines, multivitamins, or other supplements, and how climate conditions influence the use of antibiotics and painkillers in specific areas around the world. Furthermore, sociodemographic factors, including education and access to healthcare, play a significant role in shaping patterns of drug therapy usage [11]. Conducting geographically based investigations is crucial for identifying areas that require focused attention in the future and for optimizing strategies to address the impacts of climate on health outcomes. This approach is especially important in establishing a stronger understanding of the relationship between healthcare practices and geographical effects [12], ultimately contributing to better management and preparation for future climate-related health challenges.

For the purpose of this study, a comparison and discussion were conducted between two important locations: one situated in the Middle East and the other in Europe. This approach aimed to isolate the effects of geography and climate on fluctuations in drug therapy usage. Importantly, the sample focused exclusively on individuals of a single nationality residing in both regions, allowing for a clearer understanding of how environmental differences, rather than cultural or ethnic factors, influenced patterns of medication use. In this investigation, Kurdish populations were chosen to control for ethnicity/culture, enabling isolation of climate-geography effects, including those residing in Sulaimani city in Iraq and Kurdish refugees or residents living in Stockholm, Sweden. The data collection was dedicated to specific queries designed to evaluate how demographic changes, influenced by climate differences, impact healthcare practices. Particular attention was given to analyzing drug therapy usage during periods of seasonal weather fluctuations. In addition, this study focused on different cases based on previous state-of-the-art research and evaluated the data statistically to develop potential solutions and identify limitations encountered in healthcare, particularly in relation to medical therapy within the selected areas. Special attention was given to how changes in climate affect healthcare practices among individuals from Kurdish cultural backgrounds when relocating or living in different environmental conditions.

In response to this issue, the contributions of this research are outlined in the following points: (1) it offers the first comparative analysis of medication use among Kurdish populations living in distinct climatic regions – Sulaimani and Stockholm – revealing how demographic differences influence health behaviors and drug usage patterns. (2) It effectively integrates Geographic Information Systems (GISs) with statistical methods to map environmental conditions and isolate the impact of climate from cultural and geographical influences on medical practices. (3) The findings show that seasonal changes, particularly during autumn and winter, significantly affect illness rates and the consumption of antibiotics, vaccines, and painkillers, with notable differences observed between the two cities. (4) Finally, the study introduces a climate-adaptive healthcare perspective, emphasizing the importance of region-specific public health strategies, improved awareness, and tailored medication approaches in response to varying climate conditions.

The remainder of the paper, including the next section, provides a background review to enhance the state of the art related to the problem at hand. The third section details the methodology and materials, outlining the steps of the method in a step-by-step manner. The fourth section includes a discussion and comparison of the result analysis. The final section is dedicated to the conclusion and suggestions for future research.

2. BACKGROUND REVIEW

Medical geography is an old perspective and a new specialty. Today, much research in the social and behavioral sciences is shaped by its principles, which also influence studies in other fields [13]. This science determines the distribution of human diseases. It also examines the circumstances in which diseases

arise. Additionally, it explores how natural environmental factors affect human health. This is especially important following the population explosion known as population rapidity [14]. Studies have described how the environment and seasons affect health, urging practitioners to consider factors such as sun exposure, soil, altitude, climate, and geography [15], [16]. During the 16th–18th centuries, interest in climate and health effects was facilitated by the ability to measure environmental conditions with new instruments [17]. There is a clear gap in defining the relationship between city structure and efforts to reduce social inequalities, particularly through facilities tailored to specific needs during disease outbreaks [18]. The influence of these factors on the creation of a particular type of climate in the regions is evident. Therefore, one of the most important issues affected by climate is public health. Previous studies on this topic have emphasized the relationship between disease prevalence and climate in several cities [19]–[21].

Furthermore, through ecological and biological processes such as altered food supply and disease transmission, climate change may have indirect effects on health. Thus, in the study by Patz *et al.* [22], the impact of climate change on infectious diseases was further explored, including investigations [23] focused on North America [24] and Europe [25]. These important factors of temperature and seasonal variation have influenced drug use. After looking for related research, given the number of studies concerning how climate affects various illnesses, which results in the usage of various medications and therapies. All indices have improved that climatic factors have a large and different impact on disease prevalence and drug use in different parts of the world [15], [26], [27]. Table 1 provides a summary of various studies, outlining the different methods and research objectives used to investigate the selected problem. The inclusion of 2024 studies ensures the representation of the most recent and cutting-edge literature, despite their close temporal proximity to the data collection period.

Nevertheless, the findings of various studies partly correspond with each other. The influence of climate on

TABLE 1: Several studies are exploring the impact of climate on disease progression and medication usage

References	Purpose	Methodology	Key finding*	Year
Hartig *et al*. [28]	Examines if limited access to relaxing landscapes due to bad weather harms mental health and increases medication use.	Analyzes self-reported mental health, outdoor activity, medication use, and weather data.	Cold Swedish winters worsen mental health and boost antidepressant use, especially among women and urban residents.	2007
Talib and Ahmed [29]	Analyzes the link between seasonal climate change and respiratory illnesses in Sulaimani province.	A description and quantitative approach	Health climate features of Sulaimani Province and their impact on respiratory diseases.	2019
Astolphi *et al*. [30]	To examine the effects of weather and climate change on the health and well-being of elderly dementia patients. Recent 2024 studies were included to reflect the latest advancements during data collection.	A scoping review will use PRISMA-ScR and Arksey & O'Malley's framework to examine climate change effects on older adults with dementia, based on selected database findings.	There is little proof to link dementia-related health hazards to climate change; further study and action are required.	2024
Basu and Samet [31]	The study links heat-related deaths to urban heat islands and aging populations, stressing the need for predictive models to guide prevention.	The research methodology includes analyzing exposure and results, choosing a suitable study design, and analyzing the temperature-mortality link with statistical models.	Key findings focus on heat-related mortality, including risk factors like cardiovascular and respiratory conditions, hospital studies, and other temperature-related deaths.	2002
Eng and Mercer [32]	Studied the link between two climate conditions and mortality in two populated areas.	Analyzing metrological data	The effect of seasonal changes varies greatly among nations where extra winter mortality has been reported.	2016
Amuakwa-Mensah *et al*. [33]	The Relationship between Infectious Diseases and Climate Variability	Collecting data and applying statistical analysis	Draws attention to the need for personalized programs for public health.	2017

*While previous studies establish links between climate and disease, as referenced in the table, none have examined medication practices within Kurdish diaspora communities, highlighting the need for this study

disease prevalence and subsequent drug therapy use is widely acknowledged in the scientific literature. Although most studies agree on the existence of this effect, the types of diseases influenced by climate vary based on the specific climatic conditions of each region. For example, Dawson *et al.* in 2023 examined the effects of heat and cold exposure on chest pain symptoms [34]. In contrast, a study conducted in Cairo analyzed the impact of climate on human diseases, categorizing them according to seasonal variations [35]. In addition, the effects of extreme weather events on disease spread, such as those observed during Hurricane Katrina in 2005 and the major floods in Pakistan in 2010, have been well documented [36].

In addition to these studies, several other studies have investigated how climate affects therapeutic and disease utilization, focusing on environmental factors in different settings, beyond just climate. The World Health Organization (WHO) reports that there were notable cholera outbreaks in the Kurdistan Region in 2007, 2008, and 2012. A significant outbreak caused by tainted water occurred throughout Iraq in 2007 alone, sickening almost 7000 individuals and leading to 10 recorded fatalities [37].

Neira *et al.* emphasize the impact of environmental factors, including climate, on the development of respiratory diseases in Egypt, identifying both natural and human-made pollution as significant contributors [38]. In 2011, Pfister and Krämer investigated the consequences of climate change, environmental factors, and disease transmission, citing both historical and modern examples [39]. In addition, the effect of climate on diseases can vary among age groups. For example, a study conducted in Melbourne, Australia, shows how climate influences the development of chest pain in young people [21]. Another study explores how high temperatures impact the older population in Europe, particularly regarding medication use and morbidity rates [40]. Investigations emphasize the importance of this issue and its relevance for further research, particularly as climate change is associated with increased chronic disease and, in some cases, higher mortality [41].

Despite these challenges, various locations were selected as research areas to better understand and address the issue, particularly in the context of global development. The study focuses on how to manage and resolve the problem in relation to drug therapy usage, including antibiotics, multivitamins, and preventive vitamin use before the pandemic, within the broader context of climate change. The selected research areas include the city center of Sulaimani and specific districts within Stockholm. In Sulaimani, seasonal changes, prolonged summers, and high temperatures lead to respiratory diseases and seasonal epidemics, increasing the use of painkillers and anti-inflammatory drugs [29]. Furthermore, in Stockholm, the extended cold season and lower temperatures contribute to the spread of influenza, vitamin D deficiency, mental health disorders, and, in some cases, chronic conditions such as heart disease [42]–[44]. These factors are associated with an increased reliance on supplements, painkillers, and psychotropic medications [45]. Thus, studies emphasize the urgent need to adapt healthcare strategies to local climate conditions and to promptly distinguish pandemic-related infections. This approach can improve medication use, enable early identification of emerging diseases, enhance public awareness, and help prevent the spread of illness [46]–[48].

Finally, the literature highlights the long-standing importance of medical geography in linking environmental conditions to human health, including disease distribution and prevalence. Although previous studies emphasized how climate, geography, and seasonal changes affect health, they did not provide direct comparative analyses between specific cities. Our study fills this gap by uniquely comparing drug and antibiotic usage patterns between Sulaimani and Stockholm, revealing how climate differences influence public health behaviors in distinct urban contexts.

3. MATERIALS AND METHODS

In geographical studies, field research is essential because it provides micro-level insights into particular places, which help with larger planning and problem-solving initiatives for development [49]. As outlined in the initial section, the primary objective of this study is to examine the impact of climate change on the use of pharmacological therapies across different geographic contexts. The analysis focuses on how climate change influences public health, particularly through changes in the demand for and types of medications used to treat climate-sensitive illnesses.

Sulaimani, Iraq, and Stockholm, Sweden, are selected as contrasting case study locations due to their differing climates and healthcare systems. In addition, the study focuses on Kurdish residents living in Sulaimani and Kurdish individuals residing in Sweden as refugees or permanent residents. All locations were mapped based on the selected study area, with their identification and design process thoroughly discussed in the corresponding subsection on the study area. This mapping was conducted following specific selection criteria

to ensure accuracy and relevance. The comparative method employed, supported by statistical analysis, enables an in-depth examination of how regional environmental factors influence health outcomes and the demand for medication. In addition, all the steps involved in identifying the gap in the problem, as well as the change of the methodological line used to find a solution, are widely illustrated in Fig. 1. This illustration provides a clear and detailed description of the logical progression from problem identification to analytical method selection.

3.1. Problem Approaches Selection

This process involves several critical steps aimed at identifying the most effective approach for addressing the situation under study, following a systematic, step-by-step method. Due to the geographical complexity, it is challenging to directly determine how drug usage patterns vary across different locations. Therefore, this research adopts a scientific approach that relies on healthcare data to clarify the methodology. The analysis is closely linked to geographical variations, climate change, and seasonal weather fluctuations, particularly in the context of pandemics or the spread of various diseases. These approaches are outlined as follows:

- Principal approach: The focus is on describing the climatic conditions that influence disease spread and the use of therapeutic drugs in the two selected study areas. This discussion builds upon the previous sections, which addressed the impact of both natural and human factors on these regions
- Historical approach: The state-of-the-art research relevant to this study's subject, as discussed in the previous sections, highlights the latest advancements and methodologies in understanding the interplay between climatic conditions, disease spread, and therapeutic drug use
- Comparative approach: The study focuses on how different climate conditions relate to the occurrence of diseases and how they affect the use of therapeutic medications in the various study areas. By analyzing their growth and seasonal variability based on data collected from residents, this study delves deeper into the dynamics of these relationships. By highlighting regional variances and temporal trends, this comparative approach provides insights into how climate variations affect pharmaceutical use and health consequences
- Analytical approach: The research is based on questionnaires and email surveys can provide fast and cheap alternatives. The data analytic process begins with defining a clear objective or problem statement. Once the goal is established, a strategy for data collection is devised and implemented. Collected data often contains errors, duplicates, or irrelevant information, so cleaning it involves removing these issues, filling gaps, and organizing the dataset to ensure it's ready for analysis.

3.2. Problem Query Search Strategy

A comprehensive literature search has been conducted to explore the relationship between geographical climate variability and human health. Electronic databases such as PubMed, Scopus, Medline, Elsevier ScienceDirect, Springer Online Journals, Global Health, MDPI, and the Web of Science have been utilized for this investigation. The search has focused on publications from various years, with

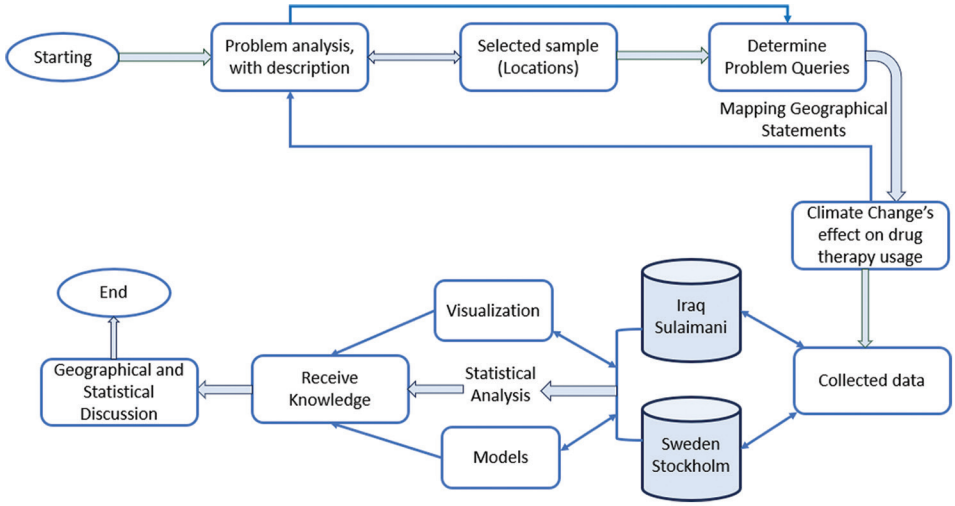


Fig. 1. Methodical flow outlining each step in the problem analysis process.

particular emphasis on the period between 2010 and 2023, using keywords such as "diseases," "climate variability," "meteorological factors," "weather," "climate and diseases," "human health," "health effects," "cardiovascular disease," "seasonal change," and "respiratory disease."

This study efforts to assess the impact of climate on health perceptions and behaviors; however, it does not include direct measures of infection rates or policy outcomes. Therefore, policy-related interpretations, such as prioritizing infection control in warmer regions, are beyond the scope of the data and should be viewed as speculative. To capture newly developed methods not identified in the initial search, we also monitored Google Scholar through daily email alerts. In addition, we reviewed reports and articles from non-biomedical sources, including publications by the World Meteorological Organization and regulatory documents from environmental and health agencies. These sources offered valuable insights into climate variables (e.g., temperature and humidity) and meteorological hazards (e.g., droughts, floods, and heatwaves).

3.3. Study Area Selection

According to the majority of geographers, despite being small, cities are important, complex, and interesting geographical areas. This is especially true for big cities, which exhibit unique urban features and the most intricate functional divisions, making it possible to identify subregions with ease [50]. Based on this approach, our investigation and results focus on a comparative study of two specific areas, which are essential for identifying the impact of geographical climate change on the use of drug therapy among Kurdish populations. The selected study areas, previously identified, are the city of Sulaimani and Stockholm. To accurately represent the study locations, GIS applications were utilized, employing Python programming in combination with the Overpass API to create detailed location maps, as illustrated in this study.

As shown in the map location in Fig. 2, Sulaimani, located in the Northeastern Kurdistan Region of Iraq, the city's coordinates are approximately longitudes 45°12'00"–45°32'00" E and latitudes 35°25'30"–35°40'00" N [51]. However, Sulaimani, known as the cultural capital, offers a stable and appealing environment for development investments. This is further supported by the district's entrepreneurial resources, which are effectively utilized in accordance with the principles of social justice. Sulaimani is renowned for its rich history, modern infrastructure, and esteemed educational institutions. The city center is situated at an elevation of approximately 870 m above sea level. Its landscape is characterized by mountainous terrain, including the Piramagroon, Azmar-Goizha, and Baranan mountains, which constitute over 20% of the city. The lower-lying areas are confined to prominent valleys, such as the Tanjero Valley, which lies at an elevation of around 675 m above sea level

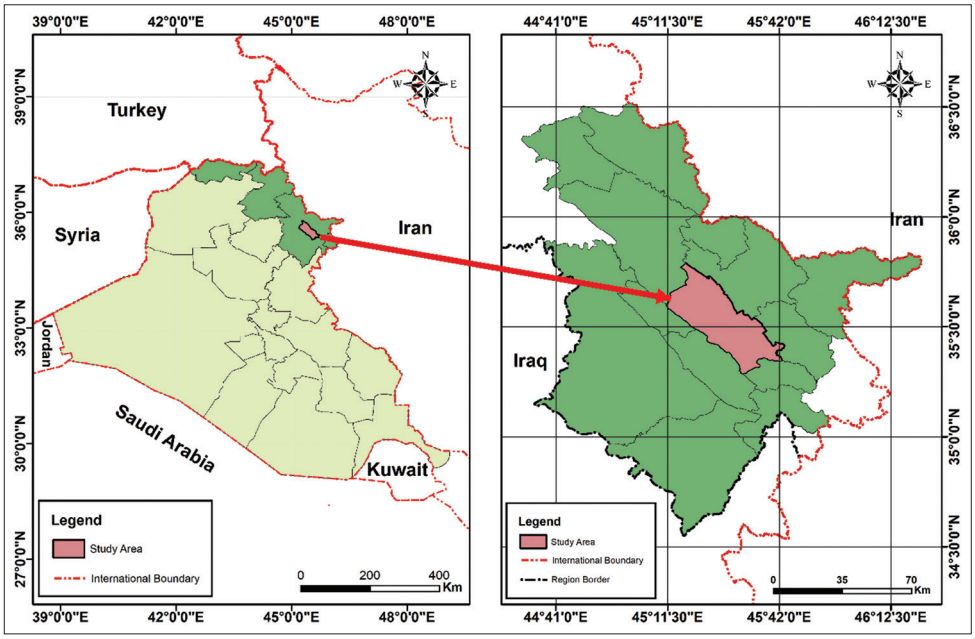


Fig. 2. Location of Sulaimani in the study area as a sample for investigation (Sulaimani: 35°25'–35°40'N, 45°12'–45°32'E.).

(m.a.s.l.). The Earth's astronomical location plays a critical role in determining the amount of solar radiation received at its surface across different seasons. This influences the length of day and night, which varies with the Earth's axial tilt and orbit around the Sun. These variations directly affect seasonal temperature patterns and contribute to fluctuations in daily, monthly, and annual temperature ranges. According to the Köppen climate classification, Sulaimani falls within the Mediterranean climate zone (Csa). This climate is characterized by hot, dry summers and mild, wet winters. The majority of rainfall occurs during the winter months, while summers are generally dry and arid [52]–[56].

On the other hand, this investigation also focuses on Stockholm, Sweden's administrative capital and economic hub, located in the southeastern part of the country, as illustrated in the location map in Fig. 3. Stockholm is one of Europe's fastest-growing cities [57], experiencing rapid development and urban expansion. It serves as a cultural and economic center, attracting people from diverse backgrounds. A significant Kurdish population resides there, contributing to the city's multicultural and dynamic community. It spans across 14 islands in the archipelago where Lake Mälaren meets the Baltic Sea. The city's coordinates are approximately 59.3293° N latitude and 18.0686° E longitude [58]. This city is known for its unique geography, blending urban areas with extensive waterways, green spaces, and historical landmarks. In addition, the city offers something for everyone, from its stunning natural beauty to its vibrant cultural scene and rich history. It falls within the humid continental climate zone (Dfb), which is characterized by cold winters, warm summers, and no distinct dry season, with precipitation distributed throughout the year [59], [60].

The topographic analysis of Sulaimani and Stockholm reveals distinct elevation characteristics. Sulaimani's mixed landscape includes plains (517–750 m), valley-hill zones (750.1–1000 m), and mountainous areas (1000.1–2210 m), leading to significant microclimatic variation that may influence health outcomes and environmental exposure. In contrast, Stockholm features a relatively flat topography, comprising plateaus (0–25 m), plains (25.1–50 m), and small hills (50.1–99 m), scattered with several water bodies. These topographic differences outline each city's air quality, humidity, and temperature, which are factors closely linked to respiratory health and patterns of related drug usage therapies. These features and characteristics are clearly illustrated in the maps shown in Fig. 4, which were generated using open-source GIS tools.

The choice of Sulaimani and Stockholm is meant to highlight the contrast between a mountainous Kurdish city and a flat Scandinavian capital. This comparison helps explore how differences in topography and climate affect seasonal drug use, especially among Kurdish communities. Sulaimani, with its high elevation and sharp seasonal temperature changes,

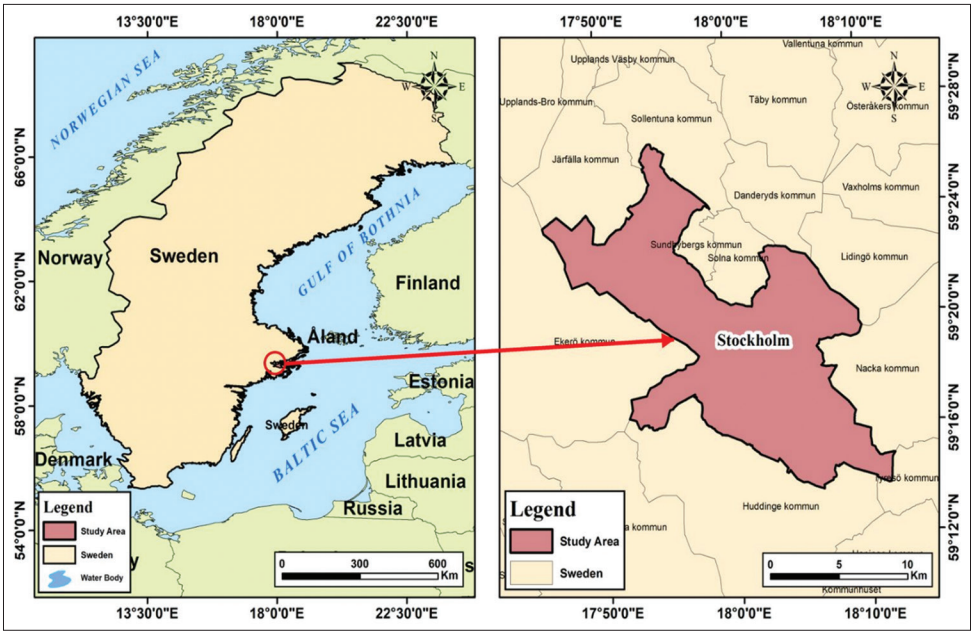


Fig. 3. Location of Stockholm in the study area as a sample for investigation (Stockholm: 59.3293°N, 18.0686°E.).

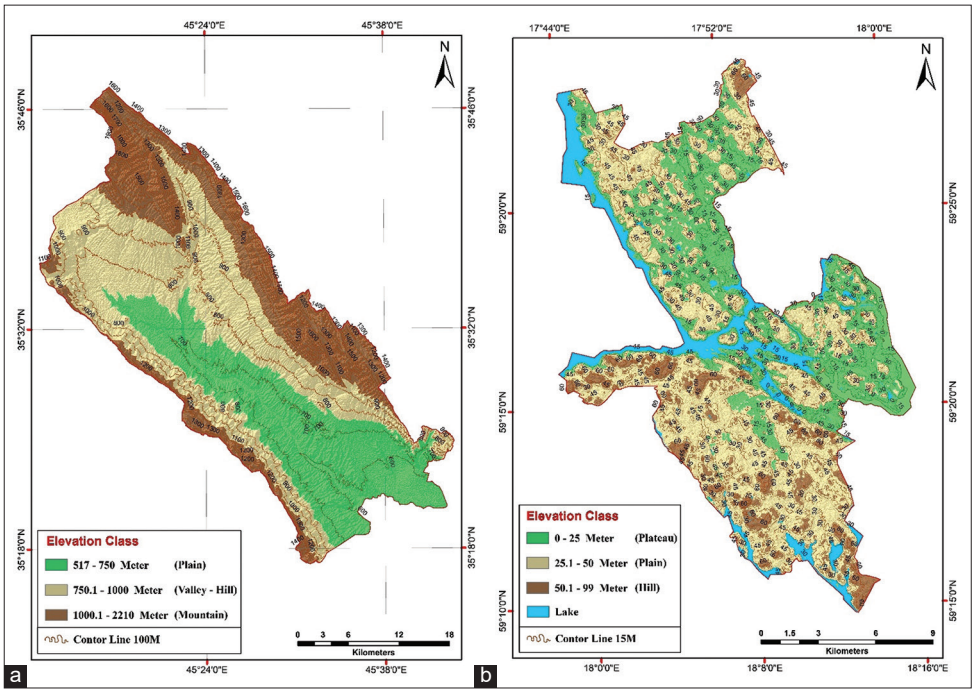


Fig. 4. Comparison of topographic layers between selected city samples, with derived explanatory features. (a) Sulaimani (35°25'–35°40'N, 45°12'–45°32'E.), (b) Stockholm (59.3293°N, 18.0686°E).

may see more respiratory infections during colder months, as reflected in the collected data. This could lead to higher use of antibiotics and other medications, as reported in the survey responses. On the other hand, Stockholm, although generally colder, has a more stable climate, well-developed infrastructure, and better healthcare access, which may contribute to more consistent and lower seasonal changes in drug use. By comparing these two cities, the study aims to better understand how geography and environment shape public health behavior, particularly among Kurdish or immigrant populations adjusting to different urban climates. All observed effects are clearly shown in the data and will be discussed in the following sections.

For a more detailed comparison between these two distinct cities, a significant contrast in temperature and humidity was analyzed using average data. The results for Sulaimani were derived from the city's meteorological station, located at latitude 35°25'–35°40'N and longitude 45°12'–45°32'E. Corresponding data for Stockholm were obtained from the Stockholm Observatory Station (lat. 59.3293°N, long. 18.0686°E), one of Sweden's oldest weather monitoring stations. The analysis focused on monthly seasonal variations over a 10-year period, from 2014 to 2024. Table 2 presents the average comparisons in temperature and humidity between the two cities.

In addition, to provide a more specific and visual understanding of the observed patterns, the data analysis focused on the relationship between seasonal time sequences, defined by specific months, and variations in temperature and humidity. This analysis aimed to improve insights into therapy-related drug usage, particularly how environmental conditions influence the timing and frequency of drug consumption. The selection of this research problem was based on its clear relevance, and the data generated from the two selected meteorological stations supports this focus.

As illustrated in Fig. 5 (Line Chart A and C), the temperature data from Sulaimani show a relatively normal distribution across the years 2014 to 2024, though the frequency patterns vary slightly from year to year. The temperature trend typically starts at its lowest in January and gradually increases, reaching a peak in June and July. However, the recorded normal temperatures in Stockholm are consistently lower than those in Sulaimani. In contrast, Fig. 5 (Line Chart B and D) presents humidity trends. In both cities, humidity generally begins high in January, decreases through the middle of the summer in June and July, and rises again toward December.

A noticeable fluctuation in Sulaimani's humidity occurred in November 2015, reflecting localized climatic instability. Overall, humidity levels in Stockholm are higher and more

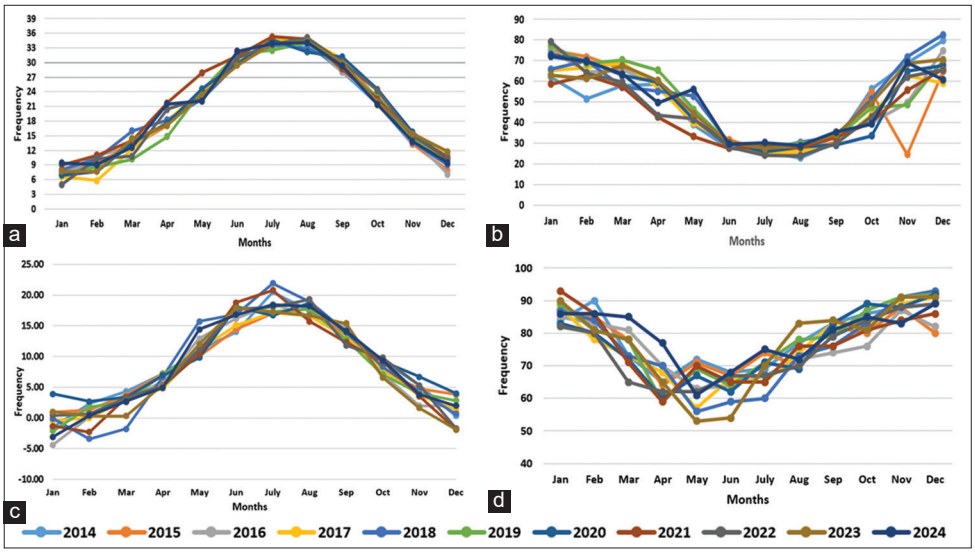


Fig. 5. Comparison of temperature and humidity frequency between Sulaimani and Stockholm from 2014 to 2024. (a) Monthly temperature frequency of Sulaimani, (b) Monthly humidity frequency of Sulaimani, (c) Monthly temperature frequency of Stockholm, (D) Monthly Humidity Frequency of Stockholm

TABLE 2: Comparison of average temperature and humidity between the cities of Sulaimani and Stockholm from 2014 to 2024

Months	Temperature		Humidity	
	Sulaimani	Stockholm	Sulaimani	Stockholm
January	7.445	−0.618	69.627	86.727
February	9.109	0.364	65.582	83.091
March	12.982	2.518	63.291	75.182
April	18.355	6.055	55.427	65.364
May	24.064	11.773	43.718	63.727
June	30.518	16.727	29.700	64.000
July	34.073	18.500	26.636	68.545
August	34.155	17.600	27.009	74.000
September	29.509	13.564	31.909	80.000
October	22.564	8.327	45.355	83.182
November	14.409	4.082	58.748	88.000
December	9.864	1.045	69.191	88.545
Average	20.587	8.328	48.849	76.697

stable compared to those in Sulaimani. This may reflect the more regulated and temperate climate of northern Europe. These seasonal movements in temperature and humidity are important climate variables that influence patterns of drug usage as therapy, particularly in relation to painkillers, antibiotics, and vaccines. The visualized patterns forecast that climate variability may serve as a useful indicator for forecasting therapeutic drug demand.

3.4. Data Collection Using Google Forms

The study examines the impact of climate on disease prevalence and drug therapy usage, focusing on public health in Sulaimani and Stockholm. Data were collected using a structured Google Form questionnaire and organized into several sections based on four specific criteria aligned with the principles of the *Simple Random Sample* model [10], [61]. The Google Form for data collection was launched in November 2024 to systematically gather responses and was completed on January 26, 2025.

Before initiating the main data collection, the survey instrument underwent pilot testing with a small group of 15 participants (10 from Sulaimani and 5 from Stockholm) to assess clarity, cultural relevance, and completion time. Based on the participants' feedback, minor linguistic and structural adjustments were made to improve item comprehension and ensure cross-cultural consistency. To evaluate internal consistency, Cronbach's alpha was calculated for multi-item sections, yielding values between 0.72 and 0.83, which reflect acceptable to good reliability. In addition, the questionnaire was reviewed and validated by three independent experts specializing in public health and climate-health dynamics, confirming the instrument's content validity and contextual appropriateness.

The first section gathered personal sociodemographic information, including age, weight, height, gender, occupation, education, and residence. These foundational data are crucial for contextualizing public health research and ensuring precise analysis of climate-related effects. The second section focused on residential area characteristics, documenting the

climate conditions of each city and their perceived impact on disease prevalence and drug use. The third section, subdivided into several parts, explored therapeutic practices for climate-related diseases. It included detailed questions about the use of vaccines, antibiotics, and paracetamol, among other treatments. The final section featured open-ended questions to capture broader perspectives and enrich the study with qualitative insights. This comprehensive approach provides a robust framework for analyzing the interplay between climate, disease, and therapeutic interventions. To further evaluate the problem, an open-ended question was provided to the resident participants, allowing them to express their perspectives and insights regarding the issue.

To address the unequal sample sizes between Stockholm (n = 150) and Sulaimani (n = 376), we acknowledge that this disparity stems from differences in accessibility and respondent availability, particularly due to the smaller Kurdish refugee population and limited digital access in Stockholm. To reduce the impact of this imbalance, we used Chi-square tests with effect size estimates (Cramér's V) to ensure that observed differences reflect meaningful associations rather than being influenced by sample size alone as discussed in the next subsection. All analyses were conducted using proportional comparisons rather than raw counts to improve interpretability. Both cohorts included Kurdish participants, which helped control for cultural differences and allowed us to focus on climatic and geographic effects. Nonetheless, we recognize that the sample size imbalance may still limit generalizability and reduce statistical power in subgroup analyses.

Participants shared the same ethnicity, but factors such as cultural adaptation, healthcare access, and socioeconomic differences were not fully controlled. Since the study focuses on climate effects, these factors are acknowledged as limitations and discussed accordingly. A specific point regarding online data collection and bias control in online surveys involves addressing sampling limitations caused by relying solely on digital methods, as noted in Section 5. This can lead to underrepresentation of groups such as the elderly or those with limited internet access, affecting the generalizability of findings. To reduce this bias, researchers have used mixed-mode data collection, applied demographic balancing techniques, increased accessibility through multiple formats or languages, and monitored response rates to correct underrepresentation and improve study validity.

3.5. Statistical Analyses

After data collection, the responses have been carefully reviewed to eliminate errors and redundancies, ensuring the accuracy and reliability of the dataset. With the cleaned dataset, statistical analysis techniques can be applied, specifically tailored to extract the desired insights and address the research objectives. After analysis, the results are presented through clear visualizations and detailed reports. Transparency is crucial – present all evidence, highlight any gaps, and avoid selective reporting to maintain credibility. Finally, the process embraces failure as an opportunity for growth. Challenges such as errors or new questions are part of the iterative nature of data analytics, requiring adaptability, creativity, and a willingness to refine the approach while staying aligned with the framework.

In addition to the statistical analysis, the sociodemographic assessment involves the calculation of Body Mass Index (BMI) [62], which is determined using the participants' weight (W) in kilograms and height (H) in meters, as shown in Equation (1).

$$BMI = \frac{W}{H^2} \quad (1)$$

The data were statistically analyzed using two software platforms. The first involved Python 3.9.7 programming language in combination with Anaconda 3.13.3 (Jupyter Notebook 7.4.0). Several Python libraries were employed, including Pandas for data analysis and manipulation and SciPy for performing Chi-square tests. The second platform utilized was IBM SPSS Statistics version 26.0.0.0, which was used to conduct cross-tabulation analyses and calculate P-values to determine statistical significance.

In numerical analysis, a categorical variable is defined as one that has a limited, typically fixed range of possible values and is used to classify data into distinct categories [63]. In this study, beyond constructing the contingency matrix, we also performed a cross-tabulation (X^2) analysis using the Chi-square test of independence. This test, as represented by the formula (2) [10], is crucial for assessing the significance of the association between categorical variables, thereby highlighting the strength and relevance of their correlation with significant values [64]. Despite variations in application, all methods are based on the same fundamental principle: outcomes are determined by comparing observed values with expected values. In this context, (i) and (j) represent a specific cell within the contingency table of the sample (N), defined by its corresponding row (r) and column (i). Here, (O) denotes the observed frequency, while (E) represents the expected frequency under the assumption of independence between the variables.

$$X^2 = \sum_{i=1}^{c} \sum_{j-1}^{r} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (2)$$

Where i denotes the row index and j denotes the column index.

4. RESULTS AND DISCUSSION

However, GIS technologies are used to map the coordinates of the research areas and visualize the spatial relationships between location, climate variables, and health indicators. Based on this mapping, certain aspects related to the impact of climate on drug therapy are finalized for the study. The collected data are then subjected to statistical analysis to identify trends and correlations between environmental factors, such as temperature fluctuations and seasonal extremes, and patterns of medication use. This analytical approach aims to assess how climate effects may influence public health practices and treatment strategies, based on data gathered from residents in the two selected study areas.

In addition, the analysis presented in this section enhances our understanding of the geographical factors influencing health and contributes to scientific efforts aimed at addressing development challenges. The findings from this research provide valuable insights to support healthcare planning and policy-making, especially in developing climate resilience and adaptation strategies. This is achieved through the integration of investigative methods, spatial analysis, and statistical evaluation. This study offers a comprehensive visualization of the complex relationships between environmental conditions and health outcomes, with a specific focus on the use of various types of therapeutic drugs in response to climate variability, weather patterns, and social factors across different geographical locations. It centers on two significant regions where Kurdish populations currently reside: Sulaimani and Stockholm.

For this purpose, to provide a more structured statistical comparison, the results were analyzed based on the impact of each modification and the validity of the dataset concerning climatic effects. To achieve this, Cramér's V, as shown in formula 3, a statistical measure used to assess the strength of association between categorical variables, was calculated to evaluate the degree of association between the outcomes from Stockholm and Sulaimani. Cramér's V values range from 0 (indicating no association) to 1 (indicating a perfect association), helping to quantify the strength of observed relationships in the context of categorical data.

$$V = \sqrt{\frac{X^2}{n(k-1)}} \quad (3)$$

Where n is the total sample size, and k is the smaller value between the number of rows and the number of columns in the contingency table.

All results were derived from a specific sample, collected randomly according to the principles of the *Simple Random Sample* model. Data collection took place in two selected locations, including 376 individuals from Sulaimani and 150 individuals from Stockholm, representing various demographic groups. This approach was intended to predict outcomes and finalize conclusions related to the research problem. The collected data were then illustrated and visualized to better understand and address key aspects of the issue, particularly those that initially appeared insignificant.

4.1. Effects of Sociodemographic Variables

Geographically, one of the key features of this study is the mathematical and statistical analysis of sociodemographic variables to identify significant aspects related to the selected demographic locations. As outlined in the methodological section, the sociodemographic variables considered include gender, age groups, BMI, education level, and employment status. To improve the accuracy of data visualization, BMI values were classified into four range levels based on the arithmetic results.

Every variable was divided into distinct groups, as Table 3 illustrates. Age was separated into five groups, for example: 18–25, 26–35, 36–45, 46–55, and 56 and over. Age distribution differed significantly between the two populations ($\chi^2 = 38.9$, $p < 0.001$). In both Sulaimani and Stockholm, the 36–45 age group represented the largest share, with 35.64% and 40.67%, respectively. Younger participants 18–25 were much fewer in Stockholm, 8% compared to Sulaimani, 25.8%. Overall, the majority of participants in both cities were under 40 years old, with Stockholm showing a stronger concentration in this age range. In addition, female participation was higher in both locations, particularly in Sulaimani. In addition, education levels differed significantly between the two groups ($\chi^2 = 84.6$, $P < 0.001$), with bachelor's degrees most common in Sulaimani and high school or diploma levels more frequent in Stockholm. Employment status also showed a significant

TABLE 3: Statistical and arithmetic comparison of sociodemographic features

Features	Categories	Iraq-Sulaimani (376)		Sweden-Stockholm (150)		Statistics	
		Count	Percentage	Count	Percentage	χ^2	*P*-value
Gender	Male	172	45.74	80	53.33	2.47E+00	1.16E-01
	Female	204	54.26	70	46.67		
Age Group	18–25	97	25.80	12	8.00	3.89E+01	7.42E-08
	26–35	105	27.93	35	23.33		
	36–45	134	35.64	61	40.67		
	46–55	30	7.98	30	20.00		
	56 or above	10	2.66	12	8.00		
BMI*	Low range	55	14.63	13	8.67	3.83E+00	2.80E-01
	Lower mid-range	270	71.81	112	74.67		
	Upper mid-range	44	11.70	22	14.67		
	High range	7	1.86	3	2.00		
Education Degree	Basic	3	0.80	7	4.67	8.46E+01	9.06E-17
	High school	35	9.31	48	32.00		
	Diploma	51	13.56	40	26.67		
	Bachelor's degree	216	57.45	40	26.67		
	Master's degree	42	11.17	14	9.33		
	Doctorate	29	7.71	1	0.67		
Employment Status	Student	94	25.00	21	14.00	6.12E+01	1.66E-12
	Employee	204	54.26	53	35.33		
	Entrepreneur (Businessman)	46	12.23	63	42.00		
	Unemployed	29	7.71	10	6.67		
	Retired	3	0.80	3	2.00		

*BMI categories were defined as follows: Low range (BMI≤22), Lower mid-range (BMI >22 and ≤30), upper mid-range (BMI >30 and ≤37), and high range (BMI >37)

difference ($\chi^2 = 61.2$, $P < 0.001$), with more employees in Sulaimani and more entrepreneurs in Stockholm.

According to Table 3, a Chi-square analysis was conducted to compare the sociodemographic variables between groups. Gender distribution did not significantly differ ($\chi^2 = 2.47$, $P = 0.116$, Cramér's V = 0.08), indicating a small and non-significant effect. Age group differences were statistically significant ($\chi^2 = 38.9$, $P < 0.001$, Cramér's V = 0.29), showing a moderate effect size. BMI categories also showed no significant variation ($\chi^2 = 3.83$, $P = 0.280$, Cramér's V = 0.07, small effect). In contrast, education levels varied markedly between groups ($\chi^2 = 84.6$, $P < 0.001$, Cramér's V = 0.43), reflecting a large effect size. Similarly, employment status was significantly different ($\chi^2 = 61.2$, $P < 0.001$, Cramér's V = 0.38), indicating a moderate to large effect.

4.2. Impact of Demographic Factors on Weather-related Descriptions

Several key inquiries and impacts were identified related to how individuals living in specific locations are influenced, particularly regarding the use of drug therapy and healthcare behaviors within a sociological context. Although there are some differences between cases, similar impacts can emerge globally, especially during widespread events like pandemics. Thus, geography, linked with demographic factors, can significantly affect lifestyle changes and cultural shifts. To explore these influences, the study gathered targeted responses from participants based on two specific questions, as presented in Table 4.

The weather description is shown in the second main part. The study emphasizes how the environment and weather have a major influence on public health in Stockholm and Sulaimani. The majority of respondents in Stockholm classified the weather as cold and humid, 73.33%, but the Sulaimani varied climate, 50.27% revealed a wider range of opinions as a median. Autumn and winter saw a rise in illnesses, especially seasonal respiratory infections. In both places, participants noted that seasonal variations affected the spread of sickness, with colder months showing higher rates of illness. Strangely, in Sulaimani, 50.27% of participants indicated that city weather had no impact on their health, whereas in Stockholm, 73.33% of respondents reported a moderate effect. In Sulaimani, seasonal variation was the most frequently reported factor affecting public health, 57.18%, followed by cold waves, 16.22%, reported by 66 participants. Similarly, in Stockholm, seasonal variation remained the leading factor, 48.00%, reported by 215 participants, while cold waves showed a comparatively higher impact, 29.33%,

TABLE 4: Comparison of weather-related health impact factors between two selected cities

Queries	Categories	Iraq-Sulaimani (376)		Sweden-Stockholm (150)		Statistics	
		Count	Percentage	Count	Percentage	χ^2	*P*-value
Weather description	Humid (Moist) and cold	27	7.18	110	73.33	267.9559	8.80E-57
	Dry and cold	96	25.53	34	22.67		
	Medium	189	50.27	6	4.00		
	Dry and hot	50	13.30	0	0.00		
	Humid and hot	14	3.72	0	0.00		
Impact of city weather on health	No effect	189	50.27	6	4.00	47.20219	1.38E-09
	Too little	96	25.53	34	22.67		
	Somewhat	27	7.18	110	73.33		
	A lot	50	13.30	0	0.00		
	Too much	14	3.72	0	0.00		
Major weather factors affecting public health	The dry period	46	12.23	0	0.00	44.93217	1.50E-08
	Seasonal variation	215	57.18	72	48.00		
	Cold waves	61	16.22	44	29.33		
	Heat waves	30	7.98	8	5.33		
	Types of precipitation (rain, snow, hail, etc.)	23	6.12	23	15.33		
	Natural disaster (flood, storm, etc.)	1	0.27	3	2.00		
Observed how weather patterns and seasonal changes influence disease spread in their cities	No noticeable effect	18	4.79	9	6.00	23.07625	1.22E-04
	Minimal impact	11	2.93	15	10.00		
	Moderate impact	122	32.45	66	44.00		
	Significant impact	174	46.28	45	30.00		
	Significantly effective	51	13.56	15	10.00		
The peak season for illness and medication depends on weather and disease prevalence.	Spring	5	1.33	6	4.00	9.499586	2.33E-02
	Summer	13	3.46	6	4.00		
	Autumn	260	69.15	85	56.67		
	Winter	98	26.06	53	35.33		
Weather-sensitive diseases include seasonal respiratory infections.	Sensitization	26	6.91	21	14.00	16.27837	2.67E-03
	Seasonal epidemics (Infectious diseases)	327	86.97	110	73.33		
	Skin diseases (Dermatosis)	11	2.93	9	6.00		
	Chronic diseases (Non-communicable diseases)	1	0.27	3	2.00		
	Sudden diseases (Acute diseases)	11	2.93	7	4.67		
The area's location and climate influence medicine usage.	No effect	25	6.65	28	18.67	28.78596	9.00E-06
	Slight effect	44	11.70	31	20.67		
	Moderate effect	220	58.51	70	46.67		
	Significant effect	75	19.95	17	11.33		
	Strong effect	12	3.19	4	2.67		

reported by 44 individuals. In Sulaimani city, 46.28% of participants, or 174 individuals, reported that weather patterns and seasonal changes had a significant impact on disease spread, while 32.45%, or 122 participants, observed a moderate impact. In Stockholm, 44.00% of participants, or 66 individuals, noted a moderate impact, and 30.00%, or 45 participants, reported a significant effect. The key findings from the data on climate impact showed that in Sulaimani, 86.97% of participants, or 327 individuals, identified seasonal epidemics (infectious diseases) as the primary weather-sensitive disease. In Stockholm, 73.33% of participants, or 110 individuals, reported the same, with a significant percentage also recognizing seasonal respiratory infections. 220 participants reported a moderate effect of the area's location and climate on medicine usage in Sulaimani, while 70 participants indicated a similar moderate effect in Stockholm. In Sulaimani, 69.15% of participants identified autumn as the peak season for illness and medication, while 56.67% of participants in Stockholm reported the same. In addition, this trend is emphasized and visualized in Fig. 6. Conversely, spring and summer had the lowest reactions and are least linked to pharmaceutical use in both regions. These patterns indicate that illness and higher medicine use are more closely associated with the colder and transitional

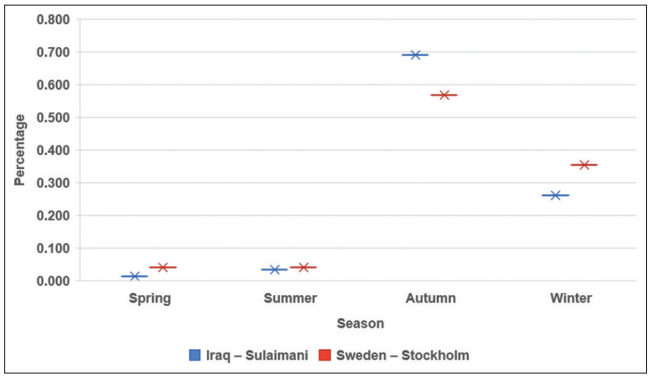


Fig. 6. The depiction of peak illness seasons and medication usage is influenced by weather and disease prevalence.

seasons. The two nations' different climates could be the cause of the differences: Stockholm's longer and colder winters could make people more open to seasonal illnesses for a longer length of time, while Sulaimani's fluctuating autumnal temperatures could cause greater health problems. The current pattern demonstrates how seasonal variations and local weather affect public health practices and attitudes toward medication.

The analysis of all queries exposed significant differences between the two locations based on the Chi-square test (χ^2) and *P*-values. In several cases, such as weather description, the impact of city weather on health, and the main weather factors affecting public health, the *P*-values were extremely low (e.g., $P < 0.001$), representing statistically significant differences between Sulaimani and Stockholm. For example, the weather description showed a clear distinction in the conditions reported, with a significant difference in the outcome of the weather on health. As well, χ^2 for factors like the impact of city weather on health and major weather factors affecting public health were also high, further supporting the statistical significance of the findings. These results highlight the substantial impact of location-specific factors on health, climate, and medication use, with clear differences experiential in the two cities.

According to Table 4, a comparative analysis was conducted to examine categorical differences between Stockholm and Sulaimani in terms of weather perception, health impact, and seasonal illness patterns. Chi-square tests revealed significant associations across all variables. Notably, the description of weather differed strongly between the two cities ($\chi^2 = 267.96$, $P < 0.001$, Cramér's V = 0.72), indicating a very large effect size. Similarly, perceptions regarding the impact of city weather on health ($\chi^2 = 47.20$, $P < 0.001$, Cramér's V = 0.38) and seasonal variation ($\chi^2 = 44.93$, $P < 0.001$, Cramér's V = 0.37) showed moderate to large effect sizes, suggesting substantial perceptual differences. Observations related to disease spread across seasons also varied significantly ($\chi^2 = 23.08$, $P < 0.001$, Cramér's V = 0.27), reflecting a moderate effect size. These findings underscore both statistical and practical differences in weather-related health perceptions and experiences between the two regions.

4.3. Therapeutic Role in Seasonal Illness Prevention

This section evaluates the use of three primary therapies, vaccines, antibiotics, and painkillers, across different locations during illness and examines how seasonal climate variations influence their application and effectiveness.

4.3.1. Assessing vaccine usage for preventing seasonal weather-related illnesses

Vaccines are a primary prophylactic intervention designed to sustain antibody-mediated immunity during emergent outbreaks and seasonal pandemic shifts. Sociocultural and demographic factors modulate vaccine uptake, as populations respond variably to climate influences and weather fluctuations at the start of each season. Such immunization programs protect against influenza and other pandemic threats; for instance, widespread obligatory vaccination was applied during the COVID-19 pandemic. Table 5 presents vaccination uptake rates alongside participants' responses on treatment utilization across fluctuating climatic conditions.

Notably, 76 participants in Sulaimani and 34 participants in Stockholm reported using vaccines to protect against seasonal weather-related illnesses. The results show a remarkably close vaccination rate in both selected cities, with approximately between 20% and 23% of participants reporting vaccine use. In addition, most participants reported a moderate influence of location and climate on vaccine usage, with 36 individuals in Sulaimani and 13 in Stockholm. In Sulaimani, 52.63% of participants (40 persons) used the vaccine for prevention, while 38.16% (29 persons) reported using it for both prevention and treatment. In Stockholm, 70.59% of participants (24 persons) used it for prevention, with 20.59% (7 persons) using it for both purposes. For vaccine use, 52.63% in Sulaimani and 70.59% in Stockholm reported using it for prevention, while 38.16% and 20.59%, respectively, used it for both prevention and treatment. Regarding the impact of climate and disease exposure on vaccine use, 52.63% in Sulaimani and 41.18% in Stockholm indicated a moderate influence. Predictably, autumn was identified as the most ideal season for using vaccines, reported by 69.74% in Sulaimani and 50.00% in Stock.

TABLE 5: Seasonal vaccination patterns and environmental health impacts in two selected cities

Queries	Categories	Iraq-Sulaimani (376)		Sweden-Stockholm (150)		Statistics	
		Count	Percentage	Count	Percentage	χ^2	*P*-value
Using vaccines helps protect against seasonal weather-related illnesses.	No	300	79.79	116	77.33	3.90E-01	5.32E-01
	Yes	76	20.21	34	22.67		
Location and climate influence vaccine usage.	No effect	13	17.11	8	23.53	5.85E+00	2.11E-01
	Slightly	9	11.84	9	26.47		
	Moderate	36	47.37	13	38.24		
	Significantly	17	22.37	4	11.76		
	Strongly	1	1.32	0	0.00		
The use of the vaccine was for prevention or treatment.	Prevention	40	52.63	24	70.59	3.99E+00	2.63E-01
	Treatment	6	7.89	2	5.88		
	Both of them	29	38.16	7	20.59		
	None of them	1	1.32	1	2.94		
The type of climate and disease exposure in the area has impacted vaccine use.	No effect	6	7.89	6	17.65	7.66E+00	1.05E-01
	Slightly	13	17.11	11	32.35		
	Moderate	40	52.63	14	41.18		
	Significantly	16	21.05	3	8.82		
	Strongly	1	1.32	0	0.00		
The most ideal season, weather, and climate for using the vaccine.	Spring	9	11.84	3	8.82	6.85E+00	7.68E-02
	Summer	2	2.63	1	2.94		
	Autumn	53	69.74	17	50.00		
	Winter	12	15.79	13	38.24		

One of the statistical issues observed is that all *P*-values indicate significance ($P < 0.05$), while the Chi-square, as illustrated by χ^2 values, is very low. However, this outcome is mainly due to the small number of participants using vaccines, making the statistical results difficult to control. However, it can be discussed that addressing these issues depends on improving vaccine implementation during climate fluctuations. Perceiving Table 5, an analysis of vaccine-related variables between groups revealed no statistically significant differences. Vaccine usage rates were comparable across locations ($\chi^2 = 0.39$, $P = 0.532$, Cramér's V = 0.04), indicating a very small effect size and suggesting minimal influence of geography on vaccination behavior. Similarly, perceptions of location and climate influence on vaccine decisions ($\chi^2 = 5.85$, $P = 0.211$, Cramér's V = 0.18) and the reported purpose for vaccine use ($\chi^2 = 3.99$, $P = 0.263$, Cramér's V = 0.20) showed small to small–moderate effect sizes, but these differences were not statistically significant. Overall, the findings suggest that vaccine-related attitudes and behaviors were largely consistent between the two populations, with only minor variations that lack statistical support.

4.3.2. Assessing antibiotics usage for preventing seasonal weather-related illnesses

One technique of protection during fluctuating weather and seasonal illnesses is the use of antibiotics as a therapeutic approach. In addition, antibiotics are often used as a seasonal therapy to treat secondary bacterial infections that may follow epidemiologic illnesses during seasonal changes. Their use typically grows during periods of heightened respiratory infections, such as in autumn and winter. However, inappropriate use against viral infections remains a significant public health concern. For this purpose, the study investigated the use of antibiotic therapy during infectious illnesses associated with seasonal weather changes.

According to Table 6, in Sulaimani, 143 participants, mean of 38.03%, used antibiotics for protection, while only 21 participants, mean of 14.00%, did so in Stockholm. This highlights the influence of demographic factors and culture on the participants, even though both groups are from the same Kurdish nation. Regarding the influence of location and climate on antibiotic use, 48.25% of participants in Sulaimani reported a moderate effect, while 24.48% noted a significant effect. In Stockholm, 38.10% observed a moderate impact, and 19.05% saw a significant effect. When asked about the purpose of antibiotic use, 73.43% of participants in Sulaimani used antibiotics for treatment, while 33.33% in Stockholm used them for both prevention and treatment. Regarding the impact of climate and disease exposure, 51.05% of participants in Sulaimani and 38.10% in Stockholm reported a moderate effect. Finally, when identifying the most ideal season for antibiotic use, 58.74% of Sulaimani participants selected autumn, and 37.06% chose winter, while in Stockholm, 57.14% chose autumn and 33.33% chose winter. This result indicates that seasonal weather is highly related to the impact of climate on the use of drug therapy for treating infections.

TABLE 6: Antibiotic patterns and environmental health impacts in two selected cities

Queries	Categories	Iraq-Sulaimani (376)		Sweden-Stockholm (150)		Statistics	
		Count	Percentage	Count	Percentage	χ^2	*P*-value
Using antibiotics helps protect against seasonal weather-related illnesses.	No	233	61.97	129	86.00	2.89E+01	7.78E-08
	Yes	143	38.03	21	14.00		
Location and climate influence antibiotics usage	No effect	15	10.49	5	23.81	4.20E+00	3.79E-01
	Slightly	20	13.99	4	19.05		
	Moderate	69	48.25	8	38.10		
	Significantly	35	24.48	4	19.05		
	Strongly	4	2.80	0	0.00		
The use of the antibiotics was for prevention or treatment	Prevention	11	7.69	5	23.81	1.17E+01	8.55E-03
	Treatment	105	73.43	8	38.10		
	Both of them	25	17.48	7	33.33%		
	None of them	2	1.40	1	4.76		
The type of climate and disease exposure in the area has impacted antibiotics use.	No effect	5	3.50	3	14.29	7.00E+00	1.36E-01
	Slightly	20	13.99	5	23.81		
	Moderate	73	51.05	8	38.10		
	Significantly	40	27.97	5	23.81		
	Strongly	5	3.50	0	0.00		
The most ideal season, weather, and climate for using the antibiotics.	Spring	2	1.40	1	4.76	1.44E+00	6.96E-01
	Summer	4	2.80	1	4.76		
	Autumn	84	58.74	12	57.14		
	Winter	53	37.06	7	33.33		

The use of antibiotics to protect against seasonal weather-related illnesses was found to be statistically non-significant, as the *P*-value was >0.05. However, in Sulaimani, individuals who reported antibiotic use demonstrated a stronger association, reflected by a favorable χ^2 value, leading to the rejection of the null hypothesis. Observing Table 6, significant differences were observed in antibiotic-related practices between the two study locations. Antibiotic usage was notably higher in Sulaimani compared to Stockholm ($\chi^2 = 28.9$, $P < 0.001$, Cramér's V = 0.27), indicating a moderate effect size and suggesting a meaningful influence of geographical context on treatment behavior. Similarly, the reported purpose of antibiotic use also differed significantly between groups ($\chi^2 = 11.7$, p = 0.0085, Cramér's V = 0.26), reflecting another moderate association. These findings imply that antibiotic consumption patterns and motivations vary substantially across the two cities, likely reflecting differences in health beliefs, access, or prescribing practices.

4.3.3. Assessing painkillers usage for preventing seasonal weather-related illnesses

The use of painkillers as a therapeutic approach was assessed for managing seasonal weather-related illnesses, highlighting painkillers as one of the protective therapies during seasonal weather fluctuations. Furthermore, painkiller therapy usage varies notably across different geographical regions, reflecting both environmental influences and cultural healthcare practices. In regions with harsher seasonal changes, such as colder winters or speedy temperature fluctuations, the use of painkillers tends to increase as individuals manage weather-related illnesses, including respiratory infections and musculoskeletal pain.

In Sulaimani, 55.05% of participants, or 207 individuals, reported using painkillers to protect against these issues, while in Stockholm, 42.00%, or 63 individuals, reported the same as stated in Table 7. Participants in both Sulaimani and Stockholm reported that climate and location moderately influenced their usage of painkillers, with 103 participants (49.76%) in Sulaimani and 29 participants (46.03%) in Stockholm admitting a moderate effect. Most individuals primarily used painkillers for treatment, with 139 participants (67.15%) in Sulaimani and 44 participants (69.84%) in Stockholm indicating this resolution, while a smaller proportion used them for both prevention and treatment, reported by 53 participants (25.60%) in Sulaimani and 15 participants (23.81%) in Stockholm. Climate and disease exposure were also noted to moderately affect painkiller usage, with 93 participants (44.93%) in Sulaimani and 32 participants (50.79%) in Stockholm reporting moderate influence. Regarding the ideal season for using painkillers, autumn was the most commonly identified season by 109 participants (52.66%) in Sulaimani and 30 participants (47.62%) in Stockholm, followed closely by winter, selected by 89 participants (43.00%) in Sulaimani and 30 participants (47.62%) in Stockholm, reflecting the seasonal rise in weather-related illnesses.

TABLE 7: Painkillers patterns and environmental health impacts in two selected cities

Queries	Categories	Iraq-Sulaimani (376)		Sweden-Stockholm (150)		Statistics	
		Count	Percentage	Count	Percentage	χ^2	*P*-value
Using painkillers helps protect against seasonal weather-related illnesses.	No	169	44.95	87	58.00	7.31E+00	6.85E-03
	Yes	207	55.05	63	42.00		
Location and climate influence painkiller usage.	No effect	26	12.56	12	19.05	9.57E+00	4.83E-02
	Slightly	48	23.19	16	25.40		
	Moderate	103	49.76	29	46.03		
	Significantly	0	0.00	0	0.00		
	Strongly	30	14.49	6	9.52		
The use of the painkillers was for prevention or treatment.	Prevention	10	4.83	4	6.35	1.85E+00	6.04E-01
	Treatment	139	67.15	44	69.84		
	Both of them	53	25.60	15	23.81		
	None of them	5	2.4	0	0.00		
The type of climate and disease exposure in the area has impacted painkillers use.	No effect	9	4.35	7	11.11	8.19E+00	8.47E-02
	Slightly	37	17.87	12	19.05		
	Moderate	93	44.93	32	50.79		
	Significantly	61	29.47	12	19.05		
	Strongly	7	3.38	0	0.00		
The most ideal season, weather, and climate for using the painkillers.	Spring	1	0.48	2	3.17	4.37E+00	2.24E-01
	Summer	8	3.86	1	1.59		
	Autumn	109	52.66	30	47.62		
	Winter	89	43.00	30	47.62		

Statistical analysis showed that the use of painkillers in relation to seasonal weather illnesses has been significantly influenced by climate and location ($P = 0.0483$). The majority used painkillers mostly for treatment purposes, with moderate climatic impact detected in both cities. The preference for using painkillers during the autumn and winter seasons further supports the seasonal pattern of weather-related illnesses.

4.3.4. Comparison discussion of usage frequency for vaccines, antibiotics, and painkillers

The differences in treatment usage frequency between Sulaimani and Stockholm highlight the impact of location and climate on health behaviors, and the collected data provide important insights for analysis. Fig. 7 presents a comparison of vaccine, antibiotic, and painkiller usage among participants from both cities, reflecting seasonal medication preferences. In both locations, autumn emerged as the most common season for medication use; however, a larger proportion of participants in Sulaimani (approximately 70%) reported taking medications during this period compared to about 55% in Stockholm. In Sulaimani, vaccine usage was notably low, with nearly 300 participants, indicating that they did not use vaccines, while only around 70 participants reported vaccine usage. Conversely, the use of antibiotics and painkillers was considerably higher in Sulaimani, with approximately 140 participants using antibiotics and slightly over 200 using painkillers. In Stockholm, vaccine usage was also limited, but proportionally, more participants used vaccines than antibiotics. Painkiller use was more prevalent than antibiotic use in Stockholm, with around 60 participants using painkillers compared to fewer than 30 for antibiotics. Overall, medication usage was higher in Sulaimani, particularly for antibiotics and painkillers, whereas Stockholm exhibited a lower but relatively more balanced pattern of medication use across the different treatments.

4.4. Impact of Health Priorities on Climate-driven Drugs

This section provides a crucial evaluation of the results related to the overall impact of climate on drug therapy usage in various ways. Furthermore, it focuses on responses that explored general opinions about weather-related health issues. In both locations, most participants assigned some degree of importance to addressing these challenges. According to Table 8, the key findings for each query are analyzed and discussed separately below, focusing on general opinions and practices related to weather-induced health challenges:

- Priority in addressing health problems was evident in both locations, where most participants emphasized the need to tackle weather-related health issues. In Sulaimani, 60.9% of respondents indicated that they assigned to some extent or to a great extent a priority to finding solutions, compared to 48% in Stockholm. Notably, a higher percentage of participants in Stockholm, 12.67%, reported never prioritizing such solutions, compared to only 4.26% in Sulaimani

TABLE 8: Compares responses on health priorities related to climate-driven drugs in these two regions

Queries	Categories	Iraq-Sulaimani (376)		Sweden-Stockholm (150)		Statistics	
		Count	Percentage	Count	Percentage	χ^2	*P*-value
The level of priority in seeking solutions to health problems caused by the weather is high.	Never	16	4.26	19	12.67	7.12E+00	6.81E-02
	Little	131	34.84	59	39.33		
	To some extent	163	43.35	51	34.00		
	To a great extent	66	17.55	21	14.00		
In general, the treatment or medication most often used is due to the local climate	Vaccine	23	6.12	14	9.33	2.86E+01	2.75E-06
	Antibiotics	118	31.38	12	8.00		
	Painkillers (Analgesics)	138	36.70	64	42.67		
	Multivitamins (Dietary supplements)	97	25.80	60	40.00		
Any public health policies or guidelines related to the climate in your city	No policy	100	26.60	15	10.00	5.93E+01	8.22E-13
	To some extent	104	27.66	42	28.00		
	In good shape	8	2.13	49	32.67		
	Unaware	164	43.62	44	29.33		
Describe the accessibility and permissibility of obtaining medicines for prevention and treatment in the area	Easy	240	63.83	52	34.67	8.99E+01	2.35E-19
	Relatively (Somewhat) easy	122	32.45	50	33.33		
	Difficult	10	2.66	28	18.67		
	Very difficult	4	1.06	20	13.33		
Types of drugs that are helpful and necessary for treating health problems related to the climate in the city include	Vaccine	76	20.21	30	20.00	1.67E+01	8.00E-04
	Antibiotics	88	23.40	13	8.67		
	Painkillers (Analgesics)	70	18.62	40	26.67		
	Multivitamins (Dietary supplements)	142	37.77	67	44.67		

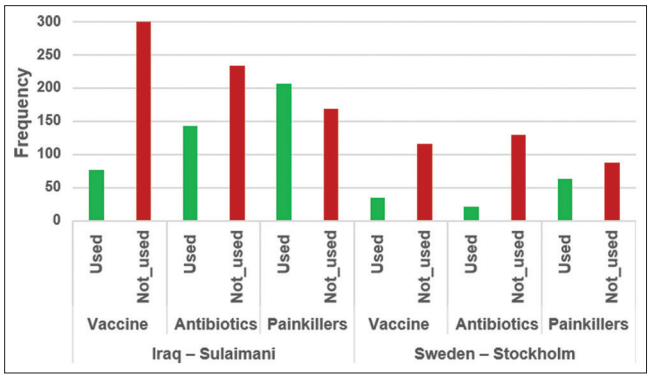


Fig. 7. The differences in frequency of treatment usage based on the impact of different locations.

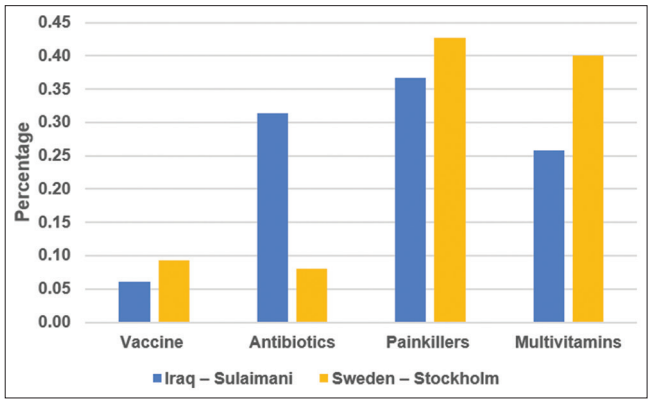


Fig. 8. The percentage of the treatment or medication most often used due to the local climate.

- Common treatments for climate-related health issues showed that painkillers, specifically analgesics, were the most commonly reported treatment in both cities, particularly in Stockholm with 42.67% and in Sulaimani with 36.7%. Multivitamins were the next most preferred treatment, with a higher usage in Stockholm at 40% compared to 25.8% in Sulaimani. The use of antibiotics was significantly higher in Sulaimani, reaching 31.38%, compared to only 8% in Stockholm. The comparison discussion, as indicated by the percentage values, is clearly illustrated in Fig. 8

- Public health policies related to climate showed considerable variation in awareness and approval between the two locations. In Sulaimani, 164 participants, representing 43.62%, reported being uninformed of any climate-related health policies, compared to 44 participants or 29.33% in Stockholm. In contrast, 49% in Stockholm, accounting for 32.67%, indicated that climate-related policies were in respectable, whereas only 8 participants or 2.13% stated the same in Sulaimani

- Accessibility of medicines showed notable differences between the two cities. In Sulaimani, 240 participants, or 63.83%, reported that obtaining medicines was easy, compared to 52 participants or 34.67% in Stockholm. Meanwhile, 48 applicants in Stockholm, representing 32%, described access as difficult or very difficult, compared to only 14 participants or 3.72% in Sulaimani
- The perceived necessity of different drug types revealed that multivitamins were the most commonly cited medications for climate-related health issues, with 142 contestants or 37.77% in Sulaimani and 67 participants or 44.67% in Stockholm. Antibiotics were reported more frequently in Sulaimani by 88 participants or 23.4%, compared to 13 contributors or 8.67% in Stockholm. The identification of this issue is highly valuable, as demonstrated in Fig. 9.

The discussion of *P*-values as discussed by statistical models, as presented in the table, highlights the statistical significance of the observed differences across various factors. The low *P*-values indicate a high level of confidence in the results, suggesting that the differences observed between the locations are not due to chance. This strengthens the reliability of the conclusions drawn from the data.

Finally, two open-ended questions were included to further explore the topic. The majority of participants reported making lifestyle or health habit changes to protect themselves against illnesses caused by adverse weather conditions. Many emphasized the importance of public education on the negative health effects of climate change, particularly through the expertise of medical professionals. Notably, in Stockholm, some respondents linked weather-related factors to mental health issues. Others highlighted that those environmental factors, such as air and water pollution, also contribute to health problems, in addition to weather-related factors. Overall, participants expressed a preference for prevention over treatment when addressing weather-induced health concerns.

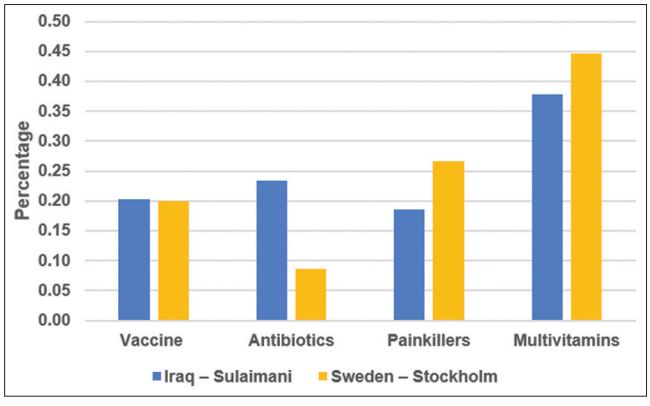


Fig. 9. The impact of using various types of drugs that are essential for treating climate-related health issues in the city includes.

5. CONCLUSION AND FUTURE WORK

This study highlights the influence of climate and geography on drug therapy practices among Kurdish residents in Sulaimani and Stockholm, including Kurdish refugees living in Stockholm. It also considers the impact of sociodemographic factors on the study outcomes and related healthcare challenges. While Stockholm shows a stronger emphasis on preventive care, evidenced by higher use of multivitamins and vaccinations, Sulaimani exhibits significantly greater antibiotic use, indicating a more treatment-oriented approach. Painkillers emerged as the most commonly used drugs in both locations, though their usage was notably higher in Stockholm, suggesting that seasonal health issues are a major concern regardless of geographic setting. According to the research, Sulaimani relies more on rapid treatment, while Stockholm's health strategies go toward prevention, showing how environmental pressures and healthcare systems adapt differently. The comparative ratio findings reveal that colder climates encourage a stronger focus on defensive health behaviors, while more variable climates drive higher rates of reactive treatments. Statistical model analysis helped isolate these trends by minimizing cultural bias; however, some effects were still clearly impacted.

Overall, the results suggest that integrating geographical and climate-sensitive strategies into healthcare planning is essential for improving medical outcomes. Future research should extend the scope to different cultural groups and regions, leading to broader validation. Understanding these regional dynamics can lead to more personalized, resilient healthcare systems that better respond to environmental and demographic shifts. The study suggests that healthcare systems in warmer, more variable climates in Sulaimani should need to emphasize infection control due to higher treatment-driven medication use, whereas colder climates might prioritize preventive care. However, these recommendations require validation in broader contexts, accounting for regional healthcare infrastructure and cultural practices.

Naturally, the study has certain limitations that should be addressed in future data collection efforts, particularly in Sulaimani. Online recruitment in this city may underrepresent individuals with limited or no internet access, potentially

affecting the inclusiveness and representativeness of the sample. While participants shared the same ethnic background, potential confounders such as cultural adaptation, healthcare access, and socioeconomic differences across regions were not fully controlled. These factors are acknowledged as limitations and may influence the interpretation of some findings. One challenge was the difficulty in selecting and identifying the necessary Kurdish residents for data collection in Stockholm. In addition, obtaining data on the weather impact over the past 10 years proved challenging, making it hard to analyze and integrate this information for improving the comparative study. Future studies could utilize Internet of Things-based (IoT-based) environmental sensors in combination with medication-tracking applications to monitor real-time interactions between climate conditions and drug usage [65], [66]. In addition, machine learning algorithms in artificial intelligence can be applied to analyze the data [67]. The data could also be classified and optimized using parameter-tuning algorithms, particularly those related to genetic algorithms and real-world bioinformatic algorithms [68], [69]. Finally, the most critical and apparent limitation is the lack of cross-cultural generalizability; extending the findings beyond Kurdish populations will require validation in more diverse ethnic and cultural groups.

REFERENCES

[1] A. C. Krefis, M. Augustin, K. H. Schlünzen, J. Oßenbrügge and J. Augustin. "How does the Urban environment affect health and well-being? A systematic review". *Urban Science*, vol. 2, no. 1, p. 21, 2018.

[2] National Research Council, Division on Earth and Life Studies, Board on Earth Sciences and Resources, Committee on Identifying Data Needs for Place-Based Decision Making and Committee on Geography. *Community and Quality of Life: Data Needs for Informed Decision Making*. National Academies Press, United States, 2002.

[3] P. Churski and R. Perdał. "Geographical differences in the quality of life in Poland: Challenges of regional policy". *Social Indicators Research*, vol. 164, no. 1, pp. 31-54, 2022.

[4] Z. Yang, B. Yang, P. Liu, Y. Zhang and X. C. Yuan. "Impact of temperature on physical and mental health: Evidence from China". *Weather Climate and Society*, vol. 13, no. 4, pp. 709-727, 2021.

[5] A. Lebano, S. Hamed, H. Bradby, A. Gil-Salmerón, E. Durá-Ferrandis, J. Garcés-Ferrer, F. Azzedine, E. Riza, P. Karnaki, D. Zota and A. Linos. "Migrants' and refugees' health status and healthcare in Europe: A scoping literature review". *BMC Public Health*, vol. 20, no. 1, p. 1039, 2020.

[6] S. C. Ng, C. N. Bernstein, M. H. Vatn, P. L. Lakatos, E. V. Loftus Jr., C. Tysk, C. O'Morain, B. Moum, J. F. Colombel and Epidemiology and Natural History Task Force of the International Organization of Inflammatory Bowel Disease (IOIBD). "Geographical variability and environmental risk factors in inflammatory bowel disease". *Gut*, vol. 62, no. 4, pp. 630-649, 2013.

[7] M. Kumar, P. Mazumder, S. Mohapatra, A. Kumar Thakur, K. Dhangar, K. Taki, S. Mukherjee, A. Kumar Patel, P. Bhattacharya, P. Mohapatra, J. Rinklebe, M. Kitajima, F. I. Hai, A. Khursheed, H. Furumai, C. Sonne and K. Kuroda. "A chronicle of SARS-CoV-2: Seasonality, environmental fate, transport, inactivation, and antiviral drug resistance". *Journal of Hazardous Materials*, vol. 405, p. 124043, 2021.

[8] K. M. Cooper, C. McMahon, I. Fairweather and C. T. Elliott. "Potential impacts of climate change on veterinary medicinal residues in livestock produce: An Island of Ireland perspective". *Trends in Food Science Technology*, vol. 44, no. 1, pp. 21-35, 2015.

[9] A. O. Abhadionmhen, E. P. K. Imarenezor, A. C. Ogodo and A. A. Ahuchaogu. "Socioeconomic and cultural factors influencing traditional medicine (TM) use in Nigeria: A systematic mixed method review". *European Journal of Integrative Medicine*, vol. 74, p. 102436, 2025.

[10] D. O. Hasan, A. M. Aladdin, A. A. H. Amin, T. A. Rashid, Y. H. Ali, M. Al-Bahri, J. Majidpour, L. Batrancea and E. S. Masca. "Perspectives on the impact of E-learning pre-and post-COVID-19 pandemic-the case of the Kurdistan region of Iraq". *Sustainability*, vol. 15, no. 5, p. 4400, 2023.

[11] A. Pietraszek, S. Agrawal, M. Dróżdż, S. Makuch, I. Domański, T. Dudzik, K. Dudek and M. Sobieszczańska M. "Sociodemographic and health-related factors influencing drug intake among the elderly population". *International Journal of Environmental Research and Public Health*, vol. 19, no. 14, p. 8766, 2022.

[12] I. Badinski, A. Finkelstein, M. Gentzkow, and P. Hull. "*Geographic Variation in Healthcare Utilization: The Role of Physicians*". *NBER Working Paper31749*, vol. 16, p. 1-10, 2023.

[13] M. Dorn. "Medical geography in historical perspective (review)". *Bulletin of the History of Medicine*, vol. 76, no. 3, pp. 617-619, 2002.

[14] K. E. Jones, N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman and P. Daszak. "Global trends in emerging infectious diseases". *Nature*, vol. 451, no. 7181, pp. 990-993, 2008.

[15] Y. Palmeiro-Silva, R. Aravena-Contreras, J. Izcue Gana, R. González Tapia and I. Kelman. "Climate-related health impact indicators for public health surveillance in a changing climate: A systematic review and local suitability analysis". *The Lancet Regional Health Americas*, vol. 38, p. 100854, 2024.

[16] A. Udali, H. J. Persson, B. Talbot and S. Grigolato. "SAR data and harvesting residues: An initial assessment of estimation potential". *Earth*, vol. 5, no. 4, pp. 945-962, 2024.

[17] P. M. Polgreen and E. L. Polgreen. "Infectious diseases, weather, and climate". *Clinical Infectious Diseases*, vol. 66, no. 6, pp. 815-817, 2018.

[18] W. Kisiała and I. Rącka. "Spatial and statistical analysis of Urban poverty for sustainable city development". *Sustainability*, vol. 13, no. 2, p. 858, 2021.

[19] V. S. Limaye, J. Vargo, M. Harkey, T. Holloway and J. A. Patz. "Climate change and heat-related excess mortality in the Eastern USA". *Ecohealth*, vol. 15, no. 3, pp. 485-496, 2018.

[20] S. M. Babin. "Weather and climate effects on disease background levels". *Johns Hopkins APL Tech Dig*, vol. 24, no. 4, pp. 343-348, 2003.

[21] L. P. Dawson, E. Andrew, Z. Nehme, J. Bloom, S. Cox, D. Anderson, M. Stephenson, J. Lefkovits, A. J. Taylor, D. Kaye, Y. Guo, K. Smith and D. Stub. "Temperature-related chest pain presentations and future projections with climate change". *Science of The Total Environment*, vol. 848, p. 157716, 2022.

[22] J. A. Patz, A. K. Githeko, J. P. McCarty, S. Hussein, U. Confalonieri and N. De Wet. "Climate change and infectious diseases". *Climate Change and Human Health Risks and Responses*, vol. 2, pp. 103-132, 2003.

[23] X. Wu, Y. Lu, S. Zhou, L. Chen and B. Xu. "Impact of climate change on human infectious diseases: Empirical evidence and human adaptation". *Environment International*, vol. 86, pp. 14-23, 2016.

[24] A. Silversides. "CMA opposes gender discrimination against doctors". *Canadian Medical Association Journal*, vol. 181, p. E311, 2008.

[25] J. C. Semenza and B. Menne. "Climate change and infectious diseases in Europe". *The Lancet Infectious Diseases*, vol. 9, no. 6, pp. 365-375, 2009.

[26] F. Messaoudene, S. Boukraa, S. C. Boubidi, A. Guerzou and A. Ouahabi. "Human cutaneous leishmaniasis in North Africa and its threats to public health: A statistical study focused on Djelfa (Algeria)". *Microorganisms*, vol. 11, no. 10, p. 2608, 2023.

[27] J. E. Bloom, Z. Nehme, E. Andrew, L. P. Dawson, H. Fernando, S. Noaman, M. Stephenson, D. Anderson, V. Pellegrino, S. Cox, J. Lefkovits, W. Chan, D. M. Kaye, K. Smith and D. Stub. "Hospital characteristics are associated with clinical outcomes in patients with cardiogenic shock". *Shock*, vol. 58, no. 3, pp. 204-210, 2022.

[28] T. Hartig, R. Catalano and M. Ong. "Cold summer weather, constrained restoration, and the use of antidepressants in Sweden". *Journal of Environmental Psychology*, vol. 27, no. 2, pp. 107-116, 2007.

[29] J. Talib and S. Ahmed. "Main effect of climate on geographical distribution of asthma in Sulaimani province". *Journal of Garmian University*, vol. 6, no. 2, pp. 10-20, 2019.

[30] C. Astolphi Lima, S. Alsunaidi, S. Lowe, D. B. Hogan, L. Dennett, C. A. Jones and S. Yamamoto. "Exploring the influence of weather variability and climate change on health outcomes in people living with dementia: A scoping review protocol". *PLoS One*, vol. 19, no. 6, p. e0304181, 2024.

[31] R. Basu and J. M. Samet. "Relation between elevated ambient temperature and mortality: A review of the epidemiologic evidence". *Epidemiologic Reviews*, vol. 24, no. 2, pp. 190-202, 2002.

[32] H. Eng and J. B. Mercer. "Seasonal variations in mortality caused by cardiovascular diseases in Norway and Ireland". *Journal of Cardiovascular Risk*, vol. 5, no. 2, pp. 89-95, 1998.

[33] F. Amuakwa-Mensah, G. Marbuah and M. Mubanga. "Climate variability and infectious diseases nexus: Evidence from Sweden". *Infectious Disease Modelling*, vol. 2, no. 2, pp. 203-217, 2017.

[34] L. P. Dawson, E. Andrew, Z. Nehme, J. Bloom, S. Biswas, S. Cox, D. Anderson, M. Stephenson, J. Lefkovits, A. J. Taylor, D. Kaye, K. Smith and D. Stub. "Association of socioeconomic status with outcomes and care quality in patients presenting with undifferentiated chest pain in the setting of universal health care coverage". *Journal of the American Heart Association*, vol. 11, no. 7, p. e024923, 2022.

[35] A. S. Zakey, M. M. Abdel-Wahab, J. B. C. PETTERSSON, M. J. Gatari and M. Hallquist. "Seasonal and spatial variation of atmospheric particulate matter in a developing megacity, the Greater Cairo, Egypt". *Atmosfera*, vol. 21, no. 2, pp. 171-189, 2008.

[36] A. J. McMichael. "Extreme weather events and infectious disease outbreaks". *Virulence*, vol. 6, no. 6, pp. 543-547, 2015.

[37] A. Jutla, E. Whitcombe, N. Hasan, B. Haley, A. Akanda, A. Huq, M. Alam, R. B. Sack and R. Colwell. "Environmental factors influencing epidemic cholera". *The American Society of Tropical Medicine and Hygiene*, vol. 89, no. 3, pp. 597-607, 2013.

[38] M. Neira, K. Erguler, H. Ahmady-Birgani, N. D. Al-Hmoud, R. Fears, C. Gogos, N. Hobbhahn, M. Koliou, L. G. Kostrikis, J. Lelieveld, A. Majeed, S. Paz, Y. Rudich, A. Saad-Hussein, M. Shaheen, A. Tobias and G. Christophides. "Climate change and human health in the Eastern Mediterranean and Middle East: Literature review, research priorities and policy suggestions". *Environmental Research*, vol. 216, p. 114537, 2023.

[39] C. Pfister and D. Krämer. "*The Relaunch of Historical Climate Impact Research-a Timely Challenge for Historical Climatology*"; 2011. Available from: https://boris.unibe.ch/id/eprint/12334 [Last accessed on 2025 May 01].

[40] F. Zanasi and R. Conte Keivabu. "Extreme temperatures and morbidity in old age in Europe". *Vienna Yearbook of Population Research*, vol. 22, pp. 1-28, 2024.

[41] H. Khraishah, B. Alahmad, R. L. Ostergard Jr., A. AlAshqar, M. Albaghdadi, N. Vellanki, M. M. Chowdhury, S. G. Al-Kindi, A. Zanobetti, A. Gasparrini and S. Rajagopalan. "Climate change and cardiovascular disease: Implications for global health". *Nature Reviews Cardiology*, vol. 19, no. 12, pp. 798-812, 2022.

[42] M. A. Mohammad, S. Koul, R. Rylance, O. Fröbert, J. Alfredsson, A. Sahlén, N. Witt, T. Jernberg, J. Muller and D. Erlinge. "Association of weather with day-to-day incidence of myocardial infarction: A SWEDEHEART nationwide observational study". *JAMA Cardiology*, vol. 3, no. 11, pp. 1081-1089, 2018.

[43] S. Holmberg, A. Rignell-Hydbom, C. H. Lindh, B. A. Jönsson, A. Thelin and L. Rylander. "High levels of vitamin D associated with less ischemic heart disease - a nested case-control study among rural men in Sweden". *Annals of Agricultural and Environmental Medicine*, vol. 24, no. 2, pp. 288-293, 2017.

[44] M. G. Abrignani, A. Lombardo, A. Braschi, N. Renda and V. Abrignani. "Climatic influences on cardiovascular diseases". *World J Cardiology*, vol. 14, no. 3, pp. 152-169, 2022.

[45] H. R. Rossi. "*The Association of Endometriosis on Body Size, Pain Perception, Comorbidity and Work Ability in the Northern Finland Birth cohort 1966: Long-Term Effects of Endometriosis on Women's Overall Health*". Oulun Yliopisto, University of Oulu; 2021. Available from: https://urn.fi/urn: isbn:9789526230023 [Last accessed on 2025 May 01].

[46] M. E. Keim. "Building human resilience the role of public health preparedness and response as an adaptation to climate change". *American Journal of Preventive Medicine*, vol. 35, no. 5, pp. 508-516, 2008.

[47] K. L. Ebi, P. Berry, D. Campbell-Lendrum, G. Cissé, J. Hess, N. Ogden and R. Schnitter. "*Health System Adaptation to Climate Variability and Change*". Washington, DC: Global Center on Adaptation; 2019. Available from: https://gca.org/wp/content/uploads/2020/12/healthsystemadaptationtoclimatevariabilityandchange_0.pdf [Last accessed on 2025 May 01].

[48] S. R. Mohammed-Taha, A. M. Aladdin, A. I. Mustafa, D. O. Hasan, R. K. Muhammed and T. A. Rashid. "Analyzing the impact of the pandemic on insomnia prediction using machine learning classifiers across demographic groups". *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 14, no. 1, p. 51, 2025.

[49] U. Jørgensen. "Mapping and navigating transitions-the multi-level perspective compared with arenas of development". *Research Policy*, vol. 41, no. 6, pp. 996-1010, 2012.

[50] R. J. Johnston. *City and Society: An Outline for Urban Geography*. Routledge, England, UK, 2013.

[51] C. I. Hussein, H. Q. Rasul and D. A. M. A. Al Manmi. “Environmental justice in water quality: Sulaimani city as a case study”. *The Scientific Journal of Cihan University Sulaimaniya*, vol. 7, no. 2, pp. 72-97, 2023.

[52] S. S. Mohammed, H. Q. Rasul and H. D. Mohammed. “GIS-based spatial analysis of the evolution of residential developments; A case-study of Sulaimani city, Iraq,” *Kurdistan Journal of Applied Research*, vol. 8, no. 2, pp. 9-32, 2024.

[53] Z. M. J. Altaee and M. J. Al-Ani. “Urban renewal as a tool of cities sustainability Sulaimani as a case study-North Iraq”. *Periodicals of Engineering and Natural Sciences (PEN)*, vol. 8, no. 2, pp. 971-990, 2020.

[54] S. H. Amin Ahmad. “Impact of climate change on heating and cooling energy consumption in Sulaimani city/Iraq-Kurdistan region”. *Journal of Garmian University*, vol. 9, no. 2, pp. 498-508, 2022.

[55] K. D. Muslih and A. M. Abbas. “*Climate of Iraq*”. Springer, Berlin, pp. 19-47, 2024.

[56] S. H. Al Jarah, B. Zhou, R. J. Abdullah, Y. Lu and W. Yu. “Urbanization and Urban sprawl issues in city structure: A case of the Sulaymaniah Iraqi Kurdistan region”. *Sustainability*, vol. 11, no. 2, p. 485, 2019.

[57] L. Vogiazides and H. Mondani. “Neighbourhood trajectories in Stockholm: Investigating the role of mobility and *in situ* change”. *Applied Geography*, vol. 150, p. 102823, 2023.

[58] H. Engström. “*Influences of Urban Geometry on Local Climate a Case Study of the City of Stockholm*.” University of Gothenburg, Göteborg; 2023. Available from: https://gupea.ub.gu.se/handle/2077/77321 [Last accessed on 2025 May 01].

[59] K. Błażejczyk, J. Baranowski, G. Jendritzky, A. Błażejczyk, P. Bröde and D. Fiala. “Regional features of the bioclimate of Central and Southern Europe against the background of the Köppen-Geiger climate classification”. *Geographia Polonica*, vol. 88, no. 3, pp. 439-453, 2015.

[60] D. Magnusson. “*Impact of Different Land Uses on Local Climate in the Stockholm Region*”. University Of Gothenburg, Gothenburg; 2023. Available from: https://gupea.ub.gu.se/handle/2077/77373 [Last accessed on 2025 May 01].

[61] D. O. Hasan and A. M. Aladdin. “Sleep-related consequences of the COVID-19 pandemic: A survey study on insomnia and sleep apnea among affected individuals”. *Insights in Public Health Journal*, vol. 5, no. 2, 2025.

[62] D. Khanna, C. Peltzer, P. Kahar and M. S. Parmar. “Body mass index (BMI): A screening tool analysis”. *Cureus*, vol. 14, p. e22119, 2022.

[63] R. Azen and C. M. Walker. *Categorical Data Analysis for the Behavioral and Social Sciences*. 2nd ed. Routledge, New York , 2021.

[64] A. A. H. Amin, A. M. Aladdin, D. O. Hasan, S. R. Mohammed-Taha and T. A. Rashid. “Enhancing algorithm selection through comprehensive performance evaluation: Statistical analysis of stochastic algorithms,” *Computation*, vol. 11, no. 11, p. 231, 2023.

[65] A. M. Aladdin, J. M. Abdullah, K. O. M. Salih, T. A. Rashid, R. Sagban, A. Alsaddon, A. Chhabra, S. Vimal and I. Banerjee. “Fitness-dependent optimizer for IoT healthcare using adapted parameters: A case study implementation”. In: *Practical Artificial Intelligence for Internet of Medical Things*, CRC Press, United States, 2023, pp. 45-61.

[66] H. S. Abdulla and A. M. Aladdin. “Enhancing design and authentication performance model: A multilevel secure database management system”. *Future Internet*, vol. 17, no. 2, p. 74, 2025.

[67] D. O. Hassan and B. A. Hassan. “A comprehensive systematic review of machine learning in the retail industry: Classifications, limitations, opportunities, and challenges”. *Neural Computing and Applications*, vol. 37, no. 4, pp. 2035-2070, 2025.

[68] A. M. Aladdin and T. A. Rashid. “*Leo: Lagrange Elementary Optimization*”. [arXiv Preprint] arXiv:2304.05346; 2023.

[69] R. Mohammed, N. K. Al-Salihi, T. A. Rashid, A. M. Aladdin, M. Mohammadi and J. Majidpour. “Artificial cardiac conduction systemsimulating heart function for advanced computational problem solving”. In: *Multi-Objective Optimization Techniques*, CRC Press, United States, p. 314-339, 2025.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Enhancing Clinical Decision Support: A Deep Learning Approach for Automated Diagnosis of Eye Diseases from Fundus Images

Ismael Abdulkareem Ali, Sozan Abdulla Mahmood

Computer Department, College of Science, University of Sulaimani, Sulaymaniyah 46001, Kurdistan, Iraq

ABSTRACT

Background and Objective: One of the most crucial sensory organs that helps the human brain receive information about the outside world is the eye. Due to its structural features, the back surface of the eye (retina) provides valuable insights about various disorders. It is essential to protect the eyes from diseases that could lead to vision impairment. If diseases affecting the retina are not identified and treated promptly, vision loss cannot be reversed. Therefore, effective automatic detection systems are necessary, as manual diagnosis is not only time-consuming, expensive, and labor-intensive but also requires a high level of expertise. To address this issue, many deep learning (DL)-based solutions have been proposed for screening retinal conditions. This study aimed to develop an effective system for the automated classification of four major eye conditions to support clinical decision-making. **Methods:** In this research, various convolutional neural network (CNN) architectures were applied to the dataset, and their performance was recorded. The CNN models are the common transfer learning pre-trained models on the ImageNet dataset. Finally, we developed a hybrid DL model combining DenseNet169 and MobileNetV1 to extract deep features from fundus images and perform multiclass classification into four categories: diabetic retinopathy, cataract, glaucoma, and normal fundus. **Results:** This hybrid approach yielded impressive results, attaining 92.99%, 93.02%, 92.85%, 92.90%, and 98.77% for accuracy, precision, recall, F1-score, and area under the curve (AUC) on a publicly available Kaggle dataset, i.e., eye disease classification. These results indicate that the hybrid approach enhances classification accuracy compared to other individual pre-trained CNN models. **Conclusion:** In summary, this study evaluated a substantial number of pre-trained models and developed a framework based on the top two optimal-performing models. Given that retinal image detection and diagnosis are critical for patient eye therapy and rehabilitation, our study offers an innovative framework that can function as a diagnostic aid for eye-related diseases.

Index Terms: healthcare, fundoscopy, retinal fundus images, multi-class classification, hybrid deep learning.

1. INTRODUCTION

Among our senses, vision is the dominant one, playing an essential role in every aspect and phase of our lives. When vision is impaired, the quality of life and the ability to perform daily tasks are also affected. The World Health Organization reports that at least 2.2 billion people worldwide suffer from near or distant vision impairment. It is either untreated or could have been avoided in at least 1 billion of these cases [1].

For ophthalmologists and healthcare centers, designing a computer-aided diagnosis strategy to identify retinal disorders is highly beneficial, as it enables early detection and ensures proper patient treatment [2]. Fortunately, automated analysis and diagnosis have been made possible by machine learning (ML) algorithms, which let ophthalmologists detect diseases early [3]. To address the practical needs of many

Access this article online

DOI: 10.21928/uhdjst.v9n2y2025.pp61-76

E-ISSN: 2521-4217

P-ISSN: 2521-4209

Copyright © 2025 Ali IA, Mahmood SA. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Ismael.ali@univsul.edu.iq

Received: 05-05-2025 Accepted: 21-06-2025 Published: 24-08-2025

patients with retinal illnesses, artificial intelligence (AI) and ophthalmology treatment are being combined. A branch of ML, i.e., deep learning (DL), finds widespread use in AI. One of its best solutions is convolutional neural networks (CNNs), which are excellent at automatically extracting features and learning [4]. Currently, CNNs have a great role in classifying and diagnosing medical images. The current success of DL technology in ML drives new research and development initiatives to enhance computer-aided diagnosis performance and expand its application to a broader range of complicated clinical tasks.

In general, retinal fundus color images are utilized to help clinicians diagnose diseases out of all types of ophthalmic data, like fluorescein angiography and optical coherence tomography. The tissue behind the eyeball, which includes the optic disc/cup, macula, and blood vessels, is called the retina. The morphological variability in retinal fundus images may illustrate several ocular diseases, such as diabetic retinopathy (DR), glaucoma, cataracts, and more [5]. Fig. 1 shows color fundus images exhibiting some retinal pathologies and morphologies. Special proteins make up the transparent lens in the front of the eye. When these proteins degrade and produce foggy areas on the lens, cataracts develop. The patches may enlarge over time, resulting in blurred vision [6]. According to "The Lancet Global Health," 100 million people in the world have cataracts, 17 million of them are blind, and 83 million have a visual impairment because of cataracts. People with diabetes are more likely to develop DR, which is an eye disorder that affects blood vessels of the retina and can lead to complete blindness and vision loss [7]. DR is a microvascular complication

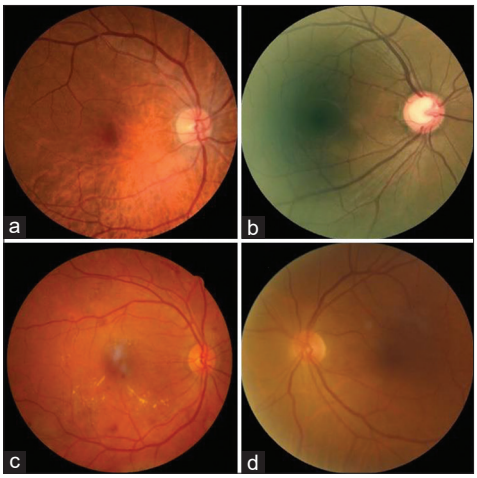


Fig. 1. Retinal fundus images in various conditions sourced from the eye disease classification dataset [45]: (a) normal; (b) glaucoma; (c) diabetic retinopathy; (d) cataract.

of diabetes mellitus (DM) that affects one in three people with DM [8]. A few of the retinal anomalies that DR may produce are microaneurysms, hard exudates, soft exudates or cotton wool spots, hemorrhages, neovascularization (NV), and diabetic macular edema (DME). DR is classified into five stages based on the presence of clinical features: mild nonproliferative DR (NPDR), moderate NPDR, severe NPDR, PDR, and DME [9], [10].

A class of disorders known as glaucoma that damages the optic nerve in the eye, which can lead to blindness or visual loss [11]. It occurs when excess fluid accumulates in the anterior region of the eye, raising intraocular pressure and harming the optic nerve [12]. Glaucoma cannot be cured; it is impossible to restore lost vision, but it is feasible to prevent further vision loss with medication and/or surgery. According to a recent study titled "Prevalence of Glaucoma among US Adults in 2022," which was published in JAMA Ophthalmology, 4.22 million Americans suffer from glaucoma [13].

Recently, numerous state-of-the-art ML and DL models or solutions have been developed for the detection, segmentation, and classification of retinal disorders. In the field of fundus image diagnosis, various studies have been conducted by researchers. More studies focused on specific eye diseases and their stages of development than on classifying different types of eye diseases. Some studies have attempted to distinguish between a healthy eye fundus and one affected by a single disorder; i.e., the authors made a binary classification [14]-[22]. In health care, binary classification is commonly used to distinguish between healthy and ill patients. Another line of research explored fundus photographs to classify several eye diseases, i.e., multiclass classification. In Nawaz *et al.* [23], the authors proposed a CNN model for classifying retinal diseases into 32 classes. Another study [24] proposed a method for automated glaucoma assessment that is made via a classification approach. Utilizing pretrained or customized Deep CNN (DCNN) models for classifying three [25], four [26]-[32], or five [33], [34] eye disease categories. Regarding DR screening, a lot of work has been carried out by researchers for grading DR into three [20] or five [35]-[38] stages. A third group established methodologies for identifying multilabel fundus diseases, i.e., multilabel multiclass classification [39]-[42]. In practice, a fundus image is likely to contain many fundus disorders, making this kind of fundus image classification a more prevalent and valuable challenge. The fourth group focused on identification and/or segmentation of retinal blood vessels or optic disk/cup, which is vital for the

diagnosis and treatment of various eye diseases [43], [44]. This paper presents a hybrid approach for classifying four classes of eye diseases using multiclass classification. We used a benchmark multi-class dataset, eye disease classification (EDC), which contains 4,217 samples of 4 eye disease types. The proposed DL framework begins with the preprocessing step, including cropping the circular region of interest, normalization, augmentation, and resizing. Subsequently, a hybrid CNN model was designed by combining two pre-trained architectures to extract spatial features from the input images. Then, outputs from both models are concatenated to combine their learned representations, resulting in a more comprehensive feature set, i.e., feature fusion. A multihead attention (MHA) layer is applied to enhance important features by allowing the model to focus on key areas in the combined feature map. Finally, fully connected (FC) layers apply the classification.

The remainder of the research paper is organized as follows. The related works are presented in Section 2. Section 3 provides a detailed description of the specific steps and methods employed in the proposed system. Results are in Section 4, discussions are in Section 5, and Section 6 presents the conclusion.

2. RELATED WORKS

In the last few years, researchers have made significant advances in the field of fundus image processing for the identification and categorization of retinal abnormalities. To minimize ophthalmologists' burden and increase diagnosis consistency, automated eye disease diagnostics were investigated. These efforts were driven by the application of sophisticated ML and DL models. More specifically, the CNN architecture and its variations were widely proposed in publications for the classification of retinal diseases using fundus images on both public and commercial datasets. This section reviews relevant studies that employed various strategies and techniques to address this challenge. Rather than classifying many eye disorders separately, there are more researchers who concentrate on particular eye diseases, their developmental stages, and multilabel retinal disease classification. For example, using publicly accessible datasets and quality evaluation, a new dataset (MuReD) including 2208 samples for 20 classes was developed for multilabel EDC by Rodriguez *et al.* [40]. In addition, the authors refined a transformer-based model through significant experimentation that can detect and classify multiple retinal diseases.

Ouda *et al.* [42] classified multilabel ocular diseases from fundus images using the retinal fundus multi-disease image dataset that contains 45 types of eye abnormalities and 3200 samples. A framework named ML-CNN was proposed, whose general architecture has FC, pooling, and convolution layers. The model's performance was evaluated using K-fold cross-validation (CV) with 2, 5, and 10 values of K to confirm the outcome attained. The average and highest accuracy rate was 94.3, obtained by 10-fold CV. Kadum *et al.* [32] designed a hybrid feature extraction methodology based on color fundus images for EDC. The dataset used in this work was EDC [45]. Three techniques were applied for the feature extraction process, after which a single vector is created by combining the extracted features. The task of classification is performed using two classifier models, which are K-Nearest Neighbors and support vector machine (SVM). The SVM classifier can classify instances with a 99.88% accuracy rate.

The study by Guo *et al.* [33] proposed a pre-trained DL model called MobileNetV2 for extracting features of fundus images and transfer learning (TL) for classifying five common labels in the eye disease dataset. A portion of the dataset's classes were utilized; in total, there are only 250 samples in the dataset [46]. According to their findings, even with a very tiny quantity of data, MobileNetV2 can classify various eye disorders with noteworthy results thanks to TL. According to experimental findings, the system's average accuracy, sensitivity, and specificity on the test data are 96.2%, 90.4%, and 97.6%, respectively.

An automated cataract detection method, developed by Junayed *et al.* [47], named CataractNet. The proposed system is based on 4 blocks of deep neural network (DNN) to analyze fundus images; convolutional and max-pooling layers make up each of the model's four blocks, a 16-layer DL neural network in total. The utilized dataset comes from a number of standard fundus imaging datasets that have been released during the last 20 years, including HRF, FIRE, ACHIKO-I, IDRiD, and DRIVE. The CataractNet model demonstrated competitive performance when compared to five pretrained CNN models, obtaining an average of 99.13% in terms of accuracy.

A hybrid DL network model for DR diagnosis is presented in this research [48]. In the segmentation phase, an open-closed watershed management strategy is used to segment the blood vessels and the optic disc. In the last phase, i.e., the classification phase, the Binocular Siamese-like hybrid (AlexNet, GoogleNet, and SVM) model is introduced to recognize the normal and DR images. The accuracy achieved

by the suggested hybrid neural network on the DB0 and DB1 datasets was 94% and 94.83%, respectively.

Mahmoud *et al*. [21] introduced an automated hybrid inductive ML algorithm for DR screening, namely HIMLA. The model only categorizes fundus images as either healthy or unhealthy. The features are extracted and classified using multiple instance learning. The suggested method obtained a 96.62% accuracy rate, a 95.31% sensitivity rate, and a 96.88% specificity rate when evaluated on the CHASE dataset. This research, conducted by Thanki [17], presents a DNN- and ML-based system to evaluate retinal fundus images for glaucomatous classification. Several ML classifiers were employed to classify and assess the deep features of fundus images that have been extracted using a DNN. According to experimental findings, the combination of a logistic regression-based classifier and DNN works better together than current glaucomatous screening methods, with increasing sensitivity and accuracy.

Nawaz *et al*. [23] used a DL-based CNN model to classify retinal diseases by making effective use of memory. A benchmark dataset called Eye Net [49], which contains 32 kinds of retinal disorders, was used to assess the model. Authors mention that the designed model, compared to other methods, uses less memory consumption and yields superior results with a 95% accuracy rate.

In this study Vadduri and Kuppusamy [26], segmentation techniques like the Tyler Coye Algorithm, Otsu thresholding, and circular hough transform are used to extract important regions of interest such as the macular region, blood vessels, and optic nerve from the raw fundus images. In addition, the comparison was performed between the optimal result among four distinct pre-trained models (Xception, VGG-16, ResNet50, and EfficientNetB7) and the newly proposed DCNN architecture for multiclass classification of the fundus dataset. EfficientNetB7 achieved an accuracy of 91.39%, while the proposed model performed even better, reaching an accuracy of 96.94%.

The work presented in Albelaihi and Ibrahim [27] used a DL model, namely the DeepDiabetic framework, for multiclass classification of four types of retinal fundus images. The authors trained five models using three different methods of image data augmentation (non-augmented, online, and offline augmented images). The performance of the architectures was examined on 1228 samples from six distinct datasets, and the EfficientNetB0 model attained an accuracy of 98.76%, which performs better than the other four specified models.

Wahab Sait *et al*. [50] proposed a DL-based EDC model for multiclass classification of fundus images. The authors employed a single-shot detection technique for feature extraction purposes and the whale optimization algorithm with the Levy flight and wavelet mutation approaches for feature selection. Furthermore, for EDC, an optimized ShuffleNetV2 model was used. The proposed model was evaluated on two benchmark datasets, ocular disease intelligent recognition (ODIR) and EDC. The accuracy and Kappa values of the suggested EDC model were 99.1 and 96.4 in the ODIR dataset and 99.4 and 96.5 in the EDC dataset, respectively. Prasher *et al*. [51] offered two CNN TL models (MobileNetV3 and EfficientNetB0) for multiclass prediction of eye disorders. EDC, the open-source Kaggle repository, is where the dataset was acquired [45]. Using an Adam optimizer with 100 epochs, MobileNetV3 obtained 73% accuracy on 15 epochs, while EfficientNetB0 achieved 94% accuracy on the remaining epochs.

Babaqi *et al*. [30] presented a traditional CNN as well as a TL method based on a pre-trained EfficientNet model for detecting and classifying four eye disease categories. This study was also evaluated on the EDC dataset [45]. The proposed EfficientNet achieved a higher accuracy rate than the traditional CNN architecture, 94% and 84%, respectively. The authors demonstrate that while CNN by itself is insufficient for classifying eye conditions, TL greatly improves its effectiveness. Applying TL to the pre-trained models has made it feasible to train the model with fewer resources and assist in preventing the need to reinvent the wheel compared to CNN models [52].

To identify the most effective algorithm for classifying eye diseases, three distinct DL-based models were employed in this research [53], including VGG-16, VGG-19, and EfficientNetB0. Training was performed on the EDC dataset [45]. The highest accuracy rate was obtained with the EfficientNetB0 model before and after applying the normalization process on the fundus images.

Vardhan *et al*. [54] explored various pre-trained models, including Inception v3, VGG19, and ResNet50, on fundus images for multiclass classification on the Kaggle EDC dataset. All three models were evaluated on the dataset without TL and achieved a validation accuracy of 66.39%, 65.50%, and 57.04%, respectively. However, the three CNN models with TL have achieved higher validation accuracy of 87.69%, 92.56%, and 83.79%, respectively.

3. MATERIALS AND METHODS

This study implemented several neural network models to develop the most effective model for EDC. These models are referred to as pre-trained models. Due to their strong generalization capabilities and training on large benchmark datasets like ImageNet, they provide a valuable starting point for new tasks. We trained CNN architectures, including Xception, VGG-16/19, ResNet50/101/152, ResNet50/101/152V2, InceptionV3, InceptionResNetV2, MobileNetV1/V2, MobileNetV3Small/Large, DenseNet121/169/201, NASNet Mobile/Large, EfficientNetB0/B1/B7, and ConvNeXtTiny/Small/Base/Large, using the TL technique with weights learned from ImageNet. Among all the evaluated models, DenseNet169 and MobileNetV1 yielded better results. Therefore, we adopted a hybrid strategy that integrated these two models to design a framework for automatically classifying fundus disorders. In the proposed hybrid model, features were extracted independently by DenseNet169 and MobileNetV1. Global average pooling occurs after each backbone, and the feature vectors are concatenated to form a single unified representation. This concatenated feature vector is further transformed via a MHA layer, which helps the model focus on different positions of the fused representation and learn inter-feature relationships. The enhanced output through attention is ultimately processed through a FC layer and a softmax function for classification. Fig. 2 shows where

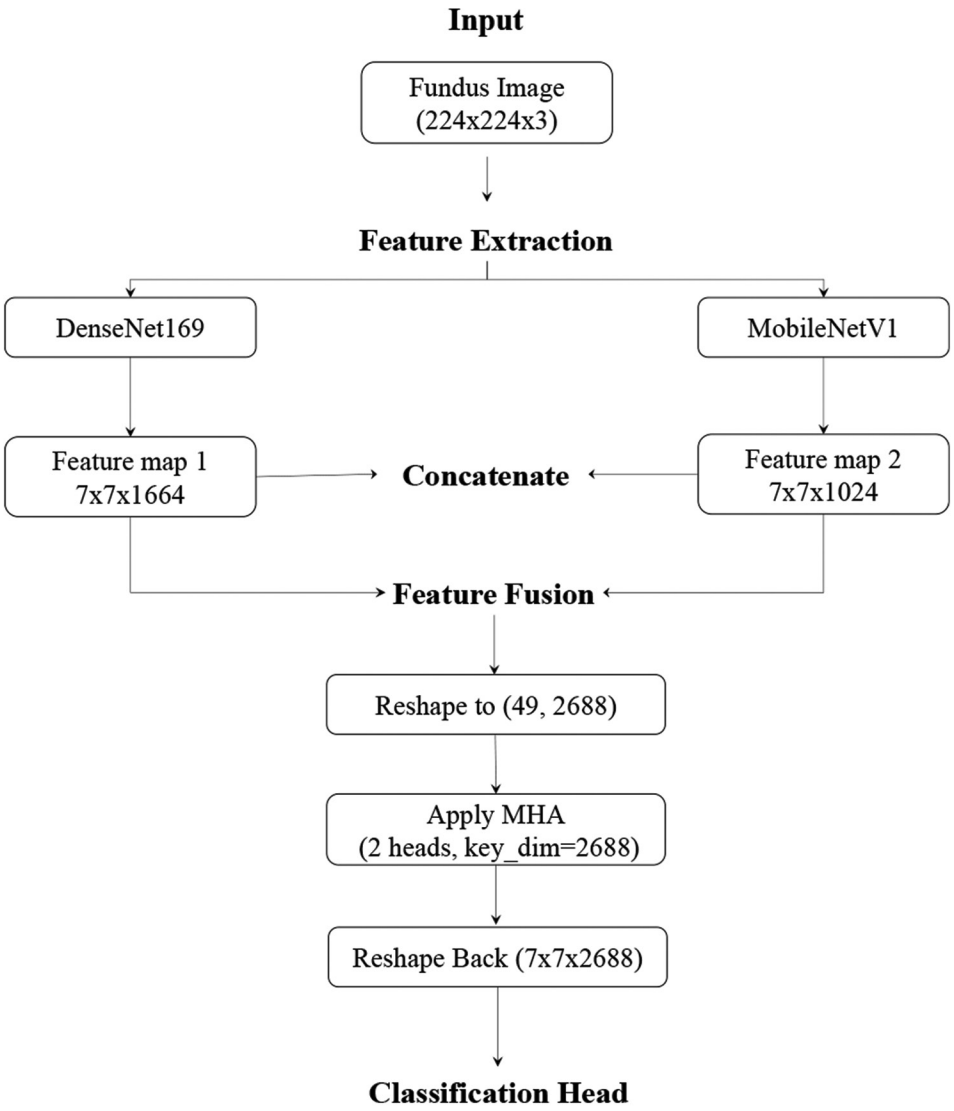


Fig. 2. Implementation of the feature fusion and attention mechanism in the proposed hybrid model.

and how features are fused and attention is applied. The methodology of the proposed system consists of three primary phases, which we explain in detail in this section. The workflow of the proposed method is presented in Fig. 3. It begins with a description of the Kaggle EDC dataset and its preparation for the preprocessing phase, followed by feeding the preprocessed images to the hybrid-TL models for feature extraction. Finally, dense layers were employed to classify the diseases.

3.1. Dataset Description

We obtained the eye_disease_classification dataset, or EDC, from the Kaggle source to evaluate the proposed system. The dataset holds 4,217 image samples for 3 types of eye diseases, including cataracts, glaucoma, and DR, as well as normal fundus cases. Fig. 1 shows one sample for each class in the dataset. The dataset is relatively balanced, with around 1,000 images in each class. These samples were gathered from a variety of sources, including HRF, Kaggle ODIR, IDRiD, and other datasets. The images in the dataset vary in resolution, but all of them have an RGB mode. Table 1 presents a detailed description of the dataset.

3.2. Dataset Pre-processing

Preprocessing describes the transformations performed on raw data to prepare the dataset before supplying it to an ML model. It is an essential phase in the ML process that can greatly improve results and is among the elements influencing ML's performance on a given task. The model's outcomes heavily depend on data diversity, quality, and quantity [55]. The process began with cropping the circular region of interest from the fundus images, specifically the retina, using a feature extraction technique known as the Hough circle transform (HCT), which is used to identify circles in an image. Although the original purpose of the Hough transform (HT) was to detect lines, it has evolved to recognize further analytical shapes, including circles and ellipses [56]. Before applying HCT, it was necessary to convert the images to grayscale and apply a Gaussian blur to eliminate noise, as excessive noise can introduce confusion and make

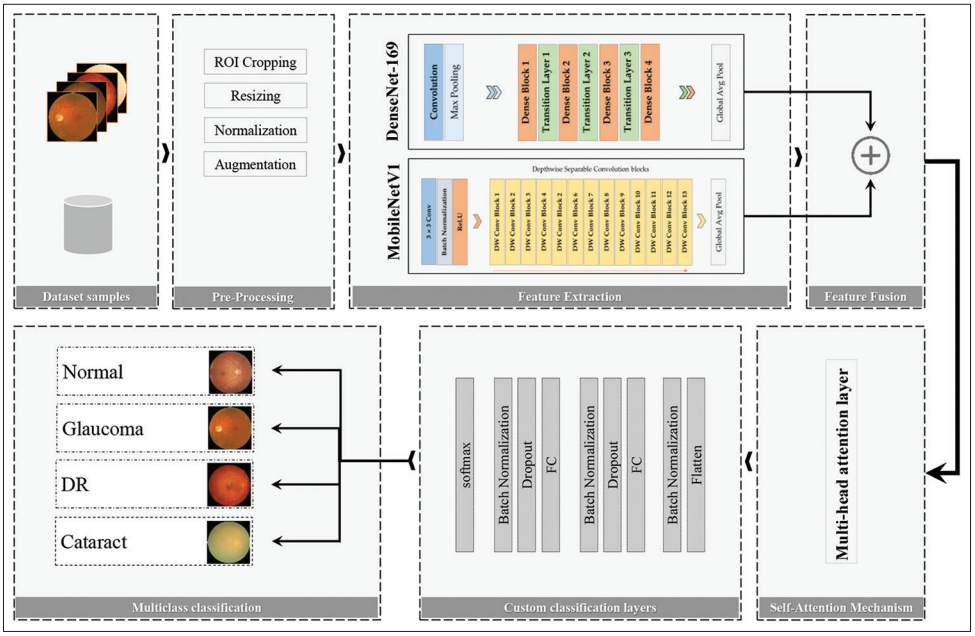


Fig. 3. Overall proposed system.

TABLE 1: Detailed information about eye disease classification dataset

Classes	Samples (percentage)	Image resolution distribution				
		256×256 (%)	512×512 (%)	1848×1224 (%)	2464×1632 (%)	2592×1728 (%)
Normal	1074 (25.5)		100			
DR	1098 (26.0)		100			
Glaucoma	1007 (23.9)	59.6	30.4	0.5	2.2	7.3
Cataract	1038 (24.6)	61.7	28.7	0.5	1.8	7.3
Total	4217 (100)	29.4	65.8	0.2	1.0	3.6

calculations difficult. HCT identifies a circle characterized by parameters (x, y, and r), where r denotes the radius and (x, y) represent the center of the fundus photographs. Finally, the fundus image was cropped based on the circle's bounding box. Originally, the images in the dataset varied in dimensions as they were sourced from multiple datasets and captured using different cameras.

In the second step, the images were resized to 224×224 pixels to standardize them. Resizing ensures uniformity in the input data and enhances feature extraction while also aiding uniformity in model learning. In the third step, the pixel values of the images were normalized before passing them to the feature extraction process. This standardizes or converts data to a common scale, typically ranging from [0–255] to [0–1]. This process ensures a comparable distribution for every input pixel, resulting in faster training convergence [47] and better performance. Augmentation is used to balance dataset samples across classes and expand the diversity of the dataset by applying various transformations. This study employed online augmentation, also known as on-the-fly augmentation, which performs a variety of dynamic transformations randomly supplied to the model during each training epoch. Since the transformations are applied dynamically rather than storing them in memory, online transformations are executed on batches of the dataset and then fed to the model during each iteration. This method boosts the diversity of the training data and helps the model generalize better to unseen data, as well as reduces the risk of overfitting. Table 2 lists the augmentation types and features employed in the proposed work. Fig. 4 presents multiple transformations of fundus images for each class to illustrate the effects of the augmentation technique.

3.3. Feature Extraction and Classification

The proposed feature extraction process employs a hybrid CNN model that combines the strengths of two CNN architectures, DenseNet169 and MobileNetV1, along with TL. DenseNet169 belongs to the Densely Connected Convolutional Network (DenseNet) family of models. It is a feed-forward CNN with a depth of 169 layers, and its parameters are relatively low compared to other models. DenseNet establishes direct connections between any two layers that share the same feature-map size, allowing all layers to access each other's features.

TABLE 2: A summary of the augmentation types with their features/ranges

Augmentation	Feature/range
Horizontal flip	100%
Vertical flip	100%
Rotation	0°, 90°, 180°, or 270°
Brightness adjustment	$\Delta \in [-0.1, 0.1]$
Contrast adjustment	Factor $\in$ [0.8, 1.2]
Zoom (Resize and Crop)	Resize to 234×234; Crop to 224×224
Pixel value clipping	[0,1]

Table 3 lists the configurations of DenseNet169 and MobileNetV1 on ImageNet. Their architectures can be seen in Figs. 5 and 6. For a detailed comparison of trainable and non-trainable parameters, as well as model sizes, including the proposed hybrid model, see Table 4. The Google team developed MobileNets, a simple and efficient CNN model for mobile vision applications, which utilizes depthwise separable convolution and performs well without requiring extensive computational resources. DenseNet is, in some ways, more complex than MobileNet due to its dense connectivity pattern, which necessitates more memory and computational power. Conversely, it simplifies feature learning by reusing features across layers, making it efficient in terms of parameter usage. Refer to Table 3, which illustrates the parameter and size comparison of both DenseNet169 and MobileNetV1 models, as well as the proposed hybrid model.

In our research, both models were initialized with pre-trained ImageNet weights, excluding the top classification layers. Both TL models were set up to pull out important details from images in the EDC dataset without changing their existing weights. The output shape of the extracted feature tensor for DenseNet169 and MobileNetV1 is (7, 7, 1664) and (7, 7, 1024), respectively. Subsequently, the outputs of both models were fused – feature fusion – to combine their learned representations and create a more comprehensive feature set. Feature fusion, in this context, means combining various visual cues or features that represent different aspects of visual characteristics, aiming to create a more comprehensive representation of features [57]. Thus, the extracted features from both models were concatenated along the channel axis to form a (7, 7, 2688) feature tensor.

Following that, we employ the self-attention mechanism using multiple attention heads. This mechanism (MHA) was originally introduced by Ashish Vaswani *et al.* [58]. It allows the model to jointly attend to information from different representation subspaces at different positions. This process improves the model's ability to identify different relationships within the features. In our work, an MHA layer was added to help the model simultaneously attend to different parts of the combined feature maps and further improve the feature representation. To make the concatenated feature tensor compatible with the MHA layer, it was reshaped into (49,

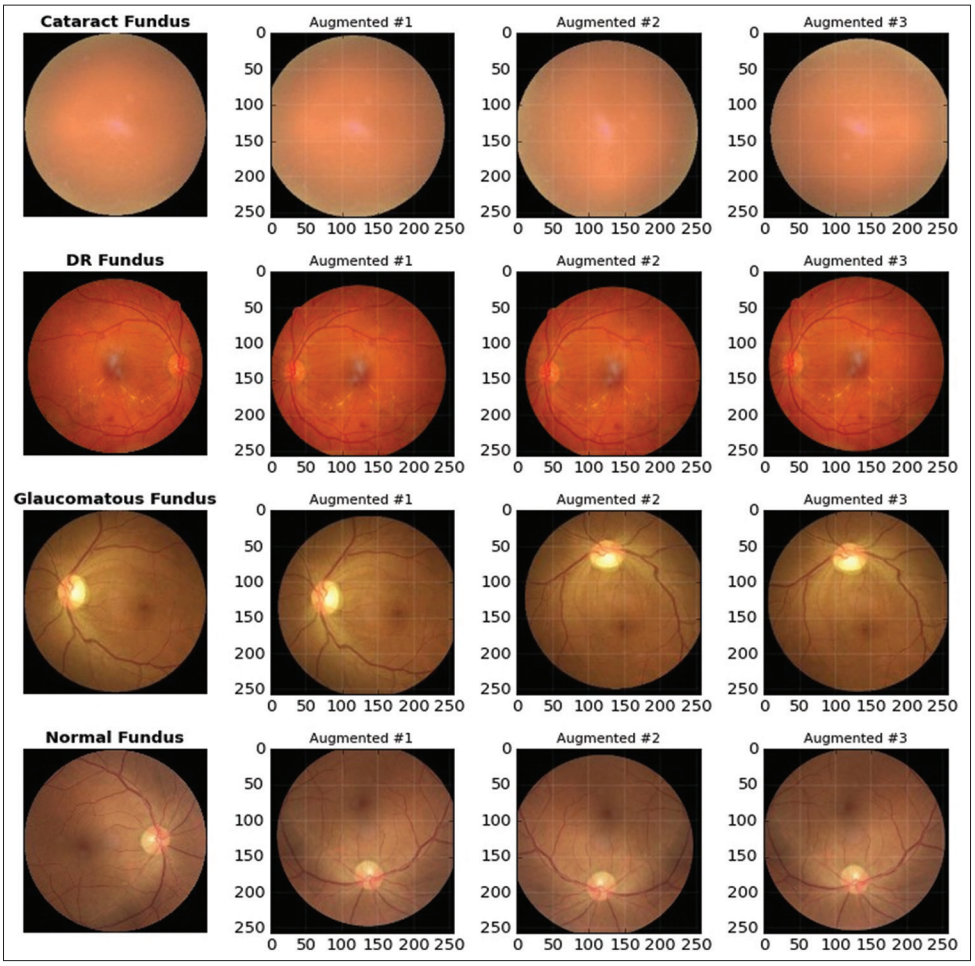


Fig. 4. Retinal fundus images with various augmented versions generated through random transformations.

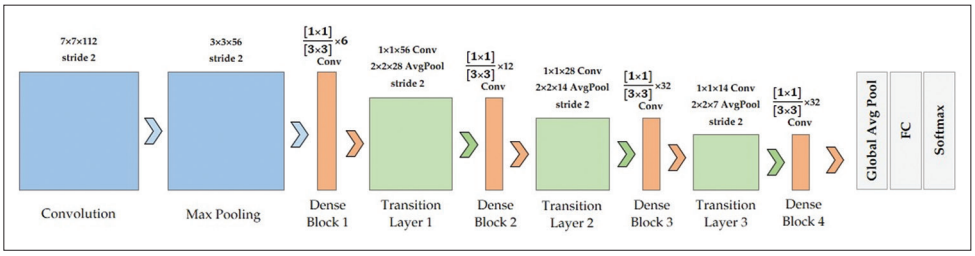


Fig. 5. DenseNet-169 architecture including layers.

2688), where "49" represents the flattened spatial dimensions of (7, 7) and "2688" denotes the total number of feature channels from both models. After applying the attention, the output was reshaped back to its original spatial dimensions.

After feature extraction and attention-based enhancement, the final step in the proposed system is to classify eye diseases by incorporating FC layers for multiclass classification. The output from the attention mechanism is passed through several layers, including dense, batch normalization, and dropout layers. The feature tensor was flattened to create a 1D vector. Two dense layers with 256 and 128 neurons and rectified linear unit (ReLU) activation were applied, followed by batch normalization and dropout layers. Finally, a dense layer with a softmax activation function was added to predict class probabilities based on the number of classes in the dataset.

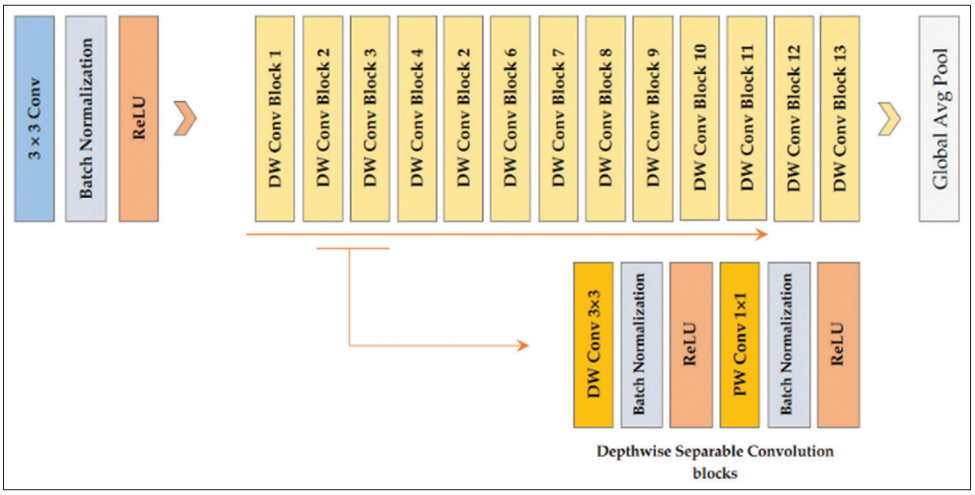


Fig. 6. MobileNetV1 architecture.

TABLE 3: DenseNet169 (left) and MobileNetV1 (right) architecture

Layers	Tensor size	Kernel size and parameters	Layers		Tensor size	Kernel size and parameters
Convolution	112×112	7×7 Conv, stride 2, ReLU	Convolution		112×112	3×3 × 3×32, stride 2
Pooling	56×56	3×3 Max Pool, stride 2	Block 1	DW Conv	112×112	3×3 × 32 DW, stride 1
Dense Block 1	56×56	[1×1, 3×3]×6 (Conv), Dropout 0.2		PW Conv	112×112	1×1 × 32×64, stride 1
Transition Layer 1	56×56	1×1 Conv	Block 2	DW Conv	56×56	3×3 × 64 DW, stride 2
	28×28	2×2 AvgPool, stride 2		PW Conv	56×56	1×1 × 64×128, stride 1
Dense Block 2	28×28	[1×1, 3×3]×12 (Conv), Dropout 0.2	Block 3	DW Conv	56×56	3×3 × 128 DW, stride 1
Transition Layer 2	28×28	1×1 Conv2×2 AvgPool, stride 2		PW Conv	56×56	1×1 × 128×128, stride 1
	14×14		Block 4	DW Conv	28×28	3×3 × 128 DW, stride 2
Dense Block 3	14×14	[1×1, 3×3]×32 (Conv), Dropout 0.2		PW Conv	28×28	1×1 × 128×256, stride 1
Transition Layer 3	14×14	1×1 Conv	Block 5	DW Conv	28×28	3×3 × 256 DW, stride 1
	7×7	2×2 AvgPool, stride 2		PW Conv	28×28	1×1 × 256×256, stride 1
Dense Block 4	7×7	[1×1, 3×3]×32 (Conv), Dropout 0.2	Block 6	DW Conv	14×14	3×3 × 256 DW, stride 2
Classifier	1×1	7×7 GlobalAvgPool		PW Conv	14×14	1×1 × 256×512, stride 1
	1000	1000D FC, Softmax	Block	DW Conv	14×14	3×3 × 512 DW, stride 1
			7–11	PW Conv	14×14	1×1 × 512×512, stride 1
			Block 12	DW Conv	7×7	3×3 × 512 DW, stride 2
				PW Conv	7×7	1×1 × 512×1024, stride 1
			Block 13	DW Conv	7×7	3×3 × 1024 DW, stride 2
				PW Conv	7×7	1×1 × 1024×1024, stride 1
			Classifier		1×1	7×7 GlobalAvgPool
					1000	1000D FC, Softmax

Conv: Convolution, ReLU: Rectified linear unit, DW: Depthwise, PW; Pointwise, FC: Fully connected

TABLE 4: Pre-trained models and proposed hybrid model parameters

Model	Trainable parameters	Non-trainable parameters	Total parameters	Model size
DenseNet169	0	12,642,880	12,642,880	48.23 MB
MobileNetV1	0	3,228,864	3,228,864	12.32 MB
Proposed (hybrid)	17,131,588	16,135,552	33,267,140	126.90 MB

3.4. Implementation Details

In our study, we experimented with both Adam and SGD optimizers. However, SGD delivered better results in enhancing the model's overall performance. The training process was configured to stop early if the validation accuracy did not improve after 5 iterations. Furthermore, other hyperparameter configurations that achieved the best accuracy for the proposed model are shown in Table 5. The

training process in the proposed framework was carried out using Tesla T4 GPU runtime on Google's cloud servers, "Colab notebook" using Python v3.10.12 and Keras API over TensorFlow v2.17.1. A personal laptop was used for all experiments with these specifications: Windows 10 Pro 64, Intel Core i7, 8 GB RAM, and NVIDIA GeForce GT 650 M.

To build the most accurate and effective system for identifying eye diseases, the fundus image dataset was exposed to various CNN architecture versions. Several models were trained along with hyperparameter tuning, and their results were recorded regardless of whether they performed poorly or positively. The CNN models used were the common TL pre-trained models on the ImageNet dataset. The best-performing models among these hyperparameter tuning models were integrated to construct a hybrid framework for extracting deep features from fundus images. Furthermore, adjustments were performed to the top layers by increasing and decreasing the number of layers and neurons, and experimenting with different learning rates, dropouts, and batch sizes was also done to achieve optimal outcomes.

TABLE 5: Hyperparameter configuration

Configuration	Specification
Input shape	(224, 224, 3)
Classifier	Softmax
Loss function	Sparse categorical crossentropy
Optimizer	SGD
Batch size	16
Learning rate	0.00001
Early stopping	Monitor='val_accuracy' Patience=5
Checkpoint	Monitor='val_accuracy' save_weights_only=True save_best_only=True
Total epochs	40

4. RESULTS

After preprocessing the dataset to prepare it for training and testing, we experimented with several data split ratios to divide the dataset into two subsets: Training and validation. These ratios included 70:30, 75:25, and 80:20, with the best results obtained using 75:25. In total, 3162 samples were used for training and 1055 samples for testing. Table 6 displays all the parameters used in the training process for each model.

Each pre-trained DL model shown in Table 6 was trained and evaluated separately with different split strategies, epochs, layers, neurons, and other parameters to accomplish the best results. Among all the pre-trained DL models, DenseNet169

TABLE 6: Pre-trained dl models with their parameters and details performance

Model	Learning rate	Batch size	dropout	Split	Epochs	ACC	V ACC
ResNet50	1e-2, 1e-3, 1e-4, 1e-5	8, 16, 32, 64	0.2, 0.5	80:20	40	71.25	79.27
ResNet50V2	1e-4, 1e-5	32	0.1, 0.2, 0.3	75:25	20	95.34	89.57
ResNet101	1e-3, 1e-4, 1e-5, 5e-6	32, 64, 128	0.2	80:20, 75:25	30	83.22	81.04
ResNet101V2	1e-3, 1e-4, 1e-5	32	0.2, 0.3	75:25	40	95.50	88.91
ResNet102	1e-5	64	-	75:25	25	74.00	72.00
ResNet102V2	1e-5	32	0.2	75:25	40	92.60	89.29
EfficientNetB0	1e-3	32	0.2	75:25	10	26.88	26.07
EfficientNetB1	1e-3, 1e-4	32	-	75:25	25	69.79	74.12
EfficientNetB7	1e-4	32	-	75:25	15	29.29	30.88
InceptionV3	1e-5, 5e-6	32	0.1, 0.2, 0.3	75:25	40	93.15	86.35
InceptionResNetV2	1e-3, 1e-4, 1e-5	32	0.1, 0.2, 0.3	75:25	30	97.29	89.48
Xception	1e-3, 1e-4, 1e-5	32	0.2	75:25	25	94.50	88.53
vgg16	1e-4, 1e-5, 5e-6	16, 32, 64	0.1, 0.2	75:25	40	84.25	89.29
vgg19	1e-5, 1e-6, 5e-6	16	0.1, 0.2	75:25	50	90.53	87.11
ConvNeXtTiny	1e-4, 1e-5	16, 32, 64	0.2	75:25	25	64.37	72.99
ConvNeXtSmall	1e-4, 5e-5	32	-	75:25	25	67.94	69.38
ConvNeXtBase	1e-3	16, 32	-	75:25	15	61.53	67.96
ConvNeXtLarge	1e-3	16, 32	-	75:25	25	65.43	69.19
DenseNet121	1e-4, 1e-5, 5e-6	32	0.1, 0.2	75:25	25	91.13	90.90
DenseNet169	1e-4, 1e-5, 5e-6, 5e-5	32, 64	0.1, 0.2, 0.3, 0.4	75:25	25	92.44	91.09
DenseNet201	1e-4, 1e-5	16	0.1, 0.2, 0.3	75:25	25	89.67	91.00
MobileNetV1	1e-3, 1e-4, 5e-4, 1e-5	16, 32, 64	0.1, 0.2, 0.3, 0.5	75:25	15	97.24	91.09
MobileNetV2	1e-3, 1e-4, 5e-4, 1e-5	32	0.1, 0.2, 0.3, 0.5	75:25	25	98.10	90.81
MobileNetV3Small	1e-3, 1e-4, 1e-5, 5e-5	8, 32	-	75:25	45	73.59	76.02
MobileNetV3Large	1e-3, 1e-4, 5e-5	8, 32	-	75:25	40	77.54	79.62
NASNetMobile	1e-3, 1e-4, 1e-5, 5e-5	32	0.1, 0.2, 0.3	75:25	20	94.52	90.33
NASNetLarge	1e-3, 1e-4, 1e-5, 5e-5	32	0.2	75:25	20	97.47	88.15

obtained 92.44 and 91.09 for training and validation accuracy, and MobileNetV1 obtained 97.24 and 91.09, respectively. Both outperformed the other competitive models in terms of classification performance. Various learning rate and batch size values were used along with each single model in the training process, as shown in Table 6. Dropout layers were applied as needed in each custom top layer, with rates of 0.1, 0.2, 0.3, 0.4, and 0.5. The Adam optimizer was used to train the pre-trained models, and for the hybrid model, we have used both Adam and SGD.

Fig. 7 demonstrates the performance of the best-performing models within each group. For example, we only illustrated the line chart of ResNet50V2 among all other versions of ResNets, as it achieved the highest accuracy.

We observed that a hybrid CNN model, designed by combining the DenseNet169 and MobileNetV1 architectures for extracting deep features from a fundus dataset, achieved a higher accuracy rate than other individual pre-trained models. The training and testing were performed repeatedly on the dataset with hyperparameter tuning. The experiments were carried out using batch sizes of 16 and 32, whereas also evaluating the performance of the hybrid model using both Adam and SGD optimizers.

First, we experimented with Adam optimization. After running several tests, we assessed the hybrid model by adding two FC layers with ReLU and the last layer using the Softmax activation function. The accuracy of 97.28% and 92.99% for training and validation was achieved in the 17th epoch. See Fig. 8a1 and a2. The second experiment was performed with the SGD optimizer. The evaluation was performed using the self-attention mechanism and FC layers. The accuracy of

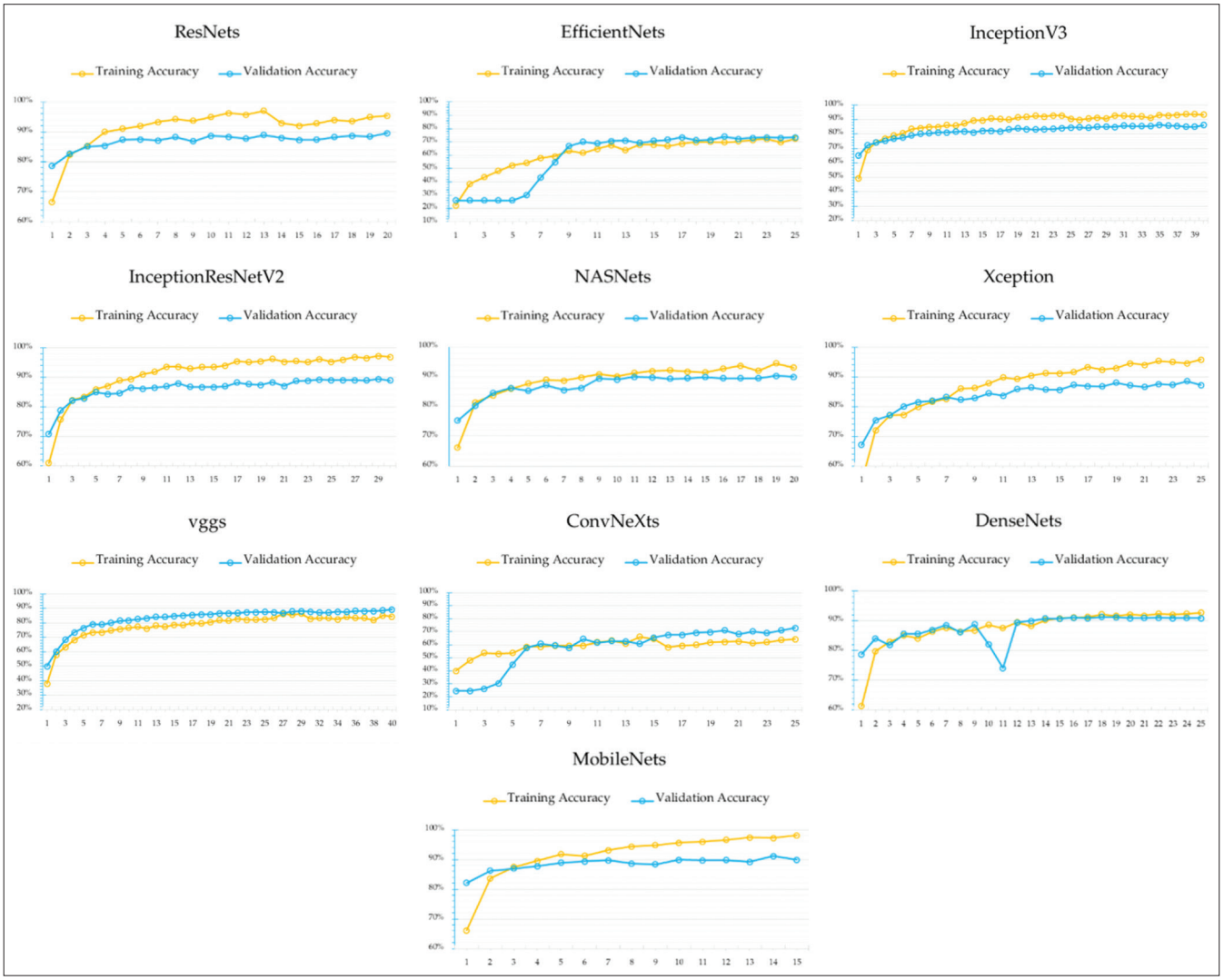


Fig. 7. Different transfer learning model's performance.

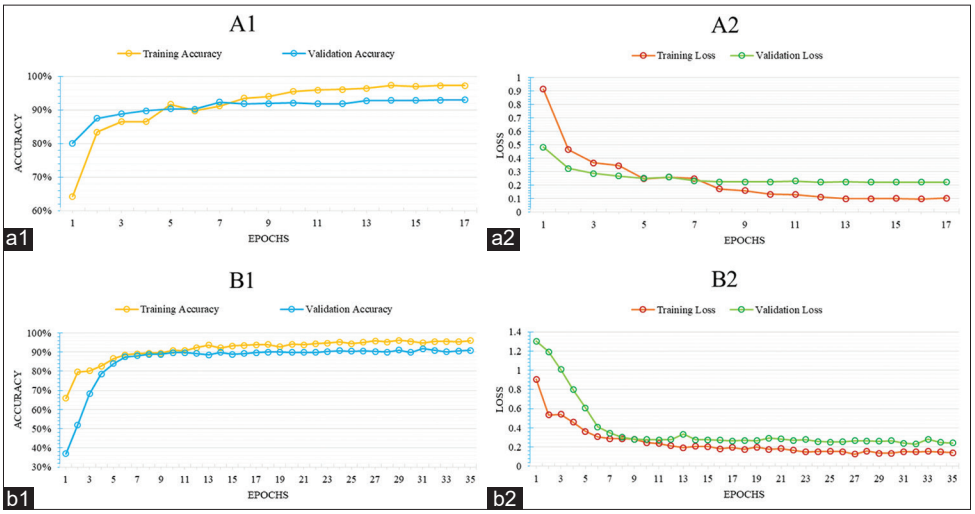


Fig. 8. Accuracy and loss of the proposed hybrid model.

97.05% and 91.75% for training and validation was achieved in the 31st epoch. See Fig. 8b1 and b2.

We noticed that the training/validation accuracy and loss curves closely overlapped when using the Adam optimizer to test the model, as illustrated in Fig. 8a1 and a2. Therefore, we decided to consider another approach to address this problem. We applied offline data augmentation to upsample the dataset to 3000 samples to balance class distributions, i.e., 1250 samples/class. We also experimented with the 70:30 ratio for data split, and the training and validation were conducted multiple times with different hyperparameters. Eventually, an efficient solution to the issue was to introduce a MHA layer and switch from the Adam to the SGD optimizer. In this way, the overlapping of curves is removed in the performance graphs (Fig. 8b1 and b2). The proposed method performed better when compared to individual TL models, with average improvements of 1.9% in terms of validation accuracy.

The proposed model's performance was evaluated using four quantitative performance metrics: accuracy, precision, recall (also called the true positive rate or sensitivity), F1-score, and area under the curve (AUC). The mathematical expression for these metrics is illustrated from Equations (1) to (4).

$$Accuracy = \frac{Tp + TN}{Tp + TN + FP + FN} \tag{1}$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - score = 2\mathrm{x}\frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

We have included a confusion matrix (Fig. 9a) that illustrates how our model performs across four categories: Cataract, glaucoma, normal, and DR. The model performs exceptionally well on DR and normal cases. It accurately identifies all DR samples (100%) and the majority of normal ones (92%). In addition, cataract cases were identified with 94% accuracy, with a small percentage misclassified as glaucoma or normal. The model demonstrates strong performance, particularly in identifying DR and normal instances, which are crucial for early diagnosis and intervention. The receiver operating characteristic curve shown in Fig. 9b clearly emphasizes how well our classification model works across different eye conditions. AUC values exhibit exceptional diagnostic efficacy, with both cataract and DR attaining a perfect AUC of 1.0, signifying impeccable differentiation between positive and negative cases. Normal fundus images scored 0.98, while glaucoma registered at 0.97 – still very good, just slightly less than perfect. Overall, these high AUC values indicate that the model is highly capable of distinguishing between the four classes and suggest it can generalize well to new data.

5. DISCUSSION

The findings from the preceding section indicate that the hybrid version of the proposed system consistently

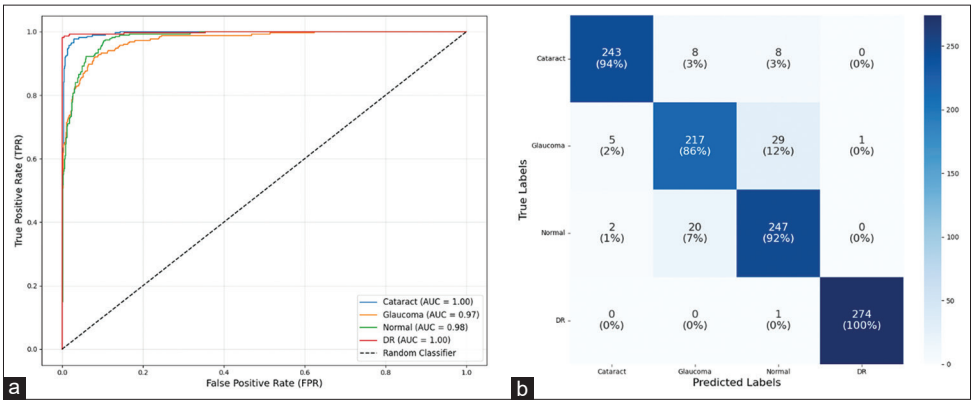
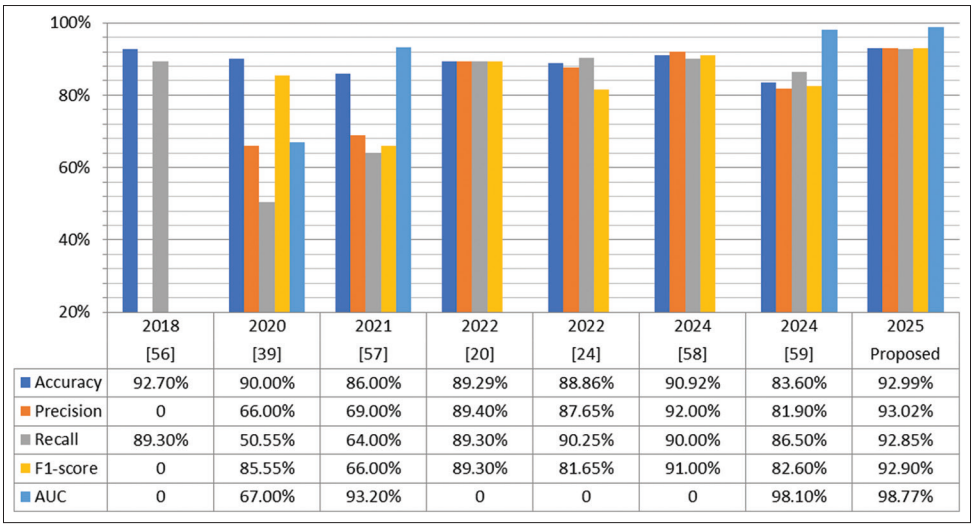


Fig. 9. Receiver operating characteristic curve (a) and confusion matrix (b) produced by the proposed model.



	2018 [56]	2020 [39]	2021 [57]	2022 [20]	2022 [24]	2024 [58]	2024 [59]	2025 Proposed
Accuracy	92.70%	90.00%	86.00%	89.29%	88.86%	90.92%	83.60%	92.99%
Precision	0	66.00%	69.00%	89.40%	87.65%	92.00%	81.90%	93.02%
Recall	89.30%	50.55%	64.00%	89.30%	90.25%	90.00%	86.50%	92.85%
F1-score	0	85.55%	66.00%	89.30%	81.65%	91.00%	82.60%	92.90%
AUC	0	67.00%	93.20%	0	0	0	98.10%	98.77%

Fig. 10. Comparison between the proposed approach and current methods.

outperforms the baseline models. This enhancement can be attributed to the complementary feature extraction capabilities of both architectures. As shown in Figs. 7 and 8, the DenseNet169-MobileNetV1 hybrid model surpasses that of individual pre-trained models in terms of classification accuracy.

Various CNN frameworks, namely, DenseNet169, MobileNetV1, EfficientNetB0, and InceptionResNetV2, were tested in this research based on fundus image-based disease classification. DenseNet169 was selected because it allows for reusing deep features and flowing gradients through densely connected layers, which is crucial in capturing fine-grained vascular and structural patterns in retinal images. MobileNetV1, by contrast, offers computational efficiency due to its depth wise separable convolutions, which can run in real-time on even resource-constrained devices. The two models had high individual performance on training and validation. Their architectural complementarity (depth of DenseNet and efficiency of MobileNet) became the driving force behind their fusion, which led to the creation of a hybrid model that demonstrates high accuracy and, at the same time, has computational feasibility. This suggested that hybrid DL techniques could successfully challenge classification problems because they had higher feature extraction capabilities, especially in multiclass disease classification scenarios.

Several recent studies have explored DL- and ML-based approaches for screening retinal diseases. For instance, Luo *et al.* [59] developed a DCNN-based DR detection methodology based on the Google Inceptionv3 network, achieving an accuracy of 83.60%. Nawaldgi and Lalitha [24] proposed a method for staging the severity of glaucoma. They extracted structural and texture features and used ML classifiers for classification, obtaining only 88.86% accuracy.

The authors of Butt *et al.* [20] achieved the maximum average accuracy of 89.29% by utilizing different ML classifiers for multiclass DR detection. Similarly, Vardhan *et al.* [54] employed three CNN models with TL approaches; the maximum accuracy reported was 92.56%, achieved by the VGG19 model. For classifying all five stages of DR using fundus images, a multitasking DNN based on the DenseNet architecture has been developed by Majumder and Kehtarnavaz [60], obtaining an accuracy of 86%. Wang *et al.* [39] utilized EfficientNet as a feature extractor and achieved 90% accuracy. Liu *et al.* [61] detected glaucomatous discs from retinal images with performance comparable to that of human experts by adopting the ResNet50 architecture with an accuracy of 92.7%. In contrast to previous works that relied solely on single-model architectures, our approach integrates multiple networks and surpasses these benchmarks by reaching 92.99% accuracy and 98.77% AUC. The superior performance suggests that combining multiple CNN architectures (like DenseNet169 and MobileNetV1) can significantly enhance feature learning and classification accuracy.

A comparative analysis (Fig. 10) further supports our findings, where our hybrid approach outperforms state-of-the-art methods across multiple metrics, such as accuracy (92.99%), precision (93.02%), recall (92.85%), F1-score (92.90%), and AUC (98.77%). These results emphasize the potential of hybrid DL models in assisting ophthalmologists with more accurate and automated diagnoses.

6. CONCLUSION

This study was conducted to classify four categories of eye diseases using most of the pre-trained DL models and to develop a framework based on the top two models that performed optimally. The EDC dataset was used to evaluate the proposed system, achieving 92.99%, 93.02%, 92.85%, 92.90%, and 98.77% for accuracy, precision, recall, F1-score, and AUC, respectively. The methodology of our framework began with a preprocessing step that includes cropping the circular region of interest, normalization, augmentation, and resizing. Subsequently, a hybrid CNN model was developed by combining the two selected architectures to extract spatial features from the dataset. Finally, after experimenting with various configurations of FC layers and neurons, the top customized classification layers were finalized for the final prediction. This study presents a methodology for detecting multi-class eye diseases, an area that previous research has not fully explored. To further validate the model's performance, we performed a comparative analysis. Our method surpasses a series of pre-trained models and findings from other studies. Our approach to classifying eye diseases based on fundus images shows significant promise for early screening and diagnosis, and we believe it could help reduce healthcare costs and streamline the eye diagnosis process. In the future, we aim to expand our research to address a wider range of eye diseases by utilizing newly available datasets or those we plan to collect independently. In addition, we plan to develop a real-time application to assist healthcare centers in the early detection and diagnosis of retinal diseases. By integrating advanced DL techniques with real-world clinical settings, we hope to enhance accessibility and efficiency in ophthalmic care.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support for this study from the Ministry of Higher Education and Scientific Research-Kurdistan Regional Government, Department of Computer, College of Science, University of Sulaimani. We sincerely appreciate the time and effort of the reviewers in evaluating our work and providing valuable insight.

REFERENCES

[1] WHO. "*Eye Care, Vision Impairment and Blindness*". Available from: https://www.who.int/health-topics/blindness-and-vision-loss#tab=tab_1 [Last accessed on 2024 Dec 07].

[2] S. Panchal, A. Naik, M. Kokare, S. Pachade, R. Naigaonkar, P. Phadnis and A. Bhange. "Retinal fundus multi-disease image dataset (RFMiD) 2.0: A dataset of frequently and rarely identified diseases". *Data,* vol. 8, no. 2, p. 29, 2023.

[3] X. Xia, Y. Li, G. Xiao, K. Zhan, J. Yan, C. Cai, Y. Fang and G. Huang. "Benchmarking deep models on retinal fundus disease diagnosis and a large-scale dataset". *Signal Processing: Image Communication,* vol. 127, p. 117151, 2024.

[4] F. Du, H. Luo, Q. Xing, J. Wu, Y. Zhu, W. Xu, W. He and J. Wu. "Recognition of eye diseases based on deep neural networks for transfer learning and improved D-S evidence theory". *BMC Medical Imaging*, vol. 24, no. 1, p. 19, 2024.

[5] N. Gour and P. Khanna. "Multi-class multi-label ophthalmological disease detection using transfer learning based convolutional neural network". *Biomedical Signal Processing and Control,* vol. 66, p. 102329, 2021.

[6] M. J. Burton, J. Ramke, A. P. Marques, R. R. A. Bourne, N. Congdon, I. Jones, B. A. M. Tong, S. Arunga, D. Bachani, C. Bascaran,… & A. Bastawrous. "The lancet global health commission on global eye health: Vision beyond 2020". *The Lancet Global Health*, vol. 9, no. 4, pp. e489-e551, 2021.

[7] National Institutes of Health. "*Diabetic Retinopathy*". Available from: https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/diabetic-retinopathy [Last accessed on 2024 Dec 07].

[8] T. Y. Wong, J. Sun, R. Kawasaki, P. Ruamviboonsuk, N. Gupta, V. C. Lansingh and M. Maia. "Guidelines on diabetic eye care: The international council of ophthalmology recommendations for screening, follow-up, referral, and treatment based on resource settings". *Ophthalmology*, vol. 125, no. 10, pp. 1608-1622, 2018.

[9] M. R. K. Mookiah, U. R. Acharya, C. K. Chua, C. M. Lim, E. Y. K. Ng and A. Laude. "Computer-aided diagnosis of diabetic retinopathy: A review". *Computers in Biology and Medicine*, vol. 43, no. 12, pp. 2136-2155, 2013.

[10] M. Z. Atwany, A. H. Sahyoun and M. Yaqub. "Deep learning techniques for diabetic retinopathy classification: A survey". *IEEE Access*, vol. 10, pp. 28642-28655, 2022.

[11] NCCDPHP. "*Vision ans Eye Health*". Available from: https://www.cdc.gov/vision-health/about-eye-disorders/glaucoma.html [Last accessed on 2024 Dec 08].

[12] K. Boyd. "*Understanding Glaucoma: Symptoms, causes, Diagnosis, Treatment*". Available from: https://www.aao.org/eye-health/diseases/what-is-glaucoma [Last accessed on 2024 Dec 08].

[13] P. Blindness. "*New Study Finds Higher Prevalence of Glaucoma Than Previously Estimated*". Available from: https://preventblindness.org/new-glaucoma-prevalence-study [Last accessed on 2024 Dec 08].

[14] M. S. M. Khan, M. Ahmed, R. Z. Rasel and M. M. Khan. "Cataract Detection Using Convolutional Neural Network with VGG-19 Model". In: *2021 IEEE World AI IoT Congress (AIIoT)*, pp. 209-212, 2021.

[15] M. R. Hossain, S. Afroze, N. Siddique and M. M. Hoque. "Automatic Detection of Eye Cataract using Deep Convolution Neural Networks (DCNNs)". In: *2020 IEEE Region 10 Symposium (TENSYMP)*, pp. 1333-1338, 2020.

[16] M. S. Khan, N. Tafshir, K. N. Alam, A. R. Dhruba, M. M. Khan, A. A. Albraikan and F. A. Almalki. "Retracted: Deep learning for ocular disease recognition: An inner-class balance". *Computational Intelligence and Neuroscience*, vol. 2023, no. 1, p. 9838475, 2023.

[17] R. Thanki. "A deep neural network and machine learning approach for retinal fundus image classification". *Healthcare Analytics,* vol. 3, p. 100140.

[18] M. ur-Rehman, S. H. Khan, Z. Abbas and S. M. D. Rizvi. "Classification of Diabetic Retinopathy Images Based on Customised CNN Architecture". In: *2019 Amity International Conference on Artificial Intelligence (AICAI)*, pp. 244-248, 2019.

[19] M. A. Syarifah, A. Bustamam and P. P. Tampubolon. "Cataract classification based on fundus image using an optimized convolution neural network with lookahead optimizer". In: *AIP Conference Proceedings*, vol. 2296, no. 1, 2020, p. 020034.

[20] M. M. Butt, D. A. Iskandar, S. E. Abdelhamid, G. Latif and R. Alghazo. "Diabetic retinopathy detection from fundus images of the eye using hybrid deep learning features". *Diagnostics,* vol. 12, no. 7, p. 1607, 2022.

[21] M. H. Mahmoud, S. Alamery, H. Fouad, A. Altinawi and A. E. Youssef. "An automatic detection system of diabetic retinopathy using a hybrid inductive machine learning algorithm". *Personal and Ubiquitous Computing,* vol. 27, no. 3, pp. 751-765, 2023.

[22] A. Aljohani and R. Y. Aburasain. "A hybrid framework for glaucoma detection through federated machine learning and deep learning models". *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, p. 115, 2024.

[23] A. Nawaz, T. Ali, G. Mustafa, M. Babar and B. Qureshi. "Multi-class retinal diseases detection using deep CNN with minimal memory consumption". *IEEE Access*, vol. 11, pp. 56170-56180, 2023.

[24] S. Nawaldgi and Y. S. Lalitha. "Automated glaucoma assessment from color fundus images using structural and texture features". *Biomedical Signal Processing and Control*, vol. 77, p. 103875, 2022.

[25] J. Verma, I. Kansal, R. Popli, V. Khullar, D. Singh, M. Snehi, and R. Kumar,"A hybrid images deep trained feature extraction and ensemble learning models for classification of multi-disease in fundus images," in *Proc. Nordic Conf. on Digital Health and Wireless Solutions,* Oulu, Finland, May 7–8, 2024, pp. 203–221. Cham, Switzerland: Springer Nature, 2024.

[26] M. Vadduri and P. Kuppusamy. "Enhancing ocular healthcare: Deep learning-based multi-class diabetic eye disease segmentation and classification". *IEEE Access,* vol. 11, pp. 137881-137898, 2023.

[27] A. Albelaihi and D. M. Ibrahim. "DeepDiabetic: An identification system of diabetic eye diseases using deep neural networks". *IEEE Access*, vol. 12, pp. 10769-10789, 2024.

[28] A. Vanita Sharon and G. Saranya. "Classification of multi-retinal disease based on retinal fundus image using convolutional neural network". In: S. Smys, A. M. Iliyasu, R. Bestak and F. Shi, Eds. *New Trends in Computational Vision and Bio-inspired Computing: Selected Works Presented at the ICCVBIC 2018, Coimbatore, India*. Springer International Publishing, Cham, pp. 1009-1016, 2020.

[29] N. Chea and Y. Nam. "Classification of fundus images based on deep learning for detecting eye diseases". *Computers, Materials and Continua*, vol. 67, no. 1, pp. 411-426, 2020.

[30] T. Babaqi, M. Jaradat, A. E. Yildirim, S. H. Al-Nimer and D. Won. "*Eye Disease Classification using Deep Learning Techniques*". [arXiv Preprint], 2023.

[31] A. Shamsan, E. M. Senan and H. S. Shatnawi. "Automatic classification of colour fundus images for prediction eye disease types based on hybrid features". *Diagnostics*, vol. 13, no. 10, p. 1706, 2023.

[32] S. A. Kadum, F. H. Najjar, H. M. Al-Jawahry and F. Mohamed. "Eye Diseases Classification Based on Hybrid Feature Extraction Methods". In: *2023 6th International Conference on Engineering Technology and its Applications (IICETA)*, pp. 402-407, 2023.

[33] C. Guo, M. Yu and J. Li. "Prediction of different eye diseases based on fundus photography via deep transfer learning". *Journal of Clinical Medicine*, vol. 10, no. 23, p. 5481, 2021.

[34] R. Sarki, K. Ahmed, H. Wang, Y. Zhang and K. N. Wang. "Convolutional neural network for multi-class classification of diabetic eye disease". *EAI Endorsed Transactions on Scalable Information Systems*, vol. 9, p. 5, 2018.

[35] H. Xu, X. Shao, D. Fang and F. Huang. "A hybrid neural network approach for classifying diabetic retinopathy subtypes". *Frontiers in Medicine (Lausanne)*, vol. 10, p. 1293019, 2023.

[36] U. Ishtiaq, E. R. Abdullah and Z. Ishtiaque. "A hybrid technique for diabetic retinopathy detection based on ensemble-optimized CNN and texture features". *Diagnostics*, vol. 13, no. 10, p. 1816.

[37] A. Ali Tabtaba and O. Ata. "Diabetic retinopathy detection using developed hybrid cascaded multi-scale DCNN with hybrid heuristic strategy". *Biomedical Signal Processing and Control*, vol. 89, p. 105718, 2024.

[38] B. Menaouer, Z. Dermane, N. El Houda Kebir and N. Matta. "Diabetic retinopathy classification using hybrid deep learning approach". *SN Computer Science*, vol. 3, no. 5, p. 357, 2022.

[39] J. Wang, L. Yang, Z. Huo, W. He and J. Luo. "Multi-label classification of fundus images with efficientNet". *IEEE Access*, vol. 8, pp. 212499-212508, 2020.

[40] M. A. Rodriguez, H. AlMarzouqi and P. Liatsis. "Multi-label retinal

disease classification using transformers". *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 6, pp. 2739-2750.

[41] N. M. Dipu, S. A. Shohan and K. Salam. "Ocular disease detection using advanced neural network based classification algorithms". *Asian Journal For Convergence in Technology (AJCT)*, vol. 7, no. 2, pp. 91-99, 2021.

[42] O. Ouda, E. AbdelMaksoud, A. A. Abd El-Aziz, and M. Elmogy. "Multiple ocular disease diagnosis using fundus images based on multi-label deep learning classification". *Electronics*, vol. 11, no. 13, pp. 1966.

[43] S. Moccia, E. De Momi, S. El Hadji and L. S. Mattos. "Blood vessel segmentation algorithms - Review of methods, datasets and evaluation metrics". *Computer Methods and Programs in Biomedicine*, vol. 158, pp. 71-91, 2018.

[44] A. Almazroa, R. Burman, K. Raahemifar and V. Lakshminarayanan. "Optic disc and optic cup segmentation methodologies for glaucoma image detection: A survey". *Journal of Ophthalmology*, vol. 2015, p. 180972, 2015.

[45] V. Doddi. "*Kaggle*". Available from: https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification [Last accessed on 2024 Dec12].

[46] L. P. Cen, J. Ji, J. W. Lin, S. T. Ju, H. J. Lin, T. P. Li, Y. Wang,… & J. Feng. "Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks". *Nature Communications*, vol. 12, no. 1, p. 4828, 2021.

[47] M. S. Junayed, M. B. Islam, A. Sadeghzadeh and S. Rahman. "CataractNet: An automated cataract detection system using deep learning for fundus images". *IEEE Access*, vol. 9, pp. 128799-128808, 2021.

[48] B. Venkaiahppalaswamy, P. Prasad Reddy and S. Batha. "Hybrid deep learning approaches for the detection of diabetic retinopathy using optimized wavelet based model". *Biomedical Signal Processing and Control*, vol. 79, p. 104146, 2023.

[49] Yang. "*EyeNet*". Available from: https://github.com/huckiyang/eyenet [Last accessed on 2024 Dec 14].

[50] A. R. Wahab Sait. "Artificial intelligence-driven eye disease classification model". *Applied Sciences*, vol. 13, no. 20, p. 11437.

[51] S. Prasher, L. Nelson and S. Gomathi. "Automated Eye Disease Classification using MobileNetV3 and Efficientnetb0 Models using Transfer Learning". In: *2023 World Conference on Communication and Computing (WCONF)*, pp. 1-5, 2023.

[52] C. Desai. "Image classification using transfer learning and deep learning". *International Journal of Engineering and Computer Science*, vol. 10, no. 9, pp. 25394-25398.

[53] B. Şener and E. Sümer. "Classification of Eye Disease from Retinal Images Using Deep Learning". In: *2023 14th International Conference on Electrical and Electronics Engineering (ELECO)*, 2023, pp. 1-4.

[54] K. B. Vardhan, M. Nidhish, C. S. Kiran, N. S. Dudekula, S. C. Varanasi and R. M. Bhavadharini. "Eye disease detection using deep learning models with transfer learning techniques". *EAI Endorsed Transactions on Scalable Information Systems*, vol. 12, pp, 1-13, 2024.

[55] K. Maharana, S. Mondal and B. Nemade. "A review: Data pre-processing and data augmentation techniques". *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91-99, 2022.

[56] S. M. A. S. Hassanein, M. Sameer and M. E. Ragab. "A survey on hough transform, theory, techniques and applications". *International Journal of Computer Science Issues*, vol. 12, no. 1, pp. 139-156, 2015.

[57] G. Cheng and J. Han. "A survey on object detection in optical remote sensing images". *Journal of Photogrammetry and Remote Sensing*, vol. 117, pp. 11-28, 2016.

[58] N. S. Ashish Vaswani, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser and I. Polosukhin. "Attention is All you Need". In: *Advances in Neural Information Processing Systems Conferences (NeurIPS)*, pp. 5998-6008, 2023.

[59] T. Zhou, X. Ye, H. Lu, X. Zheng, S. Qiu and Y. Liu. "Dense convolutional network and its application in medical image analysis". *BioMed Research International*, vol. 2022, no. 1, p. 2384830, 2022.

[60] G. Huang, Z. Liu, L. V. D. Maaten and K. Q. Weinberger. "Densely Connected Convolutional Networks". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261-2269, 2017.

[61] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto and H. Adam. "*Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications*". [arXiv Preprint], 2017.

[62] X. Luo, W. Wang, Y. Xu, Z. Lai, X. Jin, B. Zhang and D. Zhang. "A deep convolutional neural network for diabetic retinopathy detection via mining local and long-range dependence". *CAAI Transactions on Intelligence Technology*, vol. 9, no. 1, pp. 153-166, 2024.

[63] S. Majumder and N. Kehtarnavaz. "Multitasking deep learning model for detection of five stages of diabetic retinopathy". *IEEE Access*, vol. 9, pp. 123220-123230, 2021.

[64] S. Liu, A. Schulz, M. Kalloniatis, B. Zangerl, W. Cai, Y. Gao, B. Chua, H. Arvind, J. Grigg, D. Chu, A. Klistorner and Y. You. "A deep learning-based algorithm identifies glaucomatous discs using monoscopic fundus photographs". *Ophthalmology Glaucoma*, vol. 1, no. 1, pp. 15-22, 2018.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Enhanced Integer-Based Homomorphic Encryption Scheme with Windowing Mechanism for Securing Electronic Health Records

Abdulrahman Tawfeeq Jalal, Mohammed Anwar Mohammed

Department of Computer Science, College of Science, University of Sulaimani, Sulaymaniyah, Iraq

ABSTRACT

The frequent breaches of healthcare data annually make robust encryption mechanisms crucial, especially those that preserve the usefulness of the data while ensuring privacy. This study addresses specific integer-based homomorphic encryption systems and their critical vulnerabilities. The vulnerability identified in these systems is the possibility of decryption using other values, such as factors or primes, instead of the claimed unique secret key. We propose an enhanced cryptographic formula to address this vulnerability using a double random value technique that ensures decryption depends solely on the designated secret key. We also apply a windowing technique for prime selection to enhance the key properties against pattern detection attacks. Security analysis shows that the enhanced system prevents decryption using values other than the dedicated key while maintaining additive and multiplicative homomorphism. Performance evaluations show that the improved system maintains decryption times and ciphertext expansion ratios similar to the original system, with a reasonable decryption time reduction. Statistical testing results using the National Institute of Standards and Technology tests demonstrate the robustness of the proposed approach compared to the original, with the windowing technique exhibiting superior randomness properties.

Index Terms: Homomorphic Encryption, Windowing Mechanism, Electronic Health Records, Privacy Preserving, Synthetic Healthcare Data

1. INTRODUCTION

According to estimates by clinical pathology laboratories, the number of patients whose personal information was revealed reached 2.2 million, and the number of patients whose banking information was affected reached 34,500 [1]. With the increasing demand for outsourced applications, data privacy on the Internet has become crucial. While customers benefit from these services by uploading their data to be evaluated and obtaining results, their data are exposed to third-party services [2]. Hospitals and healthcare organizations increasingly collect data due to the increasing use of this data in health research and personalized medicine based on big data technologies [3]. The amount and risk of cyberattacks on patients' medical data stored in the cloud are increasing daily.

The healthcare sectors focus on digitizing healthcare information because it improves care efficiency, making it more affordable and accessible. In the healthcare sector, hospitals or healthcare facilities keep a digital record of each

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp77-91

E-ISSN: 2521-4217
P-ISSN: 2521-4209

Copyright © 2025 Jalal and Mohammed. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Abdulrahman Tawfeeq Jalal, Department of Computer Science, College of Science, University of Sulaimani, Sulaymaniyah, Iraq. E-mail: abdulrahman.jalal@univsul.edu.iq

Received: 15-05-2025 Accepted: 28-07-2025 Published: 25-08-2025

patient called an electronic health record (EHR) that contains all demographic, clinical, historical, reporting data, progress notes, problems, medications, vital signs, immunizations, laboratory data, and radiology reports [2], [4]. A large percentage of patients allow their data to be used to support research and improve the quality of healthcare, but at the same time, they are concerned about privacy, misuse, and mishandling of their data [5]. For information to be secure, it must have the security characteristics of confidentiality, integrity, and availability [6]. To prevent data breaches, encryption is a vital component of information security systems, thus protecting private data from unauthorized access and ensuring integrity [7].

2. LITERATURE REVIEW

2.1. Data Security and Privacy Challenges

In multi-party environments such as the cloud, all parties are at risk of threats, whether the provider or the user. Therefore, the minimum requirement for data security is to be stored in its encrypted state [8]. Customers are very concerned about the content of data sent to the cloud or stored in the service provider's servers because it is easily accessible to service providers. If digital records are handled in a way that ensures their privacy during acquisition, storage, and computing, then complete privacy has been achieved. However, while privacy can be achieved during acquisition and storage, achieving privacy is challenging during processing [2].

2.2. Homomorphic Encryption as a Solution

When encrypted data are required to be used, it must be decrypted, which exposes it to risk at that moment, so homomorphic encryption is a solution to ensure its privacy and security [2]. Fully homomorphic encryption (FHE) is one of the most widely adopted encryption techniques, focusing on privacy at the highest levels [9]. This technique focuses on two aspects: data encryption and the ability to handle data that has been encrypted through mathematical operations without the need to decrypt it, thus ensuring its complete confidentiality.

The main types of homomorphic encryption schemes are divided according to the number and type of operations into Partial, Somewhat, and Full. In Partial, only one operation, either addition or multiplication (not both), is applied to the encrypted data an unlimited number of times. Examples of these are Rivest-Shamir-Adleman (RSA), ElGamal, and Paillier. In Somewhat, a limited number of mathematical operations are applied to the encrypted data a limited number of times. Finally, the full technique involves both addition and multiplication utilized simultaneously and an unlimited number of times [10].

As for the functions and use cases, the benefit of FHE is that it supports functions, such as searching, sorting, max, and min. However, despite the proposed attempts that have improved the efficiency and performance of FHE schemes, it is still difficult to implement them in a real service without disturbing the user. Although somewhat homomorphic encryption or leveled-FHE has achieved acceptable performance in some applications, the bootstrapping techniques still need improvement, as do the multiplication operations, which negatively affect performance [10]. Due to this feature, data encrypted with this technique can be handed over to third parties to perform any meaningful operation [3].

2.3. Related Work

In certain cases, user data becomes vulnerable, such as during pandemics when patients cannot reach hospitals, so they contact doctors using technology over the network to obtain information about the diagnosis [11]. This is done remotely through the presence of devices that provide accurate medical data and support the Internet of Things, which can then send readings of blood pressure, oxygen, temperature, etc. [12]. Carpov *et al.* developed a mobile application that encrypts users' data with a FHE algorithm and transmits it to the cloud [9]. Their implementation focused on analyzing and evaluating cardiac risk factors in the cloud by processing encrypted health data uploaded in a secure and privacy-preserving manner. They processed the data exchange between the doctor and the patient through an architecture that enables the patient to grant authorization to the doctor to access the results through a private FHE key.

Based on the increasing need for secure remote healthcare during pandemics, Kumar *et al.* proposed a hybrid approach of multi-party computing and homomorphic encryption to secure health data [11]. Their model addressed the aspect of secure architecture that ensures the results between the patient and the doctor. This prevents eavesdropping across different systems' entities and ensures trust among the patient community. They achieved the partial homomorphic encryption approach by adopting the Paillier encryption scheme. Despite the security, reliability, and patient-centric approach, it is relatively slow due to processing long ciphertexts instead of plain data.

Scheibner *et al.* presented a system that combines homomorphic encryption and secure multiparty computation

to address the limitations of traditional approaches to enhancing privacy during medical data sharing [13]. The authors developed a new homomorphic multiparty encryption approach that enables secure, flexible processing while meeting data anonymization and legal requirements under the General Data Protection Regulation (GDPR). This is done by efficiently transitioning between homomorphic encryption, represented by local computation, and interactive protocols, defined by multiparty computation. Their analysis showed that homomorphic multiparty encryption provides an effective solution compared to using the two techniques alone, and they acknowledged the limitations of computational complexity as one of the drawbacks of full homomorphic encryption.

Guo *et al.* stated that privacy issues exist in the model training and the pre-diagnosis stage in machine learning [14]. They designed a pre-diagnosis scheme that provides healthcare services using logistic regression called Privacy-Preserving Online Medical Pre-diagnosis in a privacy-preserving manner by employing homomorphic encryption techniques. A homomorphic encryption scheme called Boneh-Goh-Nissim was applied to protect the confidentiality of the feature vector x and the pre-diagnosis model ω. Their analysis showed that the proposal resisted security threats and privacy concerns. Furthermore, it is efficient and cost-effective regarding computational and communication burdens [14].

Anwar and Salman proposed a FHE algorithm to protect data privacy on the cloud based on a Super-increasing Sequence. Each character in the plaintext is converted to its American Standard Code for Information Interchange (ASCII) code and encrypted using the $C = S\ (2 \times q \times r) + \mathrm{m}$ algorithm to produce a ciphertext. They added r as a random value for the noise added to the ciphering process and evaluated the randomness through well-known National Institute of Standards and Technology (NIST) tests [15]. Their research results showed that their algorithm performs better than other proposed algorithms, as it can encrypt larger files. As for security, a super-increasing sequence with a subset sum problem that was used provides high security. However, they faced challenges regarding the encrypted file's size, which is larger than the plaintext file. Furthermore, the decryption process takes longer than the encryption process [15].

Sinha *et al.* proposed a privacy-aware surveillance system that ensures secure patient data sharing, leading to timely and effective responses to crisis scenarios (e.g., the COVID-19 pandemic). The proposed new encryption mechanism is based on a fully homomorphic and secure encryption scheme using the ElGamal algorithm. The effectiveness of the contact tracing method was analyzed, and the results showed that the proposed solution effectively provides security and adequate support for the computational needs of contact tracing [16].

Mohammed and Abed proposed a FHE algorithm that uses n prime numbers instead of two. They convert each character in the plaintext to its ASCII code and encrypt it using the $C = M + r \times L \times n$ model [17]. The key consists of L (a large prime integer), r (a random number added as noise), and n (a set of prime numbers multiplied together). Results showed better performance in encrypting different file sizes. The proposed method was compared with the DGHV and SDC schemes, showing superior performance. The security level is high due to using n as a multiplier for the i-Prime Modular Operation. However, the authors mention limitations related to encrypted text size, affecting storage space and the decryption process [17].

2.4. Compliance with Regulations and Standards

18 elements can directly and uniquely identify an individual as defined by the Health Insurance Portability and Accountability Act (HIPAA), such as name, social security number, and phone number. Until the mid-2000s, and based on HIPAA, removing all these fields was considered the simplest and most adequate way to protect privacy and de-identify data [18]. The same case concerns the GDPR, which focuses on personal data relating to a specific, identified, or identifiable person [19]. Homomorphic encryption thus complies with data protection regulations and privacy laws such as HIPAA in healthcare and GDPR in finance [20]. The provisions of the regulation guarantee the right to access information related to processing and allow data subjects to monitor how their data are used [21].

3. PROBLEM STATEMENT

Despite encryption being a vital component of information security systems, specific existing integer-based homomorphic encryption schemes contain critical mathematical vulnerabilities. Specifically, schemes using formulations:

$$C = S \cdot (q \cdot 2 \cdot r) + M \tag{1}$$

$$C = M + r \cdot L \cdot n \tag{2}$$

$$C = M + L(rK + i) \tag{3}$$

Where:

- C = Ciphertext

- M = Message/Plaintext
- L = Large prime integer
- K = Secret key
- r = Noise/random number
- i = Counter/iteration variable
- n = Product of multiple prime numbers
- S = Parameter greater than the sum of the super-increasing sequence
- q = Sum of chosen numbers from the transformed sequence

Equation (1) in [15], Equation (2) in [17], and Equation (3) in [22], [23] are susceptible to unauthorized decryption when attackers utilize factors that exceed m, effectively bypassing the intended secret prime key. This vulnerability contradicts the fundamental security model of FHE, which aims to protect sensitive healthcare data from privacy disclosure risks during server-side processing. In addition, present implementations fail to specify secure methods for random number generation and lack resistance to pattern detection attacks. With the increasing digitization of healthcare information and application outsourcing, this security gap poses significant risks to patient data privacy. Therefore, an enhanced encryption scheme that eliminates these vulnerabilities while maintaining homomorphic properties is urgently needed.

4. THEORETICAL FRAMEWORK

4.1. Math Model

The encryption scheme is homomorphic when: E(m1) ★ E(m2) = E (m1 ★ m2), ∀m1, m2 ∈ M. Where E is the cipher algorithm, M is the plaintext to be encrypted, and ★ is the mathematical operation. This means encrypting m1 and m2 and then performing the operation ★ on their results is the same as performing the operation ★ on m1 and m2 first and then encrypting the result.

4.2. Noise and Circuit Depth

The noise component is an essential element in the FHE encryption process to ensure that many possible ciphertexts are obtained for each message that needs to be encrypted and thus ensure semantic security [9]. In FHE encryption, there is a direct proportion between noise and the number of homomorphic operations, where a substantial increase in noise has a negative effect and makes it difficult to decrypt correctly. In particular, multiplication increases noise significantly compared to addition, and the encrypted data has to be refreshed regularly by bootstrapping [2]. This depends on the circuit depth; the number of operations (addition and multiplication) can be performed in sequence while still being able to decrypt correctly.

4.3. Architecture and Quality

The security model on which FHE is based essentially consists of two parties: The user who owns the data with the privacy and the software service provider. The purpose and function of FHE is to protect user data from the risks of privacy disclosure that may be exposed to it on the server [9]. The quality and performance of this model depend on the evaluation results of three main criteria: security, speed, and simplicity. The scheme should be consistent, understandable, and within the standards and best practices followed by other practitioners in the field, and thus, be simple, applicable, and producible [10].

5. MATERIALS AND METHODS

5.1. Data Collection

There are significant challenges in accessing high-quality datasets due to data use agreements, privacy concerns, ethics reviews, and costs. Therefore, synthetic data offers an attractive alternative to address these concerns [19]. Synthetic data is artificially generated using a trained model that replicates real data according to its distributions (shape and variance) and structure (attributes' correlations) [24]. When synthetic data is derived from sensitive medical information, the challenge is to ensure that it cannot access details that would lead to the re-identification of individuals [19]. Synthea is an open-source synthetic health simulator that generates synthetic EHRs, and it was proposed by Walonoski *et al.* as an approach to simulate patient information in JSON and CSV format [25].

5.2. Original Scheme Analysis

In the proposed encryption scheme [17], the text is encrypted in a homomorphic way using equation (2), where m is the plaintext message; specifically, the character's ASCII value.

The plaintext is recovered using the modular operation. Let us derive this step-by-step:

$$C \bmod L = (M + r \cdot L \cdot n) \bmod L \tag{4}$$

Then, using the basic property of modular arithmetic:

$$(a+b) \bmod n = ((a \bmod n) + (b \bmod n)) \bmod n \tag{5}$$

We get:

$$C \bmod L = ((M \bmod L)+((r \cdot L \cdot n) \bmod L)) \bmod L \quad (6)$$

Based on the fact that any number multiplied by a coefficient is divisible by it, then the noise term becomes zero:

$$((r \cdot L \cdot n) \bmod L) = 0 \quad (7)$$

Then the equation simplifies to:

$$C \bmod L = ((M \bmod L)+ 0) \bmod L \quad (8)$$

Since m ∈ [0, L-1] according to [17], it concludes:

$$C \bmod L = (M \bmod L) = M \quad (9)$$

The advantage of this scheme is that it supports homomorphic encryption. It does not require noise management and bootstrapping to reduce noise growth, especially with the multiplication process, as the modulus operation eliminates it.

For addition, taking *mod L*, let the noise term disappear:

$$C_add = C1+C2 = (M1+r1 \cdot L \cdot n)+(M2+r2 \cdot L \cdot n) = M1+M2+nL(r1+r2) \quad (10)$$

$$C_add \bmod L = (M1+M2+nL(r1+r2)) \bmod L = M1+M2 \quad (11)$$

For multiplication, taking *mod L* eliminates any term contains noise:

$$C_mul = C1 \times C2 = (M1+r1 \cdot L \cdot n)(M2+r2 \cdot L \cdot n) \quad (12)$$

$$C_mul = (M1 \cdot M2)+(M1 \cdot r2 \cdot L \cdot n)+(M2 \cdot r1 \cdot L \cdot n)+(r1 \cdot r2 \cdot L^2 \cdot n^2) \quad (13)$$

$$C_mul \bmod L = M1 \cdot M2 \quad (14)$$

However, the authors did not specify how the random numbers are generated and how secure they are. As for security, no proven and tested method has been adopted to produce secure random numbers for cryptographic use. In addition, we identified a critical mathematical vulnerability in this scheme and similar schemes. Any factor of the terms used in the key then m can be used in decryption instead of L, which is claimed to be the only secret big prime integer key used in decryption.

Let f be any factor of the term $r \times n$ such that $f > M$. Then:

$$C \bmod f = (M+r \cdot L \cdot n) \bmod f \quad (15)$$

Since $r \cdot L \cdot n \equiv 0 \ (mod\ f)$ (as f divides r or n), this simplifies to:

$$C \bmod f = M \quad (16)$$

For example, using n = 114197120264498507, which is the product of (1009, 1013, 2153, 6553, 7919) primes, then all possible keys that can be used in decryption other than L are 31 keys (primes and products of primes). When the attacker does not need to know all the primes used, then using n of them will be a disadvantage. Unlike RSA, someone needs to get the value of p and q to calculate d and decrypt the message [26].

5.3. Enhanced Scheme Design

Encryption process:

$$C = M+(r1 \cdot L)+(r2 \cdot L \cdot W(n)) \quad (17)$$

Where:

- L = Secret large prime key
- n = Composite large value of prime numbers
- $r1$, $r2$ = Large random numbers
- $W(n)$ = Product of prime numbers in the current window

Decryption process:

$$M = C \bmod L \quad (18)$$

$$M = (M+(r1 \cdot L)+(r2 \cdot L \cdot W(n))) \bmod L \quad (19)$$

$$M = (M+(0)+(0)) = M, \text{ when } M < L \quad (20)$$

5.4. Random Number Generator (RNG)

We evaluate two approaches for generating the random values $r1$ and $r2$: First, Counter_Deterministic Random Bit Generator (CTR_DRBG) is a counter-mode deterministic random bit generator implemented based on NIST specifications. Second, Java SecureRandom is a built-in cryptographically secure pseudo-RNG in Java.

5.5. Windowing Technique for Prime Selection

One of the reasons for increasing the size of the ciphertext is to adopt a set n that contains an uncontrolled number of prime numbers. Therefore, we propose an enhanced FHE scheme that provides a technique for adopting specific prime numbers from the n set. This is done by employing a dynamic prime window mechanism while maintaining homomorphic properties. Instead of using a full fixed set of prime numbers when encrypting each character, a sliding window W of size k is defined, where the size of $k < n$. A subrange of primes is extracted for each ASCII character, and the window shifts according to a deterministic function.

We implemented a random windowing approach for the prime selection to enhance security. For each encryption operation, we select a subset of *n* primes based on:

Window size:

$$K(seed) = k_{min} + ((seed) \bmod (k_{max} - k_{min} + 1)) \quad (21)$$

Window offset:

$$O(seed) = \left[\frac{seed}{(k_{max} - k_{min} + 1)} mod \left(n - K(seed) + 1 \right) \right] \quad (22)$$

Where:

- k_{min} = Minimum window size
- k_{max} = Maximum window size
- seed = Deterministic value

5.6. Experimental Setup

To evaluate the performance and security of our enhanced scheme, we designed a comprehensive test suite with the following components:

1. Security analysis: We do a vulnerability check regarding the identified gap with a demo example
2. Homomorphic property verification: We test the enhanced scheme regarding homomorphic encryption properties (addition and multiplication)
3. We evaluate the selected RNGs randomness properties using the NIST Statistical Test Suite and choose the more secure one
4. Performance testing: We measure encryption and decryption times and ciphertext expansion ratios across file sizes ranging from 100 bytes to 10 MB
5. We check the role and effectiveness of the windowing technique
6. We apply the technique to practical use cases.

5.7. Testbed

1. Hardware Configuration
   a. Operating System: Windows 10
   b. Architecture: 64-bit operating system, x64-based processor
   c. Processor: Intel(R) Core (TM) i7-6600U CPU @ 2.60GHz 2.81 GHz
   d. Memory: 8.00 GB installed RAM (7.88 GB usable)

2. Software Environment
   a. Programming language: Java (version 20.0.2)
   b. Cryptographic libraries:
      i. Java Security API for SecureRandom implementation
      ii. Custom implementation of CTR_DRBG based on NIST SP 800-90A
   c. Analysis Tools:
      i. NIST Statistical Test Suite (STS)
      ii. Cygwin64 terminal environment for running NIST tests

6. EXPERIMENTS

6.1. Random Number Generation

Hardware or software can produce binary sequences based on cryptographic randomness for random number generation. The generated random number can be used to create keys in cryptographic applications, and common cryptosystems may generate keys randomly and employ them. However, the produced number may have patterns that can be predicted. Hence, tests must be done to check for non-randomness in the generated sequence. The NIST statistical package can be used with 15 tests to test the randomness and identify weaknesses or patterns in the produced binary sequences.

A RNG uses a non-deterministic source based on entropy and processing functions. Any resulting weakness in the entropy source is overcome using the distillation process, thus avoiding producing non-random numbers such as those containing long strings of zeros or ones. On the other hand, the pseudo-RNG relies on deterministic functions to generate pseudorandom numbers through one or more random and unpredictable seed inputs [27]. Pseudorandom numbers appear more random compared to random numbers generated from physical sources. These transformations introduce additional randomness and eliminate the associations between inputs and outputs, so the statistical properties of the output of a pseudo-RNG may be better. They can be generated faster than an RNG [27].

Key requirements such as randomness, unpredictability, uniformity, scalability, and consistency should be satisfied to develop a new random bit sequence generation technique [27]. Regarding unpredictability, there are two types: forward and backward unpredictability. Forward unpredictability means that if the seed is unknown and the previous random numbers in the sequence are known, the next number must be unpredictable. Backward unpredictability means that knowing which values were generated does not lead to the possibility of determining the seed.

Based on NIST SP 800-90A, we can use approved DRBG techniques, such as Hash, Hash-based Message Authentication Code, or CTR. We chose Counter-Deterministic Random Bit Generator for its performance; it uses an approved block cipher with Advanced Encryption Standard and runs at its native speed [28].

A special key K, a counter V, and a block cipher lock box are the basic components of CTR-DRBG. The counter V is encrypted with key K using our block cipher; the output becomes the first random number. The key is kept secret, and the counter will not be reset until it reaches the specified large number limit. In addition, knowing one random number does not lead to the next number without knowing the key. CTR-DRBG has four core functions: Update, Instantiate, Reseed, and Generate (Fig. 1) [28].

On the other hand, we have SecureRandom library that complies with the statistical RNG tests specified in FIPS 140-2 and security requirements for cryptographic modules [29]. For seed management, it is handled internally by Java SecureRandom based on platform security standards. As mentioned, even if set seed is not called, the first call to next bytes will force the SecureRandom object to create a self-seed [29]. SecureRandom automatically seeds itself from operating system entropy sources such as/dev/urandom on Linux/Unix systems and Microsoft CryptoAPI on Windows [30]. Thus, when we don't have a better seeding data source, and SecureRandom uses a highly random data source to seed automatically and securely, we can rely on it, avoiding manual seed management.

Regarding testing RNG using NIST, the recommended minimum input size for the Frequency (Monobit) Test, Runs Test, and Cumulative Sums Test is 100 bits. On the other hand, the input size for the Universal test ranges from 387,840 bits to 1,059,061,760 bits based on the block size [27]. Hence, based on that, we choose to generate sequences of 10,342,400 bits in total at least to test the generated numbers by all tests in the suite properly. For each statistical test, the minimum rate to pass the test is approximately equal to 8 for a sample size of 10 binary sequences [27].

6.2. Vulnerability Check and Security Enhancement

$$C \bmod f = (M+(r1\cdot L)+(r2\cdot L\cdot n)) \bmod f \quad (23)$$

Case 1: Decryption failure when f is a factor of n

$$C \bmod f = (M+(noise)+(0)) \neq M \quad (24)$$

Case 2: Decryption failure when f is a factor of $r1$

$$C \bmod f = (M+(0)+(noise)) \neq M \quad (25)$$

Demo Example: Message $M1 = 2$, Secret key $L = 151$ (prime number greater than $M1$), $n = 10403$ (101 × 103), $r1 = 847$, and $r2 = 548$:

$$Equation\ (17)\text{: } C1 = 2+(847\times151)+(548\times151\times10403) = 860955343$$

$$Equation\ (18)\text{: } M1 = C1 \bmod L = 860955343 \bmod 151 = 2$$

Testing decryption with potential factor values of $r1$ and $r2$ × n factors (e.g., 1, 7, 11, 77, 121, 847, 101, 103, 274) does not successfully recover $M1$. Hence, decryption now relies solely on the value of the secret key L.

6.3. Homomorphic Properties Verification

Additive Property, based on Equation (17):

$$C_add = C1+C2 = (M1+(r1\cdot L)+(r2\cdot L\cdot n))+(M1+(r3\cdot L)+(r4\cdot L\cdot n))$$

$$C_add = (M1+M2)+L(r1+r3)+nL(r2+r4)$$

Decryption of Sum, based on Equations (5), (18):

$$C_add \bmod L = ((M1+M2)+ L\cdot(r1+r3)+L\cdot n\cdot(r2+r4)) \bmod L$$

$$C_add \bmod L = ((M1+M2)+0+0) = M1+M2 \quad (assuming\ M1+M2<L)$$

Multiplicative Property, based on Equation (17):

$$C_mul = C1\times C2 = (M1+(r1\cdot L)+(r2\cdot L\cdot n))\times(M1+(r3\cdot L)+(r4\cdot L\cdot n))$$

$$C_mul = M1M2+L(r3M1+r1M2)+nL(r4M1+r2M2)+L^2(r1r3)+nL^2(r1r2+r2r3)+L^2n^2(r2r4)$$

Decryption of Product, based on Equations (5), (18):

$$C_mul \bmod L = M1M2+0+0+0+0+0 = M1M2 \quad (assuming\ M1\cdot M2<L)$$

7. RESULTS AND DISCUSSION

7.1. NIST Results for RNGs

We evaluated both CTR_DRBG and Java SecureRandom using the NIST statistical test suite. The tests were conducted using 10 separate bitstreams, each containing 1,034,240 bits in a binary file. Table 1 summarizes the comparative results.

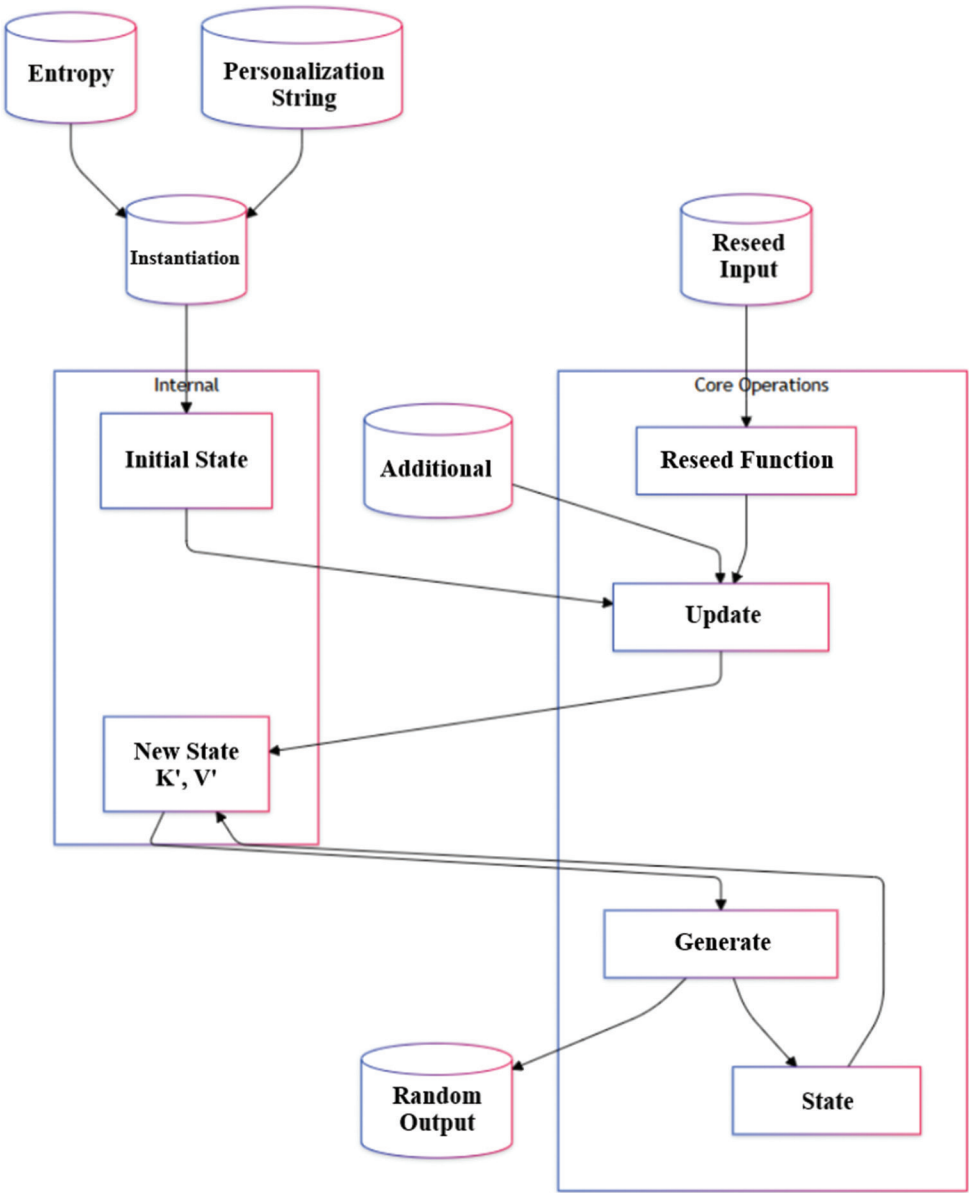


Fig. 1. Counter-deterministic random bit generator (CTR-DRBG) architecture.

A *P*-value of "0.01" is considered the minimum acceptable threshold to pass the NIST test, with stronger random properties when it is close to 1 (Fig. 2). We note that both generators passed all statistical tests successfully, demonstrating their suitability for cryptographic applications. In the Runs test, while CTR_DRBG has a higher P-value, SecureRandom passes with all 10 sequences (10/10), whereas CTR_DRBG failed in one with 9/10. Based on these results, we chose SecureRandom for implementation.

7.2. Performance Comparison

We tested the original *n*-Primes Homomorphic Encryption Algorithm for encryption and decryption times, ciphertext size, and the expansion ratio (Table 2). We use the following parameters: L = 11579208923731619542[truncated] (256 bits), n (product of primes) = 269599424803240402777[truncated] (224 bits), random number bit length = 128 bits. Hence, the combined security level is ~608 bits.

Then we tested the Enhanced Scheme using the following parameters (Table 3): L = 11579208923731619542[truncated] (256 bits), available primes pool = 32 primes (each ~75 bits), window parameters k_min = 2, k_max = 4 (uses 2-4 primes per encryption), security level of n = 151-301 bits (depending on window size), and random number bit length = 128 bits

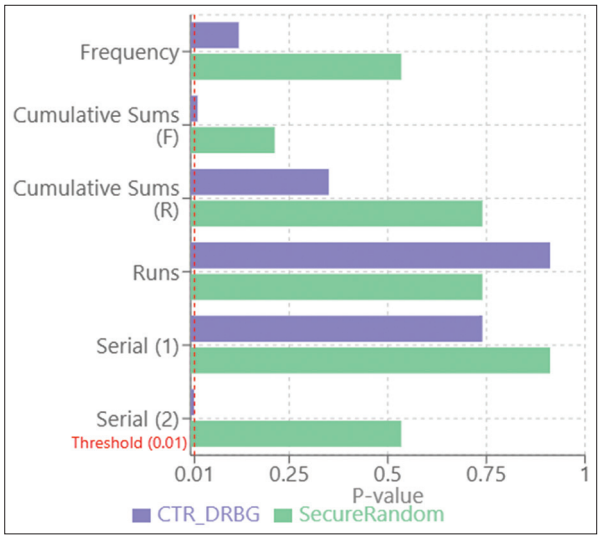


Fig. 2. National Institute of Standards and Technology statistical test *P*-values for CTR_DRBG and SecureRandom generators across selected tests (minimum threshold $P \geq 0.01$ for passing). CTR_DRBG: Counter_Deterministic random bit generator.

for each *r*1 and *r*2. Hence, the effective combined security level is ~535-685 bits and ~610 on average.

For larger files such as the tested file (10 MB), the improved system requires more time in encryption, representing a performance cost 2.89 times higher than the original system (Fig. 3). This additional processing time is due to the computational complexity of the enhanced scheme, which focuses on improving security. Two random values were used, in addition to a windowing approach, which means that the increased security impacted performance.

The decryption time is similar between the two schemes, but the enhanced scheme shows a slightly better advantage for the largest scanned file size. Consequently, decryption times remain comparable between the two schemes (Fig. 4). It should be noted that the calculated time is only for the decryption process using mod; the time required to retrieve the encrypted data from storage is not included.

As for ciphertext size, both systems achieved an expansion size of around 183× for all file sizes, indicating that the added security measures do not significantly impact this factor. The y-axis range is intentionally narrowed to (182–185) to highlight the slight differences (Fig. 5).

7.3. Windowing Technique Role

To test the effectiveness of this technique for security, we chose to test the generated keys using the NIST test suite.

TABLE 1: Comparison of NIST test results for RNGs

Statistical tests	CTR_DRBG		SecureRandom	
	P-value	Proportion	*P*-value	Proportion
Frequency (monobit)	0.122325	10/10	0.534146	10/10
Block frequency	0.739918	10/10	0.350485	10/10
Cumulative sums (forward)	0.017912	10/10	0.213309	10/10
Cumulative sums (reverse)	0.350485	9/10	0.739918	10/10
Runs	0.911413	9/10	0.739918	10/10
Longest run of ones	0.739918	10/10	0.739918	10/10
Rank	0.213309	10/10	0.350485	10/10
Fast Fourier transform	0.350485	10/10	0.534146	10/10
Overlapping template	0.534146	10/10	0.122325	10/10
Universal statistical	0.911413	10/10	0.739918	10/10
Approximate entropy	0.122325	10/10	0.213309	10/10
Serial (1)	0.739918	10/10	0.911413	10/10
Serial (2)	0.008879	10/10	0.534146	9/10
Linear complexity	0.350485	10/10	0.534146	10/10

NIST: National Institute of Standards and Technology, RNGs: Random number generators, CTR_DRBG: Counter_Deterministic random bit generator

TABLE 2: N-prime scheme performance test

File size	Encryption (ms)	Decryption (ms)	Cipher size	Expansion
100.00 Bytes	44	0	17.87 KB	183.01×
500.00 Bytes	69	2	89.35 KB	182.99×
1.00 KB	8	39	182.97 KB	182.97×
2.00 KB	24	5	365.90 KB	182.95×
5.00 KB	56	8	914.78 KB	182.96×
10.00 KB	39	7	1.79 MB	182.96 ×
20.00 KB	54	25	3.57 MB	182.95×
50.00 KB	106	31	8.93 MB	182.95×
100.00 KB	214	92	17.87 MB	182.95×
200.00 KB	635	119	35.73 MB	182.95 ×
500.00 KB	770	300	89.33 MB	182.95×
1.00 MB	1364	608	182.95 MB	182.95×
2.00 MB	3583	1356	365.91 MB	182.95×
5.00 MB	8358	3238	914.77 MB	182.95×
10.00 MB	14668	8578	1.79 GB	182.95×

For the first original *n* Prime technique, we generated keys of a 1,000,000-bit sequence binary file using the $r \times L \times n$

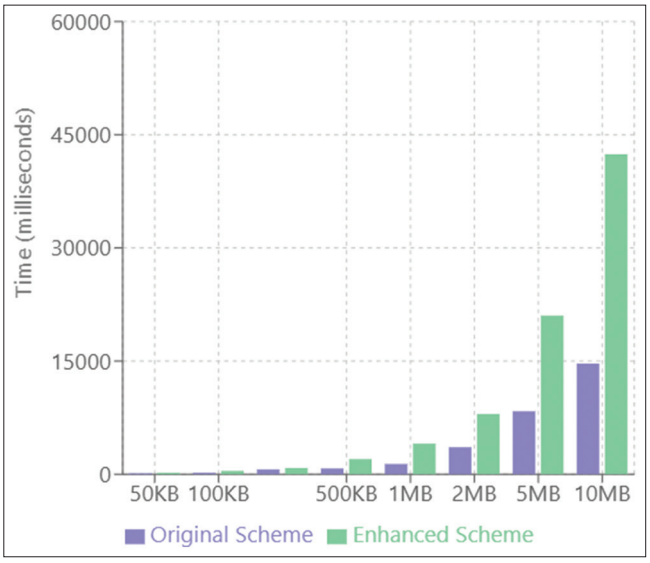


Fig. 3. Encryption time comparison between original and enhanced *n*-Primes homomorphic encryption schemes across different file sizes.

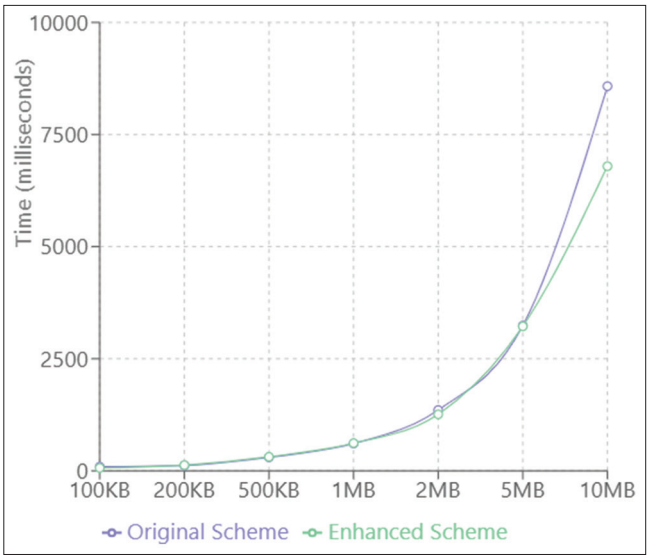


Fig. 4. Decryption time comparison between original and enhanced *n*-Primes homomorphic encryption schemes across different file sizes.

term, where n is pre-computed by multiplying three 76-bit primes. Then, we generated another similar file using the $r1 \times L + r2 \times L \times W(n)$ term of the enhanced technique, but without windowing; n is fixed, pre-computed, and equal to the one used first. Finally, we generated the third bin file based on the enhanced technique with windowing (window size = 3 primes) applied over n from a set that contains 32 different 76-bit length primes. We tested all files with 10 separate bitstreams, each containing 100,000 bits (Table 4).

TABLE 3: Enhanced scheme performance test

File size	Encryption (ms)	Decryption (ms)	Cipher size	Expansion
100.00 Bytes	71	0	17.98 KB	184.13×
500.00 Bytes	37	1	89.57 KB	183.44×
1.00 KB	22	2	183.76 KB	183.76×
2.00 KB	21	4	365.23 KB	182.61×
5.00 KB	48	6	919.09 KB	183.82×
10.00 KB	66	7	1.79 MB	183.66×
20.00 KB	108	14	3.58 MB	183.34×
50.00 KB	205	36	8.96 MB	183.52×
100.00 KB	447	65	17.90 MB	183.34×
200.00 KB	829	126	35.83 MB	183.43×
500.00 KB	2013	312	89.56 MB	183.43×
1.00 MB	4060	617	183.42 MB	183.42×
2.00 MB	7994	1260	366.86 MB	183.43 ×
5.00 MB	21024	3222	917.15 MB	183.43×
10.00 MB	42416	6793	1.79 GB	183.41×

TABLE 4: Comparison of NIST test results regarding key generation

NIST test	Fixed *n* prime	Enhanced with W (n)	Enhanced without W (n)
Frequency	0.002043 (9/10)	0.350485 (10/10)	0.122325 (10/10)
Block frequency	0.017912 (9/10)	0.017912 (10/10)	0.350485 (10/10)
CUMULATIVE sums (1)	0.008879 (9/10)	0.213309 (10/10)	0.739918 (10/10)
Cumulative sums (2)	0.002043 (10/10)	0.534146 (10/10)	0.122325 (10/10)
Runs	0.534146 (10/10)	0.739918 (10/10)	0.739918 (10/10)
Longest run	0.350485 (10/10)	0.350485 (10/10)	0.066882 (10/10)
Rank	0.534146 (10/10)	0.911413 (10/10)	0.534146 (10/10)
FFT	0.350485 (10/10)	0.534146 (10/10)	0.911413 (10/10)
Overlapping template	0.739918 (10/10)	0.911413 (10/10)	0.739918 (10/10)
Approximate entropy	0.213309 (10/10)	0.350485 (10/10)	0.066882 (10/10)
Serial (1)	0.739918 (10/10)	0.213309 (10/10)	0.035174 (10/10)
Serial (2)	0.534146 (10/10)	0.213309 (10/10)	0.213309 (10/10)
Linear complexity	0.017912 (10/10)	0.739918 (10/10)	0.035174 (10/10)

NIST: National Institute of Standards and Technology, FFT: Fast Fourier Transform

The enhanced windowing technique shows the best randomness properties of keys. It won the most tests (Fig. 6), has the highest average *P*-value (Fig. 7), shows the most consistent randomness across different tests (Fig. 8), and has the largest area on the polar graph (Fig. 9).

Fig. 5. Ciphertext expansion ratio comparison between original and enhanced *n*-Primes schemes across different file sizes.

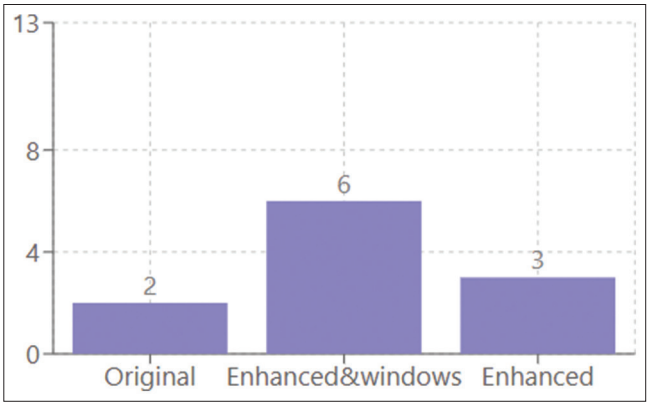


Fig. 6. Number of National Institute of Standards and Technology tests won by original, enhanced without windowing, and enhanced with windowing techniques.

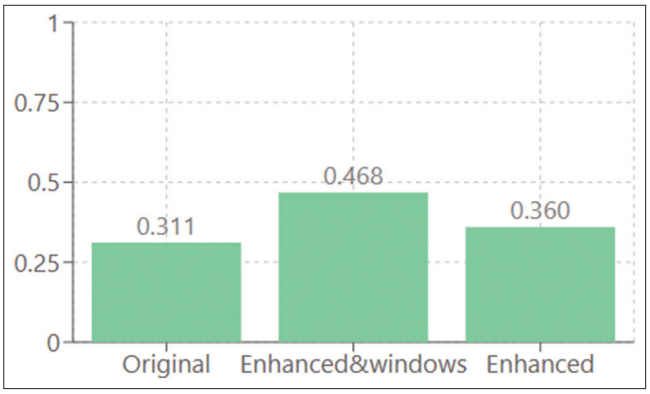


Fig. 7. Average National Institute of Standards and Technology *P*-values for original, enhanced without windowing, and enhanced with windowing techniques.

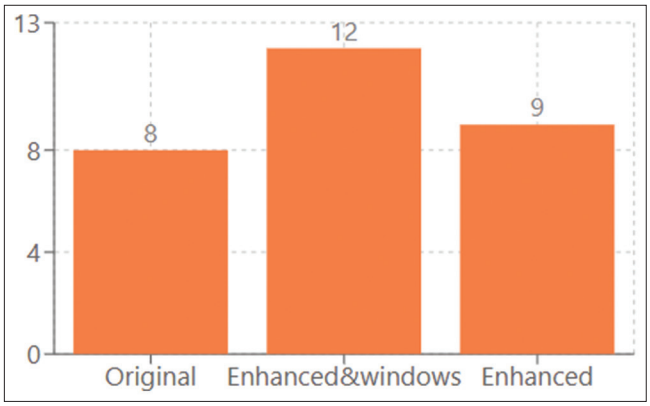


Fig. 8. Number of National Institute of Standards and Technology tests achieving solid randomness ($P > 0.1$) for each key generation technique.

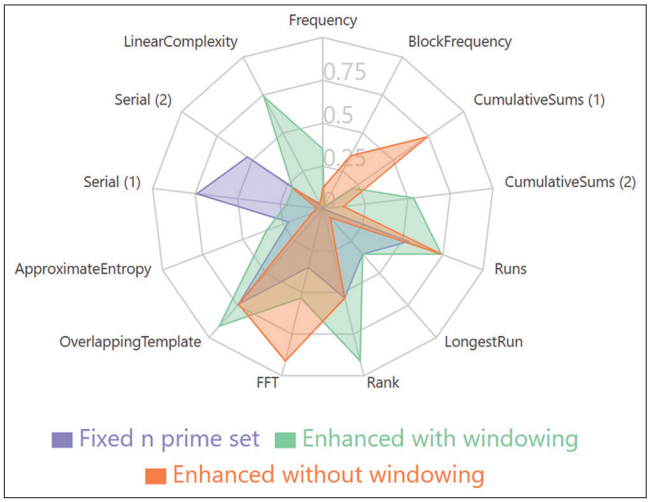


Fig. 9. Security level of the generated keys based on National Institute of Standards and Technology statistical tests.

Logically, using dynamic subsets adds variability, not just mathematical robustness. This technique adds a layer of randomness because each encryption operation uses a different value of n, making it more difficult to detect patterns in ciphertexts. We also chose to have a set of primes and then extract subsets from it instead of trying to generate the primes on the fly when needed, which is computationally expensive.

7.4. Security versus Performance Tradeoff

There is a clear trade-off between security and performance in the enhanced encryption scheme, as computational overhead is added to address vulnerability in decryption attacks using factors instead of a dedicated key. Further computational overhead is added when using windowing

techniques to make the ciphertexts' own random properties resistant to pattern detection attacks (Fig. 10). Our improved scheme offers greater security benefits and may be preferred in applications where confidentiality is the priority and performance is secondary. The windowing approach also provides a limited solution by reducing the ciphertext size to a minimum acceptable level or increasing it to a certain level by modifying the window size parameters. This allows for a tailored balance between security and performance, but its impact remains limited.

8. USE CASES

8.1. Medication Cost Analysis

Patient records contain numerical data, such as body mass index, blood pressure, glucose levels, medication doses, treatment rates, and healthcare expenditures, that a third party may need to perform operations on. Synthea™ provides a dataset in CSV and JSON files containing patient demographics, clinical notes, medications, encounters, and regulatory data [31].

Analyzing medication costs typically requires access to sensitive financial data. Research needs (such as access to spending patterns, cost variations, and improvement opportunities) conflict with privacy requirements. We implemented our integer-based homomorphic encryption to enable statistical analysis of costs, such as measuring total spending and average costs for each type of medication, without revealing individual patient data.

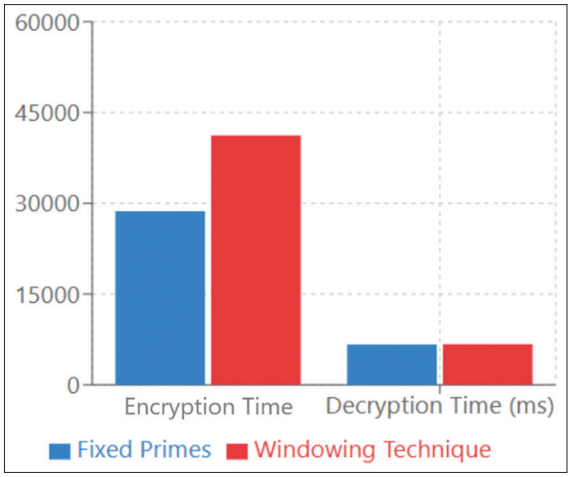


Fig. 10. Encryption and decryption time comparison between fixed primes (without windowing) and windowing approaches for the 10 MB file using the enhanced *n* Primes scheme.

The analysis was performed on a dataset of 42989 medication records, with 131 unique medication descriptions. Cost values were represented in decimals with up to two decimal places, ranging from $0.99 to $7015.8 per medication. Since our homomorphic encryption scheme operates on integers, we had to scale the data as a preprocessing step by converting the decimal cost values to integers by multiplying them by 100 (e.g., $10.25 ← 1025 cents). Then, after applying encryption to these integers and performing homomorphic operations over them, we scale back the results by dividing them by 100 (Fig. 11).

We computed the total spending cost grouping per description by applying homomorphic addition to the encrypted values. Then, we computed the average by plaintext division post-decryption. We compared the results of the homomorphic operations with calculations on plain text data to verify accuracy. As shown in Table 5, our implementation achieved perfect accuracy. As for operational efficiency, performance metrics were recorded with an encryption time of 657 ms for 42,989 records and a decryption time of 7 ms for statistical results.

8.2. CHADS2 Score Calculation

The CHADS2 score is used as a validated tool for predicting stroke risk in patients with Atrial Fibrillation (AF). Five risk factors are assessed with a score ranging from 0 to 6: Congestive heart failure (1 point), hypertension (1 point), age ≥75 years (1 point), diabetes (1 point), and prior stroke/transient ischemic attack (2 points) [32]. Scores of ≥ 3 indicate an increased risk of stroke, which may require anticoagulation [33].

$$CHADS2Score = C+H+A+D+(S\times 2) \quad (26)$$

Patients with AF were identified in the conditions dataset by searching case descriptions (34 out of 1,171 patients). Risk factors (CHADS2 components) were then extracted for each AF patient. We encoded each risk factor as 1 if present and 0 if absent, then homomorphically encrypted the values. Finally, we calculated the score using E(C), E(H), E(A), E(D), and E(S).

The results of calculating encrypted CHADS2 scores showed 100% accuracy compared to plaintext calculations for 34 AF patients identified. 85.3% of these patients were classified as high risk. The total processing time was 309 ms, with 98.4% spent on encryption, 1.3% on decryption, and <1% on calculation.

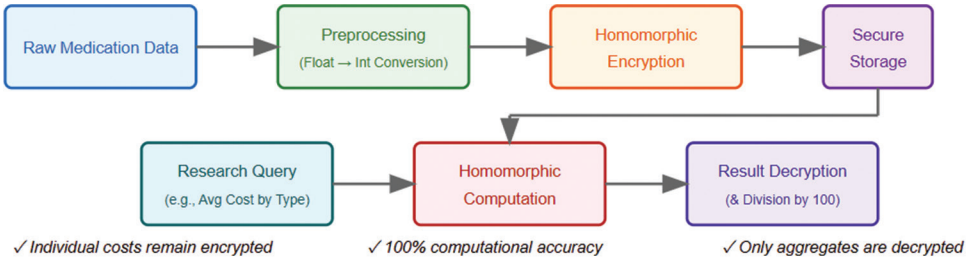


Fig. 11. Medication cost analysis workflow.

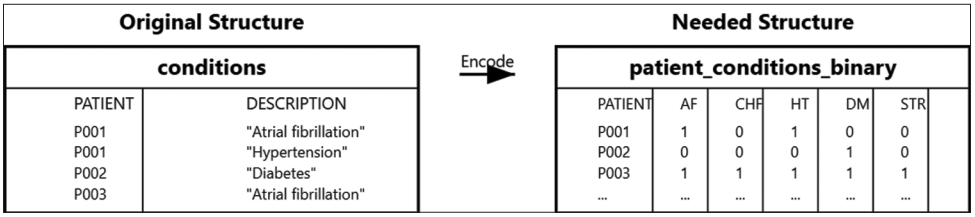


Fig. 12. Data structure required to apply homomorphic operations over binary encoded categorical data to ensure full privacy.

TABLE 5: Accuracy comparison for selected medications

Medication description	Average after decryption	Plaintext average	Error %
Insulin Lispro 100 UNT/ML	$1,410.89	$1,410.89	0.000000
Simvastatin 10 MG	$5,891.26	$5,891.26	0.000000
Acetaminophen 325 MG	$7.05	$7.05	0.000000
Hydrochlorothiazide 25 MG	$2,448.78	$2,448.78	0.000000

However, the current technique faces challenges as related math operations require dealing with categorical data. For instance, identifying a patient with AF requires decrypting this information during comparison and encoding, which compromises privacy. Similarly, identifying descriptive risk factors required matching for "hypertension," "diabetes," etc., which required displaying the data unencrypted.

Binary encoding can ensure complete privacy through additional specific structured storage (Fig. 12). One drawback of this method is its storage cost, as it requires as many columns as the expected number of descriptions—about 131 in the current conditions' dataset. For example, a patient with diabetes would have a value of 1, while the remaining 130 descriptions would have a value of 0. All these values would be homomorphically encrypted, which would greatly increase storage requirements.

9. LIMITATIONS

Integer-Only Operations: Converting floating-point values to integers requires a specific system-level scale value; otherwise, metadata for each numerical field must be retained for use in the scale-back process. Therefore, the results need careful conversion back to the original units after decryption.

Limited Operations: Supporting homomorphic addition and multiplication is insufficient for performing other statistical operations (e.g., division and ≥ in the applied use cases). Therefore, complex calculations (such as correlation, regression, variance, etc.) cannot be performed directly on the encrypted data.

Grouping Impact on Security: When operations are required on groups (e.g., aggregating and calculating the average for each medication), metadata such as descriptions should remain unencrypted. Alternatively, it should be encrypted with the same parameters and random numbers to obtain the same ciphertext for each description to enable grouping based on ciphertexts.

10. FUTURE WORK

After implementing the technique, we introduced floating-point numbers to observe how it handles them. The encryption and decryption were successful, but the decimal parts remained visible in the ciphertext. For example, 2.24 as plaintext becomes 860955343.24 ciphertext, with the fractional value "0.24" still intact. Therefore, it is important to investigate this issue further and to establish mathematical methods for securely encrypting the fractional parts in future work.

We also carried out an initial investigation into the homomorphic properties using floating-point numbers, finding that they work for addition but not for multiplication. Therefore, further mathematical analysis is necessary to determine if the current encryption formula needs specific adjustments.

When decrypting 860955343.24, the expected value was 2.24, but the actual result was 2.24000000954, with a slight difference in precision. However, identifying the cause of this discrepancy and assessing the potential for significant deviations are essential for defining acceptable error limits.

11. CONCLUSIONS

This study addressed a key vulnerability that allows using factors for unauthorized decryption instead of the assigned secret key in integer-based homomorphic encryption schemes. Security analyses demonstrated that our improved encryption formula mitigates this vulnerability by ensuring that decryption relies solely on the assigned secret key while preserving additive and multiplicative homomorphic properties. NIST analyses also demonstrated that the random window technique strengthens the scheme against pattern detection attacks by introducing variability and random properties into the ciphertext generation process. Performance evaluations showed that the enhanced scheme maintains close decryption times and identical ciphertext expansion ratios to the original scheme, with slightly improved decryption performance for larger files despite the additional security measures. The medication cost analysis use case achieved optimal accuracy in homomorphic processes with reasonable performance metrics. Future research should focus on expanding the system to support more complex statistical operations directly on encrypted data, improving computational efficiency for real-time applications, and developing support for handling floating-point values. Despite these limitations, our research contributes significantly to creating more secure homomorphic cryptosystems for healthcare applications.

REFERENCES

[1] J. R. Paragas. "*An Enhanced Cryptographic Algorithm in Securing Healthcare Medical Records. In: 2020 Third International Conference on Vocational Education and Electrical Engineering (ICVEE)*". IEEE, Piscataway, pp. 1-6, 2020.

[2] K. Munjal and R. Bhatia. "A systematic review of homomorphic encryption and its contributions in healthcare industry". *Complex and Intelligent Systems*, vol. 9, no. 4, pp. 3759-3786, 2023.

[3] J. Scheibner, M. Ienca, S. Kechagia, J. R. Troncoso-Pastoriza, J. L. Raisaro, J. P. Hubaux, J. Fellay, E and Vayena. "Data protection and ethics requirements for multisite research with health data: A comparative examination of legislative governance frameworks and the role of data protection technologies". *Journal of Law and the Bioscience*, vol. 7, no. 1, p. lsaa010, 2020.

[4] C. S. Kruse, B. Smith, H. Vanderlinden and A. Nealand. "Security techniques for the electronic health records". *Journal of Medical System*, vol. 41, no. 8, p. 127, 2017.

[5] C. Brall, C. Berlin, M. Zwahlen, K. E. Ormond, M. Egger and E. Vayena. "Public willingness to participate in personalized health research and biobanking: A large-scale Swiss survey". *PLoS One*, vol. 16, no. 4, p. e0249141, 2021.

[6] C. K. Yee and M. F. Zolkipli. "Review on confidentiality, integrity and availability in information security". *Journal of ICT Education*, vol. 8, no. 2, pp. 34-42, 2021.

[7] A. A. Abdullah, R. Khalaf and M. Riza. "A realizable quantum three-pass protocol authentication based on hill-cipher algorithm". *Mathematical Problems in Engineering*, vol. 2015, p. 481824, 2015.

[8] S. Murthy and C. R. Kavitha. "*Preserving Data Privacy in Cloud Using Homomorphic Encryption. In: 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*". IEEE, Piscataway, pp. 1131-1135, 2019.

[9] S. Carpov, T. H. Nguyen, R. Sirdey, G. Constantino and F. Martinelli. "*Practical Privacy-Preserving Medical Diagnosis Using Homomorphic Encryption. In: 2016 IEEE 9th International Conference on Cloud Computing (Cloud)*". IEEE, Piscataway, pp. 593-599, 2016.

[10] A. Acar, H. Aksu, A. S. Uluagac and M. Conti. "A survey on homomorphic encryption schemes: Theory and implementation". *ACM Computing Surveys*, vol. 51, no. 4, pp. 1-35, 2019.

[11] A. V. Kumar, M. S. Sujith, K. T. Sai, G. Rajesh and D. J. S. Yashwanth. "*Secure Multiparty Computation Enabled E-Healthcare System with Homomorphic Encryption. In: IOP Conference Series: Materials Science and Engineering*". IOP Publishing, United Kingdom, p. 022079, 2020.

[12] W. Tang, J. Ren, K. Deng and Y. Zhang. "Secure data aggregation of lightweight E-healthcare IoT devices with fair incentives". *IEEE Internet Things Journal*, vol. 6, no. 5, pp. 8714-8726, 2019.

[13] J. Scheibner, J. L. Raisaro, J. R. Troncoso-Pastoriza, M. Ienca, J. Fellay, E. Vayena and J. P. Hubaux. "Revolutionizing medical data sharing using advanced privacy-enhancing technologies: Technical, legal, and ethical synthesis". *Journal Medical Internet Research*, vol. 23, no. 2, p. e25120, 2021.

[14] W. Guo, J. Shao, R. Lu, Y. Liu and A. A. Ghorbani. "A privacy-preserving online medical prediagnosis scheme for cloud environment," *IEEE Access*, vol. 6, pp. 48946-48957, 2018.

[15] M. M. Anwar and A. F. Salman. "A new fully homomorphic cryptosystem based on a super-increasing sequence". *Telfor Journal*, vol. 12, no. 1, pp. 50-55, 2020.

[16] K. Sinha, P. Majumder and S. K. Ghosh. "*Fully Homomorphic Encryption Based Privacy-Preserving Data Acquisition and Computation for Contact Tracing. In: 2020 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*". IEEE, Piscataway, pp. 1-6, 2020.

[17] M. A. Mohammed and F. S. Abed. "An improved fully homomorphic encryption model based on N-primes". *Kurdistan Journal of Applied Research*, vol. 4, no. 2, pp. 40-49, 2019.

[18] O. Mendelevitch and M. D. Lesh. “Fidelity and privacy of synthetic medical data”. 2021. *arXiv*: arXiv:2101.08658.

[19] M. Giuffrè and D. L. Shung. “Harnessing the power of synthetic data in healthcare: Innovation, application, and privacy”. *NPJ Digital Medicine*, vol. 6, no. 1, p. 186, 2023.

[20] E. Mollakuqe, A. Parduzi, S. R. Orcid, V. Dimitrova, S. J. R. Muharremi, M. H. Orcid and J. Qarkaxhija. “Applications of homomorphic encryption in secure computation”. *Open Research Europe*, vol. 4, no. 158, p. 158, 2024.

[21] J. Scheibner, M. Ienca and E. Vayena. “Health data privacy through homomorphic encryption and distributed ledger computing: An ethical-legal qualitative expert assessment study”. *BMC Medical Ethics*, vol. 23, no. 1, p. 121, 2022.

[22] M. A. Mohammed and F. S. Abed. “Cloud storage protection scheme based on fully homomorphic encryption”. *ARO- Science Journal Koya Univercity*, vol. 8, no. 2, pp. 40-47, 2020.

[23] T. Al Attar and M. A. Mohammed. “Fully homomorphic encryption scheme for securing cloud data”. *UHD Journal Science Technology*, vol. 7, no. 2, pp. 40-49, 2023.

[24] M. Hernandez, G. Epelde, A. Alberdi, R. Cilla and D. Rankin, “Synthetic data generation for tabular health records: A systematic review”. *Neurocomputing*, vol. 493, pp. 28-45, 2022.

[25] J. Walonoski, M. Kramer, J. Nichols, A. Quina, C. Moesel, D. Hall, C. Duffett, K. Dube, T. Gallagher and S. McLachlan. “Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record”. *Journal of the American Medical Information Association*, vol. 25, no. 3, pp. 230-238, 2018.

[26] “*RSA Algorithm in Cryptography*”. Available from: https://www.geeksforgeeks.org/rsa-algorithm-cryptography [Last accessed on 2025 Jul 17].

[27] L. E. Bassham 3rd, A. Rukhin, J. Soto, J. Nechvatal, M. Smid, E. Barker, S. Leigh, M. Levenson, M. Vangel, D. Banks, A. Heckert, J. Dray and S. Vo. "*A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*". NIST Special Publication 800-22 Rev. 1a. National Institute of Standards and Technology, Gaithersburg, MD, 2010.

[28] E. B. Barker and J. M. Kelsey. “*Recommendation for Random Number Generation Using Deterministic Random Bit Generators (Revised)”*. US Department of Commerce, Technology Administration, National Institute, United States, 2007.

[29] “*SecureRandom (Java Platform SE 8)*”. Available from: https://docs.oracle.com [Last accessed on 2025 Jul 25].

[30] “MSC63-J. *Ensure that SecureRandom is properly seeded - SEI CERT Oracle Coding Standard for Java - Confluence*”. Available from: https://wiki.sei.cmu.edu/confluence [Last accessed on 2025 Jul 25].

[31] “*GitHub - Syntheticheal th/Synthea: Synthetic Patient Population Simulator*”. Available from: https://github.com/syntheticheal th/synthea [Last accessed on 2025 Jun 16].

[32] B. F. Gage, A. D. Waterman, W. Shannon, M. Boechler, M. W. Rich and M. J. Radford. “Validation of clinical classification schemes for predicting stroke: Results from the national registry of atrial fibrillation”. *JAMA*, vol. 285, no. 22, pp. 2864-2870, 2001.

[33] P. S. Goldman and M. D. Ezekowitz. “*Anticoagulation in atrial arrhythmias: Current therapy and new therapeutic options. In: Electrophysiological Disorders of the Heart”*. Elsevier, Netherlands, pp. 1175-1180, 2012.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Molecular Identification and Antibiotic Resistance Profile of Some *Pseudomonas aeruginosa* Clinical Isolates

Darun B. Mohammed[1], Ayad H. Hasan[1,2]

[1]Department of Medical Microbiology, Faculty of Science and Health, Koya University, Koya 44023, Kurdistan Region – F.R. Iraq, [2]Department of Biomedical Sciences, Cihan University-Erbil, Erbil, Kurdistan Region, Iraq

ABSTRACT

Pseudomonas aeruginosa is a Gram-negative, opportunistic bacterium being increasingly recognized as the causative agent of hospital-acquired infection, especially in immunocompromised patients. The bacterium is well known for its environmental persistence and multidrug resistance (MDR). This study aimed to characterize the antibacterial persistence profiles and genetic diversity of *P. aeruginosa* isolates from clinical settings in Sulaymaniyah city, Iraq. Twenty-eight suspected *P. aeruginosa* isolates were collected from hospitals and private laboratories from October 2024 to January 2025. The collected bacteria were identified with standard microbiological procedures, the VITEK 2 system, and confirmation through 16S RNA sequencing. Ten antibiotics were tested following the guidelines of the Clinical and Laboratory Standards Institute for antibiotic susceptibility testing. 12 out of 28 collected isolates were confirmed as *P. aeruginosa*. The antimicrobial susceptibility testing indicated that resistance to Imipenem, Ceftazidime, and Cefepime was seen in 66.7% of the isolates (MDR isolates), while Ceftolozane/Tazobactam had the lowest resistance rate (41.7%). It is observed that 66.7% of isolates subjected to MDR show resistance to three or more antibiotic classes. There is a high prevalence of *P. aeruginosa* in clinical isolates that are resistant to antibiotics. These results underscore the urgent need for improved antimicrobial stewardship programs and the development of alternative treatment options to address this rising public health concern. Through media genomics and molecular methods, reliable identification has been enhanced, which signifies the importance of both studies.

Index Terms: *Pseudomonas Aeruginosa,* Antibiotic Resistance Profile, Multidrug Resistance, 16S rRNA Sequencing, Polymerase Chain Reaction

1. INTRODUCTION

Pseudomonas aeruginosa is a Gram-negative, non-fermenting bacterium known for its adaptability and environmental persistence [1]. *P. aeruginosa* has a large genome (5.5–7 Mbp), enabling significant genetic adaptability [2]. This versatility allows the bacterium to thrive in diverse environments, including human, animal, and plant hosts, as well as non-living reservoirs such as water and soil [3], [4]. Despite its environmental persistence, *P. aeruginosa* is rarely found in the normal microbiota of healthy individuals [5]. *This bacterium* is a leading cause of nosocomial infections, contributing to ventilator-associated pneumonia, bloodstream infections, urinary tract infections, and wound infections, particularly in burn and intensive care unit patients [6]–[11]. Its ability to persist in hospital environments makes it a significant public health concern. The increasing antibiotic resistance of *P. aeruginosa* poses a major clinical challenge, and it employs multiple resistance mechanisms, including intrinsic

Access this article online

DOI: 10.21928/uhdjst.v9n2y2025.pp92-100

E-ISSN: 2521-4217
P-ISSN: 2521-4209

Copyright © 2025 Darun B. Mohammed and Ayad H. Hasan. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Ayad Hazim Hassan, Department of Medical Microbiology, Faculty of Science and Health, Koya University, Koya 44023, Kurdistan Region – F.R. Iraq. Email: ayad.hasan@koyauniversity.org

Received: 30-05-2025 Accepted: 30-07-2025 Published: 26-08-2025

resistance (low outer membrane permeability), acquired resistance (horizontal gene transfer [HGT] and mutations), and adaptive resistance (biofilm formation) [12]–[14]. *P. aeruginosa* pathogenicity is attributed to the synthesis of various virulence factors, including pyocyanin, rhamnolipids, elastase, exotoxin A, phospholipase C, and secretion systems such as the type III secretion system [15]. These factors contribute to persistence, antibiotic resistance, and evasion of host defence mechanisms [16]. As a result, *P. aeruginosa* has developed co-resistance to multiple antibiotics, increasing its potential to cause severe and potentially fatal infections [17], and consequently, it has been classified as one of the three bacterial species on the World Health Organization's list of "priority pathogens," urgently requiring the development of new antibiotics [18].

To ensure uniform resistance classification across laboratories, *multidrug resistance* (MDR) is defined as acquired non-susceptibility to at least one agent in three or more antimicrobial categories [19]. Identification and understanding of the strain diversity of MDR *P. aeruginosa* is crucial due to its ability to cause serious infections and its increasing resistance to multiple antibiotics.

Although *P. aeruginosa* has been studied on a large scale, there is still limited data on its antibiotic resistance patterns and genetic diversity in clinical settings in Sulaymaniyah, Iraq. A recent study in Sulaimanyah city by Seenaa *et al.*, in 2024 showed a prevalence of antibiotic resistance among *P. aeruginosa* isolated from the hospital environment. The resistance rates among the 26 *P. aeruginosa* isolates were as follows: 23.3% to streptomycin, 13.6% to tobramycin, 22.7% to moxifloxacin, 21.2% to levofloxacin, and 22.7% to norfloxacin, as highlighted by Ali *et al.* [20]. In addition, relying only on phenotypic identification methods can sometimes lead to misidentification. Molecular techniques, such as 16S rRNA sequencing, offer greater accuracy in bacterial classification. This study aims to address this gap by integrating both conventional and molecular identification methods to examine the resistance profiles of *P. aeruginosa* isolates from clinical samples.

2. MATERIALS AND METHODS

2.1. Sample Collection

A total of 28 bacterial isolates suspected to be *P. aeruginosa* were collected from Sulaimany Burn, Plastic and Reconstructive Surgery Hospital, Dr. Jamal Ahmad Rashid's pediatric teaching Hospital, and a private medical laboratory in Sulaymaniyah Drastic, Kurdistan region, Iraq, between October 1st, 2024, and January 1st, 2025. These bacteria were isolated from clinical samples initially collected and cultured by certain hospitals and laboratories.

2.2. Isolation and Identification of *P. aeruginosa*

2.2.1. Isolation

To obtain a pure bacterial colony, the collected samples underwent a cultivation process. The isolated bacterial strains were subcultured on a Nutrient agar plate and MacConkey agar (Liofilchem), and all plates were incubated at 37°C for 24 h. For the primary isolation, the suspected non-lactose fermented colonies grown on MacConkey agar medium were cultured on the Cetrimide agar medium (Neogen) at 37°C for 24–48 h, which is the recommended medium for the primary identification and isolation of *P. aeruginosa* [21].

2.2.2. Identification

The isolates were subjected to some physiological and biochemical tests. They were identified using conventional biochemical tests, including the Gram staining, oxidase, motility, catalase, citrate utilization, and triple sugar iron (TSI) agar tests [4], [22]. In addition, the VITEK-2 system was employed to confirm the results of the conventional identification methods [23].

2.3. Molecular Identification

2.3.1. Broth culture preparation and genomic deoxyribonucleic acid (DNA) extraction

To extract genomic DNA from representatives of each bacterial group, a loopful of specific bacteria was inoculated into a 15 mL Falcon tube containing 5 mL nutrient broth. The culture was then incubated with shaking at 150 rpm. for 24 h at 37°C. DNA was extracted using the AddPrep genomic DNA extraction kit (addbio) according to the manufacturer's guidelines. The quantity and purity of the extracted DNA were assessed using a Nanodrop (Thermo Scientific NanoDrop 2000, SN 6113). The extracted DNA was then stored at −20°C. For further analysis, 82 ng of genomic DNA was loaded onto a 1% agarose gel and run for 60 min at 80 V.

2.3.2. Amplification of 16S rRNA by standard polymerase chain reaction (PCR)

A ~1515 bp fragment of 16S rRNA was amplified using a PCR method with a final reaction volume of 30 μL. This included 15 μL of 2X Add Taq Master (Ampliqon), 1 μL (5 pmol) of each forward primer (P1F-TGAAGAGTTTGATCATGGCTCAG) and reverse primer (P1R-TTCCCCTACGGTTACCTTGT) [23], and 1 μL

(40 ng) of genomic DNA. The volume was completed by adding 12 µL of nuclease-free water. The PCR was performed using BIO-RAD and Corbett thermal cyclers, configured as follows for VITEK-2-confirmed samples: initial denaturation at 95°C for 5 min, then 27 cycles of 30 s at 95°C for denaturing, 27 s for annealing at 58°C, 1 min for the extension at 72°C, and 5 min for the final extension at 72°C.

2.3.3. Agarose gel electrophoresis

To verify the presence of the correct PCR amplification, 2 µL of the amplicons were electrophoresed on a 1% agarose gel containing 0.07% EtBr, alongside a 100 bp DNA ladder (Genedrix). The gel was run in 1X Tris-borate-ethylenediaminetetraacetic acid buffer at 80 V for 60 min to confirm the proper amplification of the targeted gene. After the run, the DNA molecules were visualized and photographed using a ultraviolet Gel Imager (SynGene 1409).

2.3.4. Partial 16S rRNA sequencing

The resulting approximately 1515 bp PCR amplicon was sent to Macrogen Inc., a South Korean company, for sequencing using P2R (TCTAATCCTGTTTGCTCCCCA) [24], to include the V4 region.

2.3.5. Quality of the sequenced products

The DNA base assembler program was used for sequence quality analysis and editing. The beginning and end of the sequence were trimmed to evaluate its quality.

3. RESULTS

3.1. Isolation and Identification of *P. aeruginosa*

After isolation, the characteristics of the isolates were examined by culturing them on two media: MacConkey agar and Cetrimide agar. The isolates appeared as non-lactose fermenting colonies, producing greenish-blue colonies. In addition, they emitted a characteristic fishy odor and were identified as Gram-negative rods. Of the 28 samples, only 20 isolates grew on Cetrimide agar, which is a selective medium for the isolation and identification of *P. aeruginosa*. Further confirmation of these isolates was carried out using the biochemical tests and VITEK2 system.

3.2 Biochemical Test for Bacterial Isolates

P. aeruginosa was positive for the oxidase, catalase, and motility tests, shows an alkaline slant and alkaline butt on TSI (no acid or gas production), and is positive for the Simmons citrate test (utilizes citrate) [4]. Of the 20 samples isolated from Cetrimide agar, only 15 isolates were positive for oxidase, catalase, citrate utilization, and motility tests, and also appeared as an alkaline slant and alkaline butt on TSI. These bacteria were further conferred by VITEK2 system.

3.3. VITEK2- System

P. aeruginosa was identified in only 12 out of the 20 isolates, as detailed in Tables 1 and 2. *P. aeruginosa* isolates, denoted by their (A numbers).

The analysis of the 12 *P. aeruginosa* isolates revealed distinct patterns of antibiotic susceptibility, classifying the samples into three categories: MDR, sensitive, and intermediate (mixed resistance). Eight samples (A1, A3, A4, A5, A9, A10, A11, and A12) were identified as MDR, showing resistance to antibiotics from at least 3 different classes [25]. These isolates represent a significant concern due to the limited treatment options available for infections caused by such strains. In contrast, three samples (A2, A6, and A8) were

TABLE 1: VITIK-2 system's result for confirmation of *P. aeruginosa* identification

Sample ID	Expectation	Growth on cetrimide agar	VITEK result
PA1	*P. aeruginosa*	Positive	*P. aeruginosa*
PA2	*P. aeruginosa*	Positive	*Enterobacter cloacae complex*
PA3	*P. aeruginosa*	Negative	
PA4	*P. aeruginosa*	Positive	*Proteus mirabilis*
PA5	*P. aeruginosa*	Negative	
PA6	*P. aeruginosa*	Positive	*P. aeruginosa*
PA7	*P. aeruginosa*	Positive	*P. aeruginosa*
PA8	*P. aeruginosa*	Positive	*P. aeruginosa*
PA9	*P. aeruginosa*	Negative	
PA10	*P. aeruginosa*	Positive	*Morganella morganii*
PA11	*P. aeruginosa*	Negative	
PA12	*P. aeruginosa*	Positive	*Proteus mirabilis*
PA13	*P. aeruginosa*	Negative	
PA14	*P. aeruginosa*	Positive	*P. aeruginosa*
PA15	*P. aeruginosa*	Positive	*Morganella morganii*
PA16	*P. aeruginosa*	Positive	*P. aeruginosa*
PA17	*P. aeruginosa*	Positive	*Escherichia coli*
PA18	*P. aeruginosa*	Positive	*P. aeruginosa*
PA19	*P. aeruginosa*	Positive	*P. aeruginosa*
PA20	*P. aeruginosa*	Positive	*P. aeruginosa*
PA21	*P. aeruginosa*	Positive	*P. aeruginosa*
PA22	*P. aeruginosa*	Positive	*Morganella morganii*
PA23	*P. aeruginosa*	Negative	
PA24	*P. aeruginosa*	Positive	*Enterobacter cloacae complex*
PA25	*P. aeruginosa*	Negative	
PA26	*P. aeruginosa*	Negative	
PA27	*P. aeruginosa*	Positive	*P. aeruginosa*
PA28	*P. aeruginosa*	Positive	*P. aeruginosa*

P. aeruginosa: Pseudomonas aeruginosa

TABLE 2: List of confirmed *Pseudomonas aeruginosa* isolates, identified by A numbers

PA	A
PA1	A1
PA6	A2
PA7	A3
PA8	A4
PA14	A5
PA16	A6
PA18	A7
PA19	A8
PA20	A9
PA21	A10
PA27	A11
PA28	A12

found to be fully sensitive to all tested antibiotics, indicating a favourable susceptibility profile. These isolates are expected to respond well to conventional treatments. Only one sample (A7) displayed a limited resistance pattern, showing a combination of resistance and sensitivity across different antibiotics. Sample A7 exhibited resistance to two antibiotics, but remained sensitive to the remaining antibiotics, showing resistance to 2 classes as outlined in Tables 3 and 4.

3.4. Bacterial Identification at the Molecular Level

3.4.1. Genomic DNA extraction

The bacterial sample number from which the genomic DNA was extracted is indicated by lanes 1–12; M; Genedirex 1 kb DNA marker, C–; negative control that was performed without a DNA template (Fig. 1). The wells contain high molecular weight and genomic DNA with yields of on average 82 µg and average purities of 1.85 at (A260/A280) (Table 5). No degradation was observed in all samples. The purity of the DNA was confirmed by the negative control result.

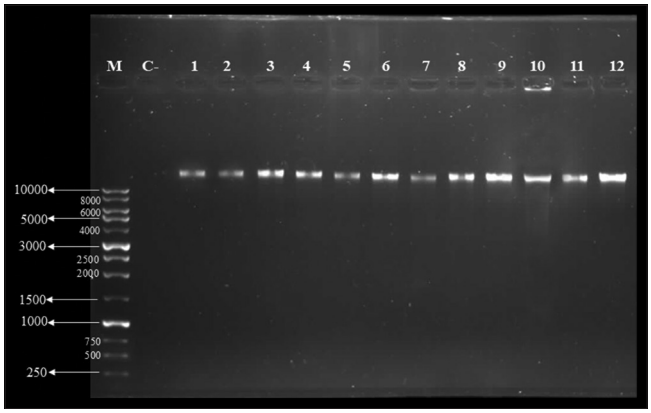


Fig. 1. Agarose gel electrophoresis analysis of genomic deoxyribonucleic acid (DNA) extracted from *Pseudomonas aeruginosa* isolates. From left to right: M is the Genedirex 1 kb DNA marker, C-is the negative control (reaction without DNA template), and lanes 1–12 are the bacterial samples from which genomic DNA was extracted.

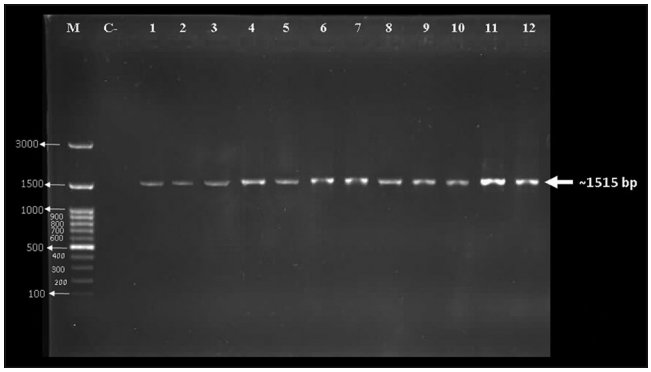


Fig. 2. 16S rRNA partial amplification with P1F and P1R primers. Lanes M and C- represent a 100 bp deoxyribonucleic acid (DNA) marker (Genedirex) and a negative control without a DNA template, respectively. Approximately 1515 bp polymerase chain reaction amplicons produced using DNA templates from A1 through A12 are displayed in Lanes 1–12, respectively.

3.4.2. Amplification of 16S rRNA by standard PCR

The bacterial isolate template DNA was successfully amplified to the expected size of the DNA fragment (~1515 bp), while the negative controls showed no PCR results (Fig. 2).

3.4.3. Partial 16S rRNA sequencing

To identify the species of each bacterial isolate, the reverse primer P2R, which targets the V4 region, was used to sequence the 16S rRNA gene PCR amplicons generated from each isolate using P1F and P1R primers, because the V4 region of the 16S rRNA gene, a hypervariable region, is useful for bacterial identification and phylogenetic analysis because it provides high-resolution differentiation of microbial species. To specify the taxonomic origin of the twelve 16S rRNA sequences, independent computerized alignments were performed versus accessible prokaryotic sequences of 16S rRNA employing EzBioCloud [26]. All of the twelve sequences were highly comparable to reference strains of *P. aeruginosa*, ranging from 99% to 100%. Each isolate had top alignment hits only with *P. aeruginosa*, with query coverage and identity scores compatible with species identification. These molecular results demonstrate that all twelve isolates were *P. aeruginosa* detailed in (Table 6).

3.4.4. Quality of the sequenced products

Depending on the reliability of the DNA sequencing, the 12 *P. aeruginosa* isolates were confirmed to have high-quality values, all exceeding 40, as presented in Table 7 and Fig. 3.

TABLE 3: Antibiotic susceptibility profiles of 12 *Pseudomonas aeruginosa* isolates using the VITEK2 antibiotic panel

Sample ID	VITEK antibiotics panel against *Pseudomonas aeruginosa*									
	AK	GEN	PTZ	IMP	MRP	CIP	CAZ	CPM	CZA	C/T
A1	R	R		R	R	R	R	R	S	R
A2	S	S	S	S	S	S	S	S	S	S
A3	R	R	R	R	R	R	R	R	R	R
A4	R	R	R	R	R	R	R	R	R	R
A5	R	R	R	R	R	R	R	R	S	R
A6	S	S	S	S	S	S	S	S	S	S
A7	S	S	S	R	S	R	S	S	S	S
A8	S	S	S	S	S	S	S	S	S	S
A9	S	S	R	S	S	S	R	R	S	S
A10	R	R	R	R	R	S	R	R	R	R
A11	R	R	R	R	R	R	R	R	R	R
A12	R	R	R	R	R	R	R	R	R	R

R: Resistant, S: Sensitive, I: Intermediate, AK: Amikacin, GEN: Gentamicin, PTZ: Piperacillin/Tazobactam, IMP: Imipenem, MRP: Meropenem, CIP: Ciprofloxacin, CAZ: Ceftazidime, CPM: Cefepime, CZA: Ceftazidime/Avibactam, C/T: Ceftolozane/Tazobactam

TABLE 4: Antibiotic class resistance and susceptibility classification of *Pseudomonas aeruginosa* strains

Sample ID	No. of antibiotics resistant	No. of classes resistant	Susceptibility profile
A1	8	5	MDR
A2	0	0	Sensitive
A3	10	6	MDR
A4	10	6	MDR
A5	9	6	MDR
A6	0	0	Sensitive
A7	2	2	Intermediate
A8	0	0	Sensitive
A9	3	3	MDR
A10	9	5	MDR
A11	10	6	MDR
A12	10	6	MDR

TABLE 5: Quantity and purity of the extracted DNA by nanodrop

	DNA integrity by nanodrop	
Sample ID	A260/A280	Concentration (ng/μL)
A1	1.931	28.0
A2	1.951	28.0
A3	1.861	33.5
A4	1.869	28.5
A5	1.838	34.0
A6	1.841	40.5
A7	1.882	48.0
A8	1.808	47.0
A9	1.862	60.5
A10	1.831	59.5
A11	1.804	46.0
A12	1.809	42.5

DNA: Deoxyribonucleic acid

TABLE 6: Identification of selected isolates by molecular approaches

Sample ID	Suspicion	Per identity (%)	Accession no
A1	*P. aeruginosa*	100	MN606210
A2	*P. aeruginosa*	100	PP515656
A3	*P. aeruginosa*	99.01	CP142446
A4	*P. aeruginosa*	99.58	CP143908
A5	*P. aeruginosa*	99.86	OR452246
A6	*P. aeruginosa*	99.58	ON721331
A7	*P. aeruginosa*	99.73	MG396991
A8	*P. aeruginosa*	100	PP515656
A9	*P. aeruginosa*	100	EF556270
A10	*P. aeruginosa*	99.86	MF144483
A11	*P. aeruginosa*	100	PP515656
A12	*P. aeruginosa*	100	PP515656

P. aeruginosa: Pseudomonas aeruginosa

3.5. Statistical Analysis of Antibiotic Resistance Patterns

According to the level of antibiotic resistance in each bacterial isolate, the bacterial isolates were classified. MDR isolates exhibited resistance to 9.33 distinct antibiotics on average, with a range of 8–10 and a slight variation (standard deviation [SD]: 0.78), as illustrated in Table 8). The susceptible group(s) exhibited no resistance at all, whereas the intermediate-resistant strains were only able to resist two antibiotics (mean: 0.00, SD: 0.00). This analysis demonstrates the differences in how various bacterial strains react to antibiotic therapy. Using a Chi-square test of independence, we found a significant difference in resistance patterns between antibiotics ($\chi^2 = 32.14$, df = 9, $P < 0.001$), indicating that resistance patterns differed considerably between sources (Table 9).

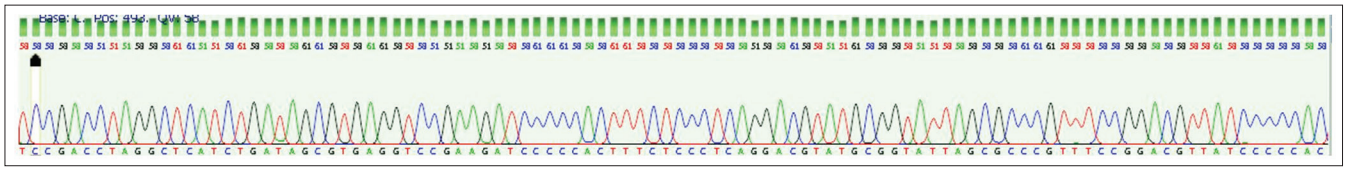


Fig. 3. An example of the 16s rDNA sequence. The quality value of a single base is shown as a numerical value beneath the top green bars. The peaks are evenly spaced, and there is no noise. The baseline remains flat and constant throughout the whole sequence chromatogram. The base is called precisely under each peak (DNA Sequence Assembler v4 (2013), Heracle BioSoft, www.DnaBaser.com) [27].

TABLE 7: Quality value of the sequenced result

Sample ID	Bases left after trimming (%)	Average quality before trimming	Average quality after trimming	Trusted bases (QV >26, %)	Quality remarks
A1	682 (45)	29	55	99.1	Very good
A2	697 (45)	27	56	99.3	Very good
A3	571 (38)	20	42	84.2	Very good
A4	696 (43)	28	53	98.6	Good trust
A5	723 (44)	24	43	92.3	Very good
A6	688 (44)	28	54	98.5	Very good
A7	1159 (94)	35	39	85.6	Good
A8	724 (46)	28	56	99.2	Very good
A9	715 (46)	31	56	99.9	Very good
A10	725 (47)	28	55	98.3	Very good
A11	713 (43)	29	57	99.9	Very good
A12	714 (44)	28	57	100	Very good

TABLE 8: Mean, standard deviation, and range of resistance levels within each susceptibility category (MDR, sensitive, intermediate)

Susceptibility profile	Mean resistant antibiotics	Min	Max	Standard deviation
MDR	9.33	8	10	0.78
Sensitive	0.00	0	0	0.00
Intermediate	2.00	2	2	0.00

MDR: Multidrug resistant

TABLE 9: The Chi-square test to assess the antibiotic resistance differs

Chi-square value	Degrees of freedom (df)	*P*-value
32.14	9	<0.001

3.5.1. Correlation analysis of antibiotic resistance

Using a correlation matrix, the relationships between antibiotic resistance were assessed. Positive correlations (Strong Associations) were found between meropenem (MRP) and imipenem (IMP) (0.92). If a strain demonstrates resistance to one antibiotic, it is highly probable that it will exhibit resistance to other antibiotics as well. Gentamicin (GEN) and Amikacin (AK) (0.87) were strongly linked. Ciprofloxacin (CIP) and Ceftazidime (CAZ) (0.79) were linked to showing resistance as well. CAZ/Avibactam (CZA) showed weak correlations with the other antibiotics, which suggests that CZA might be effective against MDR *P. aeruginosa* that is resistant to other antibiotics (Table 10). Furthermore, to demonstrate the relationships between patterns of antibiotic resistance, a heatmap was generated (Fig. 4).

A strong correlation between AK and GEN (0.87), which reflects shared aminoglycoside resistance mechanisms, and CIP and CAZ (0.79), which may indicate plasmid-mediated co-resistance. MRP and CZA (0.92), which is a striking association that raises concerns about cross-resistance in carbapenem-resistant infections, possibly due to metallo-β-lactamases. However, no correlations (0.00) were observed for other combinations, suggesting independent resistance mechanisms (Fig. 4).

4. DISCUSSION

This study was conducted to assess the prevalence and antibiotic resistance profiles of *P. aeruginosa* isolates from clinical cases in Sulaymaniyah. 20 (71.4%) of the 28 isolates were found to grow on cetrimide agar, a selective medium used for *Pseudomonas aeruginosa* [28]. The results confirmed that although the medium is selective, more accurate techniques such as the VITEK2 automated system and 16S *rRNA* gene sequencing should be employed for identification. Nonetheless, the VITEK2 system only

TABLE 10: Antibiotic resistance correlations among *Pseudomonas aeruginosa* isolates. Strong positive correlations indicate possible co-resistance between specific antibiotic classes, while weak or negative correlations suggest independent resistance patterns

Correlation type	Antibiotic pair	Correlation coefficient (r)	Interpretation
Positive (Strong)	MRP and IMP	0.92	Strong co-resistance to carbapenems
	GEN and AK	0.87	Strong association between aminoglycoside resistance
	CIP and CAZ	0.79	Linked resistance between fluoroquinolone and cephalosporin
Negative or Weak	CZA versus others	Mostly low or negative	Suggests CZA might remain effective against MDR strains

MRP: Meropenemm, IMP: Imipenem, GEN: Gentamicin, AK: Amikacin, CIP: Ciprofloxacin, CAZ: Ceftazidime CZA: Ceftazidime/Avibactam, MDR: Multidrug resistance

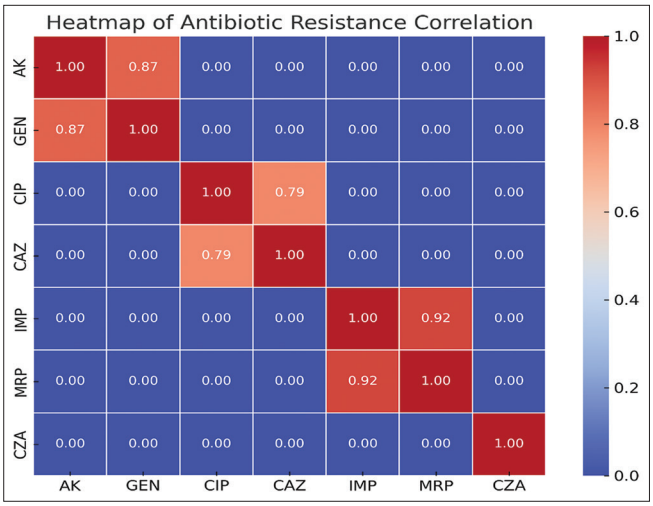


Fig. 4. Heatmap of correlation coefficients between antibiotic resistance patterns among 12 clinical isolates of *Pseudomonas aeruginosa*. The strength of correlation (Pearson's r) is color-coded, with red indicating strong positive correlations and blue indicating weak or no correlation. Notable associations include Meropenem and Imipenem (r = 0.92), and Gentamicin and Amikacin (r = 0.87). Ceftazidime/Avibactam (CZA) showed low or negative correlation with others, indicating potential effectiveness against multidrug-resistant strains.

confirmed 12 as *P. aeruginosa*, indicating a difficulty in the identification of non-fermenting Gram-negative bacteria. Results from the VITEK2 System misidentified many isolates. Possible phenotypic overlap is likely due to similar appearances of *Pseudomonas* spp. and other Gram-negative bacteria, including *Enterobacter cloacae* and *Proteus mirabilis*. This emphasizes the significance of 16S RNA as an identification tool [21], [29], [30]. The differences in identification demonstrated the need for ongoing training and updates on laboratory procedures as bacterial profiles continue to evolve.

The confirmed *P. aeruginosa* isolates showed a tendency towards MDR in the antibiotic susceptibility profiles. Eight MDR isolates were found to be resistant to all (66.7%) and pose a great challenge for antibiotic treatment options. The findings of the current study are aligned with several regional studies that report high levels of MDR. For example, the study by Alkhulaifi and Mohammed reported 72.63% MDR in Basra, Iraq [31]. Similarly, Li *et al.*, reported a rate of 87.5% in Iran and 55.9% in Turkey [32]. As per Al-Orphaly *et al.*, the MDR rates of the neighbouring countries Bahrain (86%), Egypt (75.6%), Lebanon (64.5%), and Jordan (52.5%), were also high. Similar findings indicated that similar causes may exist for the globally common *P. aeruginosa* resistance in the Middle East, such as antibiotic consumer behavior and medical practices [33]. Lower resistance rates were also reported at 17.6% in Thailand, 17.7% in the US, and 22.2% in Croatia [32]. In Europe, the total prevalence estimated of MDR is 29.9 [32]. The fairly low numbers could be a result of more controlled prescription practices as well as better hospital hygiene standards and stronger antibiotic stewardship policies [34]. Statistical analysis supported these findings. Most MDR isolates showed resistance to at least 9 antibiotics. The factors that may be responsible for the high occurrence of antibiotic resistance observed in this study are interrelated. A major contributor to antibiotic resistance is the indiscriminate use of antibiotics, which refers to the overprescribing of these drugs by health care providers and self-medication practices in the community. Excessively using antibiotics forces the bacteria to mutate. Besides, the lack of public knowledge about the proper use of antibiotics–like not finishing the prescribed course or using them for viral infections–is a contributor to the emergence and spread of resistance. The problem is made worse by the suboptimal infection control in the clinical and community settings that favour the transmission of resistant pathogens. The increase in drug resistance, the lack of new antibiotic discoveries, and the worldwide public health burden imposed by bacterial resistance jointly call for antibiotic stewardship programs and education, and monitoring.

The resistance patterns were not randomly distributed, as the Chi-square test results show dependency on the antibiotic class. Carbapenems (MRP and IMP) and aminoglycosides (GEN and AK) showed a high correlation, indicating potential

for the shared mechanism of resistance or common pressure. On the other hand, CAZ/Avibactam's weak correlations with other antibiotics indicate a potential efficacy against MDR strains, allowing for successful treatment. In hospitals, patients are frequently immunocompromised [35], [36]. The resistance patterns or intermediate resistance showed (1 isolate) 8.3% further complicating the treatment regimens [37]. Surprisingly, 3 isolates (%25) exhibited complete sensitivity to all antibiotics, suggesting that not all *P. aeruginosa* are resistant. As for the subject matter experts, this finding is quite important. It shows that effective treatment options are still available for some strains. Thus, the clinicians must focus on precise identification and susceptibility testing. The identification of sensitive strains can assist clinicians in determining the correct choices of therapy. Molecular identification of the isolates using 16S rRNA sequencing gave a firm confirmation of the species and showed genetic diversity among the isolates [38]. This study highlights the importance of molecular methods in their diagnostic application in antibiotic resistance characterization. Testing organisms in culture often do not have the same sensitivity as molecular methods and may miss bacterial pathogens with even high relative abundance. On the other hand, the molecular approach (e.g., 16S *rRNA* gene sequencing) improves the accuracy of identifications. The reliability of these methods is further confirmed by the high percentage of sequence identity with established *P. aeruginosa* reference strains. In addition, it was found that genomic analysis of *P. aeruginosa* is important to elucidate the evolution of infection epidemiology and possible HGT mechanisms involved in the transmission of antibiotic resistance genes [39].

5. CONCLUSION

This study investigated the antibacterial resistance profiles and genetic diversity of Pseudomonas aeruginosa. A notable prevalence of multidrug resistance (MDR) was detected in eight (66.7%) of the isolates, with four isolates (33.3%) exhibiting resistance to all tested antibiotics. Resistance against imipenem, ceftazidime, and cefepime was at its highest level (58.3%), while ceftolozane/tazobactam showed the lowest resistance rate (41.7%). A significant correlation was identified between carbapenems and aminoglycosides, suggesting the involvement of similar resistance mechanisms. Conversely, the weak correlations of CAZ/avibactam with other antibiotics indicated its potential efficacy against MDR strains, which could facilitate successful therapeutic interventions. Furthermore, three isolates (25%) showed complete susceptibility to all antibiotics tested, indicating that not all P. aeruginosa strains are resistant and that viable treatment options remain available for certain strains. Thus, clinicians must prioritise accurate identification and susceptibility testing. This study highlights the necessity for refined methodologies to identify antibiotic resistance, including the VITEK2 automated system and molecular techniques using 16S rRNA gene sequencing.

REFERENCES

[1] N. Sathe, P. Beech, L. Croft, C. Suphioglu, A. Kapat and E. Athan. "*Pseudomonas aeruginosa*: Infections and novel approaches to treatment "Knowing the enemy" the threat of *Pseudomonas aeruginosa* and exploring novel approaches to treatment". *Infectious Medicine (Beijing)*, vol. 2, no. 3, pp. 178-194, 2023.

[2] F. Giovagnorio, A. De Vito, G. Madeddu, S. G. Parisi and N. Geremia. "Resistance in *Pseudomonas aeruginosa*: A narrative review of antibiogram interpretation and emerging treatments". *Antibiotics*, vol. 12, no. 11, pp. 1621-1621, 2023.

[3] P. Pachori, R. Gothalwal and P. Gandhi. "Emergence of antibiotic resistance *Pseudomonas aeruginosa* in intensive care unit; a critical review". *Genes and Diseases*, vol. 6, no. 2, pp. 109-119, 2019.

[4] M. Al-Saffar and E. Jarallah. "ISOLATION AND CHARACTERIZATION OF PSEUDOMONAS AERUGINOSA FROM BABYLON PROVINCE." *Biochemical and Cellular Archives*. vol. 19, no. 1, pp. 203-209, 2019.

[5] K. Telling, M. Laht, A. Brauer, M. Remm, V. Kesand, M. Maimetis, T. Tenson and I. Lutsar. "Multidrug resistant *Pseudomonas aeruginosa* in Estonian hospitals". *BMC Infectious Diseases*, vol. 18, no. 1, p. 513, 2018.

[6] Y. W. Yang, Y. Z. Jiang, C. M. Hsu and L. W. Chen. "*Pseudomonas aeruginosa* ventilator-associated pneumonia induces lung injury through TNF-α/c-Jun NH2-terminal kinase pathways". *PLoS One*, vol. 12, no. 1, p. e0169267, 2017.

[7] J. A. Lopes, A. D. Rghei, B. Thompson, L. Susta, C. M. Khursigara and S. K. Wootton. "Overcoming barriers to preventing and treating *P. aeruginosa* infections using AAV vectored immunoprophylaxis". *Biomedicines*, vol. 10, no. 12, p. 3162, 2022.

[8] J. Newman, R. Floyd and J. Fothergill. "Invasion and diversity in *Pseudomonas aeruginosa* urinary tract infections". *Journal of Medical Microbiology*, vol. 71, no. 3, p. 001458, 2022.

[9] F. F. Tuon, L. R. Dantas, P. H. Suss and V. S. Tasca Ribeiro. "Pathogenesis of the *Pseudomonas aeruginosa* biofilm: A review". *Pathogens*, vol. 11, no. 3, p. 300, 2022.

[10] O. Ozlu and A. Basaran. "Infections in patients with major burns: A retrospective study of a burn intensive care unit". *Journal of Burn Care and Research*, vol. 43, no. 4, pp. 926-930, 2021.

[11] A. Kour, S. Jaglan, S. Sharma and S. Sharma. "A new strategy to treat *Pseudomonas aeruginosa* infected burn wounds: Antimicrobial cocktails as potent topical therapy". *Medical Hypotheses*, vol. 180, pp. 111167-111167, 2023.

[12] L. R. Mulcahy, V. M. Isabella and K. Lewis. "*Pseudomonas aeruginosa* biofilms in disease". *Microbial Ecology*, vol. 68, no. 1, pp. 1-12, 2013

[13] Z. Pang, R. Raudonis, B. R. Glick, T. J. Lin and Z. Cheng. "Antibiotic resistance in *Pseudomonas aeruginosa*: Mechanisms and alternative therapeutic strategies". *Biotechnology Advances*,

vol. 37, no. 1, pp. 177-192, 2019.

[14] M. Fernández-Billón, A. E. Llambías-Cabot, E. Jordana-Lluch, A. Oliver and M. D. Macià. "Mechanisms of antibiotic resistance in *Pseudomonas aeruginosa* biofilms". *Biofilm*, vol. 5, p. 100129, 2023.

[15] K. C. F. De Almeida, M. A. Calomino, G. Deutsch, S. R. De Castilho, G. R. De Paula, L. M. R. Esper and L. A. Teixeira. "Molecular characterization of multidrug-resistant (MDR) *Pseudomonas aeruginosa* isolated in a burn center". *Burns*, vol. 43, no. 1, pp. 137-143, 2017,

[16] F. C. Ugwuanyi, A. Ajayi, D. A. Ojo, A. I. Adeleye and S. I. Smith. "Evaluation of efflux pump activity and biofilm formation in multidrug resistant clinical isolates of *Pseudomonas aeruginosa* isolated from a federal medical center in Nigeria". *Annals of Clinical Microbiology and Antimicrobials*, vol. 20, no. 1, p. 11, 2021.

[17] C. B. Adjei, U. Govinden, K. Moodley and S. Y. Essack. "Molecular characterisation of multidrug-resistant *Pseudomonas aeruginosa* from a private hospital in Durban, South Africa". *Southern African Journal of Infectious Diseases*, vol. 33, no. 2, pp. 38-41, 2017.

[18] J. Gómez-Martínez, R. D. C Rocha-Gracia, E. Bello-López, M.A. Cevallos, M. Castañeda-Lucio, Y. Sáenz, G. Jiménez-Flores, G. Cortés-Cortés, A. López-García and P. Lozano-Zarain. "Comparative genomics of *Pseudomonas aeruginosa* strains isolated from different ecological niches". *Antibiotics*, vol. 12, no. 5, p. 866, 2023.

[19] A. P. Magiorakos, A. Srinivasan, R. B. Carey, Y. Carmeli, M. E. Falagas, C. G. Giske, S. Harbarth, J. F. Hindler, G. Kahlmeter, B. J. C. M. Olsson-Liljequist and D. L. Paterson. "Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: An international expert proposal for interim standard definitions for acquired resistance". *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, vol. 18, no. 3, pp. 268-281, 2012.

[20] S. M. Ali, T. A. H. Soor, G. A. Ahmed, G. A. Mhdin, G. A. Othman and S. M. Faiq. "Distribution and molecular characterization of antibiotic-resistant *Pseudomonas aeruginosa* in hospital settings of Sulaymaniyah, Iraq". *Polish Journal of Microbiology*, vol. 73, no. 4, pp. 467-473, 2024.

[21] A. Gamze Yilmaz. "Development of a new *Pseudomonas* agar medium containing benzalkonium chloride in cetrimide agar". *Food and Nutrition Sciences*, vol. 8, no. 4, pp. 367-378, 2017.

[22] S. Su, K. Z. U. Lae and H. Ngwe. "Isolation and identification of *Pseudomonas aeruginosa* from the clinical soil". *University of Yangon Research Journal*, vol. 8, pp. 271-275, 2018.

[23] S. S. Hasan, L. A. Said and S. L. Hamed. "Identification of Clinical *Pseudomonas* spp. by Vitek 2 compact system and species-specific polymerase chain reaction assay for identification of *Pseudomonas aeruginosa*". *International Journal of Drug Delivery Technology*, vol. 10, no. 3, pp. 383-388, 2020.

[24] M. A. Hama-Ali and A. H. Hasan. "Profiling of bacterial species from covid-19 faecal samples in Kurdistan Region-Iraq". *Science Journal of University of Zakho*, vol. 11, no. 1, pp. 11-15, 2023.

[25] H. Saderi and P. Owlia. "Detection of multidrug resistant (MDR) and extremely drug resistant (XDR) *P. aeruginosa* isolated from patients in Tehran, Iran". *Iranian Journal of Pathology*, vol. 10, no. 4, pp. 265-271, 2015.

[26] S. H. Yoon, S. M. Ha, S. Kwon, J. Lim, Y. Kim, H. Seo and J. Chun. "Introducing EzBioCloud: A taxonomically united database of 16S *rRNA* gene sequences and whole-genome assemblies". *International Journal of Systematic and Evolutionary Microbiology*, vol. 67, no. 5, pp. 1613-1617, 2017.

[27] "*DNA Sequence Assembler v4. Heracle BioSoft*", 2013. Available from: https://www.dnabaser.com/download/DNA-Baser-sequence-assembler/index.html

[28] Y. Tang, Z. Ali, J. Zou, G. Jin, J. Zhu, J. Yang and J. Dai. "Detection methods for *Pseudomonas aeruginosa*: History and future perspective". *RSC Advances*, vol. 7, no. 82, pp. 51789-51800, 2017.

[29] J. W. Chen, Y. Y. Lau, T. Krishnan, K. G. Chan and C. Y. Chang. "Recent advances in molecular diagnosis of *Pseudomonas aeruginosa* infection by state-of-the-art genotyping techniques". *Frontiers in Microbiology*, vol. 9, p. 1104, 2018,

[30] L. H. Moehario, E. Tjoa, H. Putranata, S. Joon, D. Edbert and T. Robertus. "Performance of TDR-300B and VITEK®2 for the identification of *Pseudomonas aeruginosa* in comparison with VITEK®-MS". *Journal of International Medical Research*, vol. 49, no. 2, pp. 1-12, 2021.

[31] Z. M. Alkhulaifi and K. A. Mohammed. "Prevalence and molecular analysis of antibiotic resistance of *Pseudomonas aeruginosa* isolated from clinical and environmental specimens in Basra, Iraq". *Iranian Journal of Microbiology*, vol. 15, no. 1, pp. 45-54, 2023.

[32] Y. Li, J. A. Roberts, M. M. Walker, A. T. Aslan, P. N. A. Harris and F. B. Sime. "The global epidemiology of ventilator-associated pneumonia caused by multi-drug resistant *Pseudomonas aeruginosa*: A systematic review and meta-analysis". *International Journal of Infectious Diseases*, vol. 139, pp. 78-85, 2024.

[33] M. Al-Orphaly, H.A. Hadi, F.K. Eltayeb, H. Al-Hail, B.G. Samuel, A.A. Sultan and S. Skariah. "Epidemiology of multidrug-resistant *Pseudomonas aeruginosa* in the Middle East and North Africa Region". *mSphere*, vol. 6, no. 3, pp. e00202-e00221, 2021.

[34] D. Baur, B. Gladstone, F. Burkert, E. Carrara, F. Foschi, S. Döbele and E. Tacconelli. "Effect of antibiotic stewardship on the incidence of infection and colonisation with antibiotic-resistant bacteria and Clostridium difficile infection: A systematic review and meta-analysis". *The Lancet Infectious Diseases*, vol. 17, no. 9, pp. 990-1001, 2017

[35] A. A. Al-Naqshbandi, H. A. Hassan, M. A. Chawsheen and A. Qader. "Categorization of bacterial pathogens present in infected wounds and their antibiotic resistance profile recovered from patients attending rizgary hospital-erbil". *Journal of the Association for Research in Otolaryngology*, vol. 9, no. 2, pp. 64-70, 2021.

[36] P. Hernández-Jiménez, F. López-Medrano, M. Fernández-Ruiz, J. T. Silva, L. Corbella, R. San-Juan, M. Lizasoain, J. Díaz-Regañón, E. Viedma and J. M. Aguado. "Risk factors and outcomes for multidrug resistant *Pseudomonas aeruginosa* infection in immunocompromised patients". *Antibiotics (Basel, Switzerland)*, vol. 11, no. 11, p. 1459, 2022

[37] A. Karruli, C. Catalini, D. D'Amore, F. Foglia, F. Mari, A. Harxhi, M. Galdiero and E. Durante-Mangoni. "Evidence-based treatment of *Pseudomonas aeruginosa* infections: A critical reappraisal". *Antibiotics*, vol. 12, no. 2, p. 399, 2023.

[38] G. S. Watts, K. Youens-Clark, M. J. Slepian, D. M. Wolk, M. M. Oshiro, G. S. Metzger, D. Dhingra, L. D. Cranmer and B. L. Hurwitz. "16S *rRNA* gene sequencing on a benchtop sequencer: Accuracy for identification of clinically important bacteria". *Journal of Applied Microbiology*, vol. 123, no. 6, 1584-1596, 2017.

[39] O. A. Eremwanarue, S. U. Nwawuba and O. H. Shittu. "Characterisation of the prevailing multidrug *Pseudomonas aeruginosa* strains from surgical wound using 16S rRNA Sequencing technique". *Malaysian Journal of Medical Sciences*, vol. 28, no. 4, pp. 37-49, 2021.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Enhancing Pelican Optimization Algorithm with Differential Evolution: A Novel Hybrid Metaheuristic Approach

Rebin Abdulkareem Hamaamin[1], Omar Mohammed Amin Ali[2]

[1]Department of Computer Science, College of Sciences, Charmo University, Chamchamal, Sulaymaniyah, Iraq.
[2]Department of Information Technology, Chamchamal Technical Institute, Sulaimani Polytechnic University, Sulaymaniyah, Iraq.

ABSTRACT

In the field of solutions for composite objective functions, the problem of identifying a proper trade-off between exploitation and exploration is still urgent. Classical methods can hardly avoid early iteration convergence or be insufficient in terms of searching throughout the space of potential solutions, especially when dealing with multi-variate multi-dimensional problems. To overcome this problem, this work proposes a combination of the pelican optimization algorithm (POA) and differential evolution (DE), known as the POA-DE metaheuristic method, which comprises the explorative characteristic of POA and the exploitative feature of DE. The main issue dealt with in this work relates to the conflict of global search and local exploitation in the context of solving complex optimization tasks. In global exploration, the POA technique is applied to improve the performances of the search in the large area, and the DE method is used in the local search space for improving the solution. To this end, the proposed solution hybrid model tries to avoid the shortcomings associated with using either of the two key aspects when used independently. To support the results obtained through POA-DE, it is necessary to perform the intensive empirical examination of several benchmark functions. The results also show that the proposed method has achieved better stability, efficiency, and convergence speed than the basic POA. Therefore, extending the hybrid optimization techniques is significant in enhancing the meta-heuristic algorithms that form a powerful tool to solve the optimization problems.

Index Terms: Pelican Optimization Algorithm, Differential Evolution, Hybrid Metaheuristic, Exploration and Exploitation, Optimization Benchmark Functions

1. INTRODUCTION

Optimization plays a key role in solving complex real-world problems across various disciplines, including engineering, medicine, and economics. One of the most pressing challenges in this domain is finding a suitable balance between exploration (searching new regions of the solution space) and exploitation (refining current solutions). Many metaheuristic algorithms suffer from premature convergence or inefficient space traversal, particularly when handling high-dimensional or multi-modal functions. This paper addresses this critical issue by proposing a hybrid approach that combines the pelican optimization algorithm (POA), known for its strong global search capabilities, with differential evolution (DE), a method renowned for effective local optimization. By combining these two useful methods, the proposed POA-DE algorithm seeks to improve results in tackling difficult optimization problems and to get past the drawbacks of using each method on its own. Optimizing complex non-linear functions with constraints remains a critical challenge in engineering and computational science.

Access this article online	
DOI:10.21928/uhdjst.v9n2y2025.pp101-114	**E-ISSN:** 2521-4217 **P-ISSN:** 2521-4209

Copyright © 2025 Hamaamin RA, Amin Ali OM. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Rebin Abdulkareem Hamaamin, Department of Computer Science, College of Sciences, Charmo University, Chamchamal, Sulaymaniyah, Iraq. E-mail: rebin.abdulkarim@chu.edu.iq

Received: 30-04-2025 Accepted: 14-08-2025 Published: 08-09-2025

Existing metaheuristics often face a trade-off between exploration and exploitation, which limits their effectiveness in high-dimensional and noisy environments. This paper addresses this challenge by proposing a novel hybrid algorithm that balances these aspects to achieve improved optimization performance.

This paper proposes a novel hybrid metaheuristic, the POA-DE, which synergistically integrates the exploratory capabilities of POA with the exploitative advantages of DE. This integration leads to improved convergence speed and solution quality on complex optimization problems compared to the original POA and other metaheuristics.

This optimization is the systematic process of finding the most favorable solution within a reasonable timeframe. This region has seen significant changes with the introduction of a genetic algorithm (GA) and DE. As a result, the number of optimization challenges is increasing in complexity. Therefore, the resolution of these issues requires the application of more efficient optimization techniques [1]. There are several effective algorithms available for solving a certain issue. Nevertheless, it is premature to designate any one of them as the superior option until this study has conducted a comprehensive evaluation of their relative performance in addressing the problem under consideration. Optimization algorithms are capable of efficiently solving many difficulties [2]. Optimization problem-solving methodologies can broadly be classified into two group's deterministic methods and stochastic methods [3]. When trying to solve difficult optimization problems with objective functions that are discontinuous, high-dimensional, non-convex, and non-derivative, deterministic approaches have trouble. Stochastic methods, unlike deterministic methods, can effectively address the challenges of optimization problems by utilizing random search in the problem-solving space. These methods do not rely on derivative and gradient information from the optimization problem's objective function [4]. People often classify stochastic as a heuristic or metaheuristic. Nature-inspired metaheuristic algorithms are capable of efficiently solving both real-world issues and traditional mathematical functions throughout their exploration and exploitation stages. However, achieving a balance between these two stages is a critical challenge that metaheuristic optimizations struggle with [5]. Various optimization issues have been tackled using metaheuristic methods. The objective of using these algorithms is to determine the maximum or lowest value of a certain function, such as minimizing the time required for a specific journey or minimizing the cost of completing a task [6]. However, these algorithms do have some shortcomings when it comes to achieving global optima since they must balance the competing goals of exploration and exploitation. Due to their excellent performance, metaheuristic algorithms tackle real-world issues. The problem appears to have its roots in electromagnetics [7], engineering design problems [8], constrained optimization problems [9], economic problems [10], medical problems [11], and task planning problems [12]. These methods are effectively utilized in a wide range of engineering and scientific applications, such as optimizing power generation in electrical engineering, designing bridges and buildings in civil engineering, performing data mining tasks such as classification, prediction, clustering, and system modeling, as well as designing radars and networking systems in communication [13].

In 2022, Trojovský and Dehghani [14] introduced a new optimization technique called the Pelican optimization technique POA. They have modeled the design after the foraging actions of pelicans. Compared to eight well-known SI optimization methods, the POA achieves outstandingly comparable performance by effectively balancing exploration and exploitation. Therefore, many practical applications utilize the POA. Although the standard POA is valuable, it is also susceptible to local optimization. To overcome this problem, several academics have proposed alternate, improved methodologies. They implemented tent chaos to improve population diversity and incorporated a dynamic weight factor to enable the pelican's continuous position updates. These approaches surpass classic POAs in performance and provide better outcomes in 10 benchmark functions. However, they did not compare the execution times of various algorithms [15]. Each particle modifies its path toward its past optimal position and the current optimal position achieved by any other member in its local area [16]. Particle swarm optimization (PSO) has the advantage over GA in that it is theoretically straightforward, requires little calculation time, and has a limited number of parameters to change. Nevertheless, the primary drawback of PSO is the potential for premature search convergence, particularly in intricate multi-peak search issues. In this study have developed a hybrid approach that combines PSO with DE to address this issue and enhance the efficiency of the PSO algorithm, this work proposes the PSO-DE method. DE is an enhanced iteration of GA, first introduced by Abualigah *et al.* [17]. A hybrid PSO-DE technique is proposed to tackle a global optimization problem. A hybrid PSO-DE technique is presented, combining the speed of PSO with the exploration capabilities of DE. The hybrid approach employs the PSO algorithm to identify the best solution area, followed by a

mix of PSO and DE algorithms to locate the ideal point [18]. The POA-DE distinguishes itself from previous POAs because it uses DE in its hybridization process. This way, the hybridization process can combine the characteristics of DE such as mutation and crossover with those of the Population-based Optimization Algorithm POA, which works by repeating itself. The proposed hybrid method incorporates an additional DE phase into the primary loop of the PSO. This novel addition aims to enhance both the exploration and exploitation capabilities of the original POA. The integration of DE's methodology for generating a pool of prospective solutions and POA's methodology for modifying positions led to the development of POA-DE, which offers a more balanced and comprehensive approach to global search and local fine-tuning. The POA-DE is a novel kind of hybridization that significantly enhances the optimization outcomes of the original POA.

Our research is organized around a POA framework, with its mechanism detailed in Section 2. Section 2 then elucidates the POA Work and Flowchart. Section 3 describes DE. Section 4 provides a comprehensive explanation of our proposed POA-DE approach. Section 5 involves a comparison between POA-DE and 23 benchmark test functions [19], statistical results are documented and presented. Finally, the conclusion is presented, and suggestions for future research are offered.

2. LITERATURE REVIEW

Stochastic population-based optimization algorithms are among the best methods for addressing optimization problems. Based on the primary concepts and sources of inspiration that shaped their design, optimization algorithms may be broadly divided into four groups: Swarm-based optimization techniques that are game-, physics-, and evolutionary-based. Natural phenomena, such as the swarm behaviors of insects, animals, and other living things, are considered while developing swarm-based optimization algorithms. One of the first and most widely used swarm-based algorithms is PSO, which draws inspiration from how birds forage for food. The best position each population member has encountered and the best position the whole population has experienced are used to update each member's status in the PSO [20]. The modeling of a classroom environment and student-teacher interactions serves as the foundation for teaching-learning-based optimization (TLBO). Population members in TLBO exchange information with one another and receive updates as part of teacher training [21]. The social behavior and hierarchical structure of gray wolves while hunting serve as the inspiration for gray wolf optimization (GWO). Alpha, beta, delta, and omega wolves are the four wolf types employed in GWO to simulate the hierarchical leadership of gray wolves. Simulations update population members by modeling the three primary hunting stages: Searching for prey, surrounding prey, and attacking prey [22]. The Whale optimization algorithm (WOA) is a swarm-based optimization algorithm inspired by nature that models the social behavior of humpback whales and their bubble-net hunting technique. WOA uses three hunting phases searching for prey, surrounding prey, and humpback whale bubble-net foraging behavior to update population members [23]. In this research, a Tunicate Swarm Algorithm (TSA) is developed by simulating the swarm behavior and jet propulsion of tunicates during feeding and navigation. Four phases avoidance of search agent (SA) conflicts, convergence toward the best SA, movement toward the best neighbor, and swarm behavior are used by TSA to update the population [24]. The movement strategies used by marine predators to catch their food in the ocean served as the model for the marine predators algorithm (MPA). The population update process in MPA is divided into three stages due to the different speeds of the predator and prey: (i) The predator is faster, (ii) the predator and prey are equal in speed, and (iii) the prey is faster [25]. The introduction of evolutionary-based optimization algorithms is predicated on models of genetic, biological, and other evolutionary processes. One of the first and most popular evolutionary algorithms is the GA, which draws inspiration from Charles Darwin's idea of natural selection and the reproductive process. In [26], the authors employ three primary operators' selection, crossover, and mutation to update the population members. The artificial immune system (AIS) algorithm, a revolutionary approach, is based on the immune system's response to viruses and bacteria. Cognitive, activation, and effector stages all have an impact on the population updating process in AIS [27]. The modeling of the many physics laws serves as the foundation for the development of physics-based optimization methods. The metallurgical melting and cooling process inspired the physics-based approach known as "simulated annealing." To lessen its flaws, the material is heated and then softly cooled under carefully monitored circumstances. The SA optimizer was designed using mathematical modeling of this process [28]. The modeling of the gravitational attraction between objects at varying distances from one another served as the inspiration for the gravitational search algorithm (GSA). The GSA updates its population members by modeling Newtonian principles of motion and calculating gravitational force [29]. The construction of game-based optimization algorithms is based on modeling player behavior and the

rules of various solo and multiplayer games. A game-based algorithm called the football game-based optimizer (FGBO) simulates player behavior and club relations in a football league. The four stages of league holding, training, player transfers between teams, and club promotion and relegation form the basis of FGBO's population update procedure [30]. The foundation of Tug of War Optimization (TWO) is modeling player behavior in a tug of war. Modeling the tensile force between population members who compete with one another serves as the foundation for TWO's population member update method [31].

While numerous metaheuristic algorithms such as GA, PSO, and DE have been proposed, most prior studies provide primarily descriptive overviews of their applications. A critical comparison reveals that many of these methods suffer from premature convergence or lack the balance between exploration and exploitation necessary for complex, high-dimensional optimization problems. The proposed POA-DE addresses this gap by integrating the exploratory nature of the POA with DE's exploitative strengths, aiming to improve convergence speed and solution quality.

3. THE POA ALGORITHM AND DE

3.1. Biological Inspiration and Mathematical Modeling of POA

The pelican is a large avian creature known for its prominent beak and throat pouch, which it uses to capture and consume food [32]. This species congregates in communal roosts, where groups of pelicans, sometimes numbering in the hundreds, meet together. We can describe the appearance of pelicans as follows: Pelicans have a weight range of approximately 2.75–15 kg, a height range of 1.06–1.183 m, and a wing span of 0.5–3 m. They primarily consume fish but also feed on frogs, turtles, and other crustaceans. In times of extreme hunger, they may even consume shellfish. Pigeons, particularly pelicans, often exhibit collective behavior when foraging for their prey. On locating their prey, they swiftly submerge into the water, descending a distance of 10–20 m. Indeed, some predators descend to lower regions to capture their prey. Pelicans are considered to be very skilled hunters based on their intellect, hunting habits, and techniques. The suggested approach's modeling primarily inspires the design of the intended POA [14], [33].

The proposed POA is population-based, including pelicans as well. Within population-based algorithms, every individual serves as a prospective solution. Every individual in the population suggests values for the variables of the optimization problem depending on their location in the search space. At first, individuals in the population are allocated randomly within the specified range of values using Equation (1).

$$X_{ab} = Y_b + rand.\left(Z_b - Y_b\right), a = 1,2,\ldots\ldots,M,\ b = 1,2\ldots\ldots N \quad (1)$$

where X_{ab} is the value of the bth variable specified by the ath candidate solution, M is the number of population members, N is the number of problem variables, rand is a random number in interval [0, 1], Y_b is the bth lower bound, and Z_b is the bth upper bound of problem variables. The suggested POA replicates pelican behavior and strategy for approaching and hunting prey, updating potential solutions.

3.2. DE and its Role in POA-DE

The DE algorithm, introduced by Storn and Price [17], is a powerful evolutionary optimization technique widely used for solving continuous and multimodal optimization problems. DE operates through three primary operations: Mutation, crossover, and selection. These steps allow the population to evolve better solutions over iterations.

In the context of the hybrid POA-DE algorithm, DE enhances the global search ability of POA by diversifying the candidate solutions early in the search process, helping avoid premature convergence. The new trial vector is generated using DE's mutation and crossover operations. Depending on the nature of the problem, various mutation and crossover strategies can be employed [34] [35].

This hybridization leverages the global exploration strength of DE and the problem-specific exploitation ability of POA, leading to a more robust and efficient optimization process.

3.3. Parameter Settings

Table 1 summarizes the key parameters used in the POA-DE algorithm. These parameters control the behavior and performance of the algorithm during optimization. Selecting appropriate values is essential to balance exploration of the search space and convergence speed, ensuring effective and efficient optimization results.

4. HYBRID APPROACH: POA-DE

The suggested study introduces a novel metaheuristic algorithm called POA-DE, which combines the features of

TABLE 1: Parameter settings

Parameter	Description	Value	Justification
Population size	Number of pelican agents	30	Balances search diversity and computational cost
Maximum iterations	Max iterations for algorithm	500	Empirically sufficient for convergence in tests
DE mutation factor (F)	Controls mutation step size	0.5	Standard value that balances exploration and exploitation
DE crossover rate	Probability of crossover in DE	0.9	Encourages recombination of solutions

DE: Differential evolution

POA and DE. The purpose of this hybridization is to merge DE's global search capability with POA's local exploitation capability to achieve a balance between exploration and exploitation in the optimization process. The POA-DE algorithm operates in two distinct stages. The first stage employs the DE algorithm. Therefore, DE starts by selecting three pairs of distinct candidate solutions from the population in a random manner. The mutation process generates a third candidate solution by calculating the weighted difference between the first two solutions and adding it to the third solution. This modified candidate subsequently mates with the aforementioned candidate, resulting in the creation of a novel solution. This implies that the population includes the new solution only if its fitness exceeds that of the current population. This phase fortifies the algorithm's learning capabilities as it searches for the optimal solution to a given issue, shielding it from becoming stuck in suboptimal solutions. The POA assumes the role during the second phase. Pelicans' hunting technique is the source of the term "POA." During exploration, they categorize each potential solution as either moving closer to or farther away from a reference solution, also known as a food source. This categorization is based on the fitness value compared to the reference values. Put simply, the candidate approaches the reference point if its fitness is greater, and it goes away if the candidate's ability is lower. It is particularly helpful in investigating the search space. Following the exploration phase, the exploitation phase improves the solutions by including a small random increment that progressively decreases during the exploitation phase. Optimizing this fine-tuning process is crucial for enhancing the exploration of the local search space and improving the quality of the acquired solutions.

The hybrid POA-DE algorithm has been tested using a collection of benchmark functions (F1–F23) to compare the results with the original POA. Experiments were conducted to evaluate the performance and adaptability of POA-DE across various population sizes and iteration counts. The outcomes assessment demonstrated that POA-DE exhibited a superior level of accuracy when compared to the original POA in several functions, thereby confirming the successful integration of DE into the POA framework.

In the DE phase of the proposed POA-DE algorithm, the mutation factor (F) and crossover rate (CR) play crucial roles in controlling the search dynamics. In this study, F is set to 0.5 and CR to 0.9, which are commonly used default values known to provide a good trade-off between exploration and exploitation. To ensure the robustness of these parameters, a sensitivity analysis was conducted by varying F within the range [0.4, 0.9] and CR within [0.5, 1.0]. The results confirmed that the algorithm maintains stable performance under these variations, supporting the suitability of the chosen parameter values for the benchmark functions used in this work.

Algorithm (1) displays the Pseudocode and flowchart for POA-DE.

4.1. Mathematical Equations for Hybrid POA-DE

Equation (2) in the DE phase, which is also called the mutation equation, yields a new candidate solution by introducing diversity into the population. This step is done for three randomly selected solutions, though adding to the third solution a scaled difference between two other solutions, this process creates a mutant vector that has the potential to explore a new area of the search space. Consequently, it aids the algorithm in evading local optima and amplifies its capacity for global search.

$$V = X_a + \mathrm{F}.(\mathrm{X_b} - \mathrm{X_c}) \quad (2)$$

Here, V is the mutant vector, F is the differential weight, X_a+F.(X_b−X_c) are three randomly selected individuals from the population.

Crossover Equation (3) uses the generated mutant vector during the previous step and combines it with the current solution to generate a trial solution. This is achieved through random selection of some of the components either from the mutant vector or the current solution. This process also guarantees that the trial solution has some of the features

ALGORITHM 1: Pseudo code of pelican optimization algorithm-differential evolution

```
Algorithm: POA-DE
1. Input:
  a. Define the objective function f (x)
  b. Set bounds for each decision variable
  c. Set population size N and maximum iterations T
2. Initialization:
  a. Randomly initialize positions of N pelicans within bounds → X
     = {x1, x2, ., xn}
  b. Evaluate fitness of each pelican: f (xi) ∀ i ∈ [1, N]
3. For t = 1 to T do:
  a. Step 1: Update global best
    i. Identify best pelican: X_best = argmin (f (xi))
    ii. Store best fitness value: f_best = f (X_best)
  b. Step 2: Differential Evolution Phase (DE Phase)
    i. Set DE parameters: F (mutation factor), CR (crossover rate)
    ii. For each pelican i = 1 to N do:
      1. Mutation:
        - Randomly select three distinct pelicans: x_r1, x_r2, x_r3
        - Compute mutant vector: vi = x_r1 + F * (x_r2 - x_r3)
      2. Crossover:
        - Generate trial vector ui by mixing vi and xi based on CR
      3. Selection:
        - If f (ui) < f (xi), then xi ← ui
  c. Step 3: Pelican Optimization Algorithm Phase (POA Phase)
    i. Select a random pelican as prey: X_FOOD
    ii. Phase 1: Exploration (Moving Towards Prey)
      For each pelican i = 1 to N:
        - Update position of xi based on movement toward X_FOOD
    iii. Phase 2: Exploitation (Winging on Water Surface)
      For each pelican i = 1 to N:
        - Refine position of xi using water surface dynamics
  d. Step 4: Update and store best solution found in iteration t
4. Output:
  - Best candidate solution X_best and corresponding fitness f_best
End Algorithm
```

of the original and the mutant solutions, hence diversifying the population to increase the possibility of arriving at an even better solution.

$$U_{i(j)} = \begin{cases} X_{i(j)} & if\ rand(j) \le CR\ or\ j = j_{rand} \\ X_{i(j)} & otherwise \end{cases} \quad (3)$$

$U_{(i(j))}$ is the trial vector, $X_{(i(j))}$ is the target vector, CR is the crossover rate, and jrand is a randomly chosen index.

The exploration phase in Equation (4) determines the direction of moving a candidate solution in the search space forward or backward depending upon whether the current candidate solution performs better or worse than a randomly selected "food." If the solution becomes worse, it approaches the food in an attempt to better itself. If it's better, it steps slightly back, which makes one search for other options with more fervor. This allows the algorithm to extend its search in multiple directions and not to be confined to areas with lower utility levels.

$$X_i^{new} = \begin{cases} X_i + rand().(X_{Food} - I.X_i) & if\ f(X_i) > f(X_{FOOD}) \\ X_i + rand().(X_i - X_{FOOD}) & otherwise \end{cases} \quad (4)$$

Here, X_i^{new} is the updated position of an individual, X_{FOOD} is the location of a random individual considered as "food," and I is a random binary value.

The exploitation phase in Equation (5) refines the candidate solution by adjusting the position slightly in a way that reduces over time. Initially, to search the space of the search space, the changes made are more significant, whereas as the algorithm proceeds, the changes made are comparatively smaller, trying to get as close to the optimal value as possible. This makes sure that the algorithm is able to converge with the right solution within a short span of time.

$$X_i^{new} = X_i + 0.2.\left(1 - \frac{t}{Max_iterations}\right).(2.rand() - 1).X_i \quad (5)$$

This equation introduces a small random perturbation to X_i as the algorithm iteratively refines the solution towards the optimal value.

4.2. Statistical Validation

The Wilcoxon rank-sum test was carried out with a significance threshold of 0.05 to provide statistical evidence that POA-DE is better than other evaluation methods. The findings indicate that POA-DE performs much better than the metaheuristics that were chosen for comparison in the majority of benchmark functions. This substantiates the fact that the observed increases in performance are statistically significant and are not the product of random chance.

5. EXPERIMENTAL SETUP AND RESULTS

A novel POA-DE method is provided in this study, followed by a comprehensive empirical analysis using 23 benchmark functions. This assessment is part of the comprehensive experimental design used to compare the effectiveness of the hybrid technique. The 23 benchmark functions are a standard set of test problems that can be used to see how the suggested changes to the POA-DE algorithm stack up against the original POA method. The test results are very important for supporting the idea that using both DE and

POA together makes the balance between exploration and exploitation better, as well as the optimization performance in many situations. The below tables and figures provide a performance comparison between the original POA and the hybrid POA-DE. The results are calculated using two parameters: The number of SA and the maximum number of iterations (MI). In cases where POA-DE performs better than POA, significant improvements (SIs) are indicated. In addition to the highest scores, each table displays the benchmark functions on which both algorithms have been evaluated. The tables are accompanied by charts that facilitate comparison. Table 2 displays the results of the POA versus POA-DE comparison for 20 SA. This table presents a comparison between the results of the POA and the outcomes of the POA-DE in situations involving 20 SA. The findings include a range of iteration numbers. They are calculating the square values of 100, 500, 800, and 1000. In addition, for each iteration count, the top results for each algorithm are shown for each benchmark function, highlighting significant improvements where POA-DE outperforms POA. To be more precise: Out of 100 trials, there are 17 instances where the performance of POA-DE exceeds that of POA in terms of SI. In 500 iterations, there are 16 improvements, with POA-DE surpassing POA by a substantial margin. Among the 800 iterations, there are 15 occurrences when the POA-DE has a superior index (SI) in comparison to the POA. Among the 1000 repetitions, there are 14 instances of SI. These findings allow for a comparison of the relative efficiency of the studied algorithms in terms of their ability to operate with a smaller group of persons. They also provide an evaluation of how many times the algorithm hybridization based on the POA improves its performance compared to the original version.

Table 3 presents the results of the comparison between the original POA and the hybrid POA-DE for scenarios with 30 SA. It includes results for various iteration counts: 100, 500, 800, and 1000. For each iteration count, the table shows the best scores achieved by each algorithm for each benchmark function and indicates significant improvements where POA-DE outperforms POA. Specifically: For 100 iterations, there are 17 significant improvements where POA-DE performs better than POA. For 500 iterations, there are 14 significant improvements where POA-DE performs better than POA. For 800 iterations, there are 14 significant improvements where POA-DE performs better than POA.

TABLE 2: POA versus POA-DE results for 20 SA

Function	Maximum number of iterations							
	100		500		800		1000	
	POA	POA-DE	POA	POA-DE	POA	POA-DE	POA	POA-DE
F1	8.44242E-19	2.607E-19	5.3111E-103	1.0405E-128	5.6066E-180	2.5394E-212	6.0024E-232	2.9654E-259
F2	1.02422E-08	1.67645E-11	1.602E-54	2.75195E-68	3.89621E-86	1.906E-111	1.2771E-111	5.7266E-147
F3	4.8004E-18	1.34672E-20	1.0435E-114	1.2884E-119	3.8238E-155	8.9934E-187	3.6462E-226	1.5807E-240
F4	1.83678E-10	4.41394E-10	3.90111E-52	5.80852E-62	6.50224E-87	3.9976E-100	1.8618E-106	3.6818E-127
F5	28.8410256	28.40746944	28.83434241	25.72787702	26.46808462	24.36385211	27.96069942	24.13425128
F6	0	0	0	0	0	0	0	0
F7	0.001842118	0.002489956	0.000811606	0.0002196	0.00018826	0.000251286	0.000195164	0.000185465
F8	−6752.338401	−8973.535957	−7986.181885	−8658.249743	−7206.595666	−9666.603987	−7690.545649	−9739.742638
F9	0	0	0	0	0	0	0	0
F10	7.1642E-10	2.5695E-12	4.44089E-15	4.44089E-15	4.44089E-15	4.44089E-15	4.44089E-15	4.44089E-15
F11	0	0	0	0	0	0	0	0
F12	0.379335727	0.110466078	0.20717646	0.006759976	0.163103261	0.011113309	0.307773401	0.004877733
F13	2.993078621	1.886155787	2.485823119	0.828283394	2.974257347	0.722645709	1.996425926	0.593715012
F14	0.998003838	0.998003838	0.998003838	0.998003838	0.998003838	0.998003838	0.998003838	0.998003838
F15	0.000308452	0.001223173	0.000307486	0.000307486	0.000307487	0.000307486	0.000307505	0.020363339
F16	−1.031628453	−1.031628453	−1.031628453	−1.031628453	−1.031628453	−1.031628453	−1.031628453	−1.031628453
F17	0.397887372	0.397887358	0.397887358	0.397887358	0.397887358	0.397887358	0.397887358	0.397887358
F18	3	3	3	3	3	3	3	3
F19	−3.862782044	−3.862782148	−3.862782148	−3.862782148	−3.862782148	−3.862782148	−3.862782148	−3.862782148
F20	−3.321732438	−3.321995121	−3.321990444	−3.20310205	−3.32199453	−3.321995172	−3.321994009	−3.321995172
F21	−10.14786467	−10.15319968	−10.15307188	−10.15319968	−10.15313299	−10.15319968	−5.055197391	−10.15319968
F22	−10.39811953	−10.40294057	−10.40283234	−10.40294057	−10.402923	−10.40294057	−10.40289467	−10.40294057
F23	−10.45228277	−10.53640982	−10.53628678	−10.53640982	−10.53639666	−10.53640982	−10.53638089	−10.53640982
SI	17		16		15		14	

POA: Pelican optimization algorithm, DE: Differential evolution, SA: Search agent

TABLE 3: POA versus POA-DE results for 30 SA

Function	Maximum number of iterations							
	100		500		100		1000	
	POA	POA-DE	POA	POA-DE	POA	POA-DE	POA	POA-DE
F1	2.44686E-21	3.24652E-21	1.8596E-100	2.8411E-129	7.1447E-166	6.2389E-202	1.0337E-225	1.689E-261
F2	6.30687E-10	6.92141E-13	5.36129E-52	6.45989E-69	3.11624E-86	6.1841E-110	3.3754E-110	6.1431E-133
F3	1.49479E-18	2.15422E-19	2.1028E-111	3.8733E-127	1.5153E-172	2.2947E-192	4.3141E-216	6.8049E-241
F4	6.17297E-09	1.05369E-08	2.94986E-52	6.17893E-63	6.5319E-91	5.0517E-102	7.6164E-113	2.1422E-122
F5	28.9155397	28.5032141	28.83833557	24.70437222	27.97057944	24.06465201	25.859068	25.00836819
F6	0	0	0	0	0	0	0	0
F7	0.001223655	0.001004973	0.000328992	6.25394E-05	0.000140125	0.00022994	0.000429553	0.000181634
F8	−6427.415167	−8416.720434	−7084.244045	−8835.40723	−8185.882717	−9316.476719	−8513.675101	−8658.286337
F9	0	0	0	0	0	0	0	0
F10	8.89046E-10	2.33826E-10	4.44089E-15	8.88178E-16	4.44089E-15	4.44089E-15	4.44089E-15	4.44089E-15
F11	0	0	0	0	0	0	0	0
F12	0.57331789	0.056721115	0.176276713	1.57179E-06	0.145433174	1.16027E-06	0.1679032	3.03096E-07
F13	2.990963261	1.009436307	2.620938032	0.297529501	2.982695256	0.108381791	2.980186413	0.308401823
F14	0.998003838	0.998003838	0.998003838	0.998003838	0.998003838	0.998003838	0.998003838	0.998003838
F15	0.000308878	0.000307486	0.000307486	0.000307486	0.000307487	0.000307486	0.000307486	0.000307486
F16	−1.031628453	−1.031628453	−1.031628453	−1.031628453	−1.031628453	−1.031628453	−1.031628453	−1.031628453
F17	0.397887358	0.397887358	0.397887358	0.397887358	0.397887358	0.397887358	0.397887358	0.397887358
F18	3	3	3	3	3	3	3	3
F19	−3.862782087	−3.862782148	−3.862782148	−3.862782148	−3.862782148	−3.862782148	−3.862782148	−3.862782148
F20	−3.321708176	−3.20310205	−3.321993603	−3.321995172	−3.321994662	−3.20310205	−3.321991814	−3.20310205
F21	−5.055164449	−10.15319968	−10.1531993	−5.055197729	−10.15319967	−10.15319968	−10.15315661	−10.15319968
F22	−10.40167737	−10.40294057	−10.40292935	−10.40294057	−10.40281433	−10.40294057	−10.40294057	−10.40294057
F23	−10.53138633	−10.53640982	−10.5363856	−10.53640982	−10.53624667	−10.53640982	−5.128480787	−10.53640982
SI	17		14		14		15	

POA: Pelican optimization algorithm, DE: Differential evolution, SA: Search agent

For 1000 iterations, there are 15 significant improvements where POA-DE performs better than POA. These results offer a comparative view of the algorithms' performance with a medium-sized population and highlight how the iteration count influences the effectiveness of the hybrid approach over the original POA.

Furthermore, Table 4 presents the results of the comparison between the original POA and the hybrid POA-DE for scenarios with 50 SA. It includes results for various iteration counts: 100, 500, 800, and 1000. For each iteration count, the table shows the best scores achieved by each algorithm for each benchmark function and indicates significant improvements where POA-DE outperforms POA. Specifically: For 100 iterations, there are 16 SI where POA-DE performs better than POA. For 500 iterations, there are 14 SI where POA-DE performs better than POA. For 800 iterations, there are 16 SI where POA-DE performs better than POA. For 1000 iterations, there are 13 SI where POA-DE performs better than POA. This table provides insights into how a larger population size affects the performance improvements of POA-DE relative to POA and the impact of iteration count on optimization results.

In last table, Table 5 presents the results of the comparison between the original POA and the hybrid POA-DE for scenarios with 80 SA. It includes results for various iteration counts: 100, 500, 800, and 1000. For each iteration count, the table shows the best scores achieved by each algorithm for each benchmark function and indicates significant improvements where POA-DE outperforms POA. Specifically: For 100 iterations, there are 17 SI where POA-DE performs better than POA. For 500 iterations, there are 13 SI where POA-DE performs better than POA. For 800 iterations, there are 13 SI where POA-DE performs better than POA. For 1000 iterations, there are 14 SI where POA-DE performs better than POA. These results reflect the performance of the algorithms with a larger population and provide a comparison of how different iteration counts affect the relative success of the hybrid POA-DE approach over the original POA.

The graphical comparisons between the POA and the hybrid POA-DE, with a focus on significant improvements, are illustrated in four figures: These are Figs. 1-4. For either figure, in this research report the results of a total of 23 test functions for given values of SAs and four iteration counts of

TABLE 4: POA versus POA-DE results for 50 SA

Function	Maximum number of iterations							
	100		500		100		1000	
	POA	POA-DE	POA	POA-DE	POA	POA-DE	POA	POA-DE
F1	2.49576E-15	8.25975E-24	6.4535E-113	1.7909E-129	1.5327E-175	3.1429E-201	2.788E-225	2.7614E-260
F2	3.883E-12	5.58281E-11	2.10498E-56	4.55152E-62	2.78613E-91	2.0983E-105	4.5724E-118	7.365E-131
F3	3.82949E-18	2.40486E-17	1.1427E-106	7.8247E-121	1.6074E-168	7.2797E-191	2.1796E-216	3.4427E-243
F4	4.78844E-09	2.42029E-10	4.97603E-51	3.37586E-58	1.05869E-85	5.01127E-98	1.4792E-106	8.7552E-127
F5	28.70319141	28.41110891	28.02387665	24.66922068	28.81369914	23.19309231	28.55704875	23.64091015
F6	0	0	0	0	0	0	0	0
F7	0.000272223	0.001243581	0.000168009	0.000191357	2.52893E-05	2.42018E-05	4.91583E-05	0.000110357
F8	−6585.887624	−6602.34042	−8065.980309	−9206.070817	−8220.373482	−9350.304939	−6827.237998	−10094.43083
F9	0	0	0	0	0	0	0	0
F10	1.02749E-10	8.60156E-11	4.44089E-15	4.44089E-15	4.44089E-15	4.44089E-15	4.44089E-15	4.44089E-15
F11	0	0	0	0	0	0	0	0
F12	0.395989404	0.008500108	0.114831156	7.42445E-07	0.223673566	4.61258E-10	0.105689638	3.94851E-08
F13	2.995377503	0.195516639	2.979128781	0.097371549	2.976471966	0.010992274	2.9762112	0.240193412
F14	0.998003838	0.998003838	0.998003838	0.998003838	0.998003838	0.998003838	0.998003838	0.998003838
F15	0.000516476	0.000307486	0.000307486	0.000307486	0.000307486	0.000307486	0.000307486	0.001223173
F16	−1.031628453	−1.031628453	−1.031628453	−1.031628453	−1.031628453	−1.031628453	−1.031628453	−1.031628453
F17	0.397887358	0.397887358	0.397887358	0.397887358	0.397887358	0.397887358	0.397887358	0.397887358
F18	3	3	3	3	3	3	3	3
F19	−3.862782148	−3.862782148	−3.862782148	−3.862782148	−3.862782148	−3.862782148	−3.862782148	−3.862782148
F20	−3.321263723	−3.20310205	−3.321989525	−3.20310205	−3.321994373	−3.321995172	−3.203100353	−3.20310205
F21	−10.14426658	−10.15319968	−5.055197729	−10.15319968	−10.15319967	−10.15319968	−10.15319968	−10.315319968
F22	−10.38733962	−10.40294057	−10.40294057	−10.40294057	−10.40294057	−10.40294057	−10.40294057	−10.40294057
F23	−10.53576953	−10.53640982	−10.53640982	−10.53640982	−10.53640982	−10.53640982	−10.53640284	−10.53640982
SI	16		14		16		13	

POA: Pelican optimization algorithm, DE: Differential evolution, SA: Search agent

TABLE 5: POA versus POA-DE results for 80 SA

Function	Maximum number of iterations							
	100		500		100		1000	
	POA	POA-DE	POA	POA-DE	POA	POA-DE	POA	POA-DE
F1	1.10087E-18	6.35165E-21	1.5474E-110	1.4112E-124	3.5709E-176	4.7273E-203	3.2398E-224	1.3994E-255
F2	2.66845E-10	2.18443E-10	4.80324E-53	2.51132E-60	2.11834E-89	6.4643E-109	4.9799E-112	8.4084E-130
F3	7.89952E-18	7.25637E-21	1.4617E-114	1.0093E-123	1.458E-174	2.241E-189	3.9415E-228	1.6495E-237
F4	4.06298E-09	1.62514E-10	1.68652E-54	2.84401E-58	2.39672E-84	1.54687E-98	3.4455E-106	1.2882E-119
F5	28.78665385	27.75183764	27.96056999	23.72066394	28.58014219	23.35552802	28.80080497	22.5407497
F6	0	0	0	0	0	0	0	0
F7	0.000183214	0.00046826	5.21957E-05	0.000267952	1.93001E-05	4.18375E-05	4.34409E-05	2.89613E-06
F8	−6611.55237	−7473.179641	−7089.851035	−9527.994796	−8402.996499	−10631.92435	−7543.986097	−10393.55237
F9	0	0	0	0	0	0	0	0
F10	1.03606E-11	4.31521E-11	4.44089E-15	4.44089E-15	4.44089E-15	4.44089E-15	4.44089E-15	4.44089E-15
F11	0	0	0	0	0	0	0	0
F12	0.434235978	0.009326528	0.186761494	8.6929E-11	0.134381986	2.3543E-12	0.204381175	1.25909E-20
F13	2.608189819	0.062913512	1.538478871	4.31207E-10	2.172645757	6.71697E-12	2.975194915	0.09737116
F14	0.998003838	0.998003838	0.998003838	0.998003838	0.998003838	0.998003838	0.998003838	0.998003838
F15	0.001275838	0.000307486	0.000307486	0.000307486	0.000307486	0.000307486	0.001223173	0.000307486
F16	−1.031628453	−1.031628453	−1.031628453	−1.031628453	−1.031628453	−1.031628453	−1.031628453	−1.031628453
F17	0.397887358	0.397887358	0.397887358	0.397887358	0.397887358	0.397887358	0.397887358	0.397887358
F18	3	3	3	3	3	3	3	3
F19	−3.862782148	−3.862782148	−3.862782148	−3.862782148	−3.862782148	−3.862782148	−3.862782148	−3.862782148
F20	−3.321919745	−3.321995172	−3.321994915	−3.321995172	−3.321994196	−3.321995172	−3.32199447	−3.20310205
F21	−10.152767	−10.15319968	−10.15319965	−10.15319968	−5.055197729	−10.15319968	−10.15319968	−10.15319968
F22	−10.39964661	−10.40294057	−10.40294057	−10.40294057	−10.40293405	−10.40294057	−10.40294057	−10.40294057
F23	−10.53569662	−10.53640982	−10.5364098	−10.53640982	−10.53640982	−10.53640982	−10.53640982	−10.53640982
SI	17		13		13		14	

POA: Pelican optimization algorithm, DE: Differential evolution, SA: Search agent

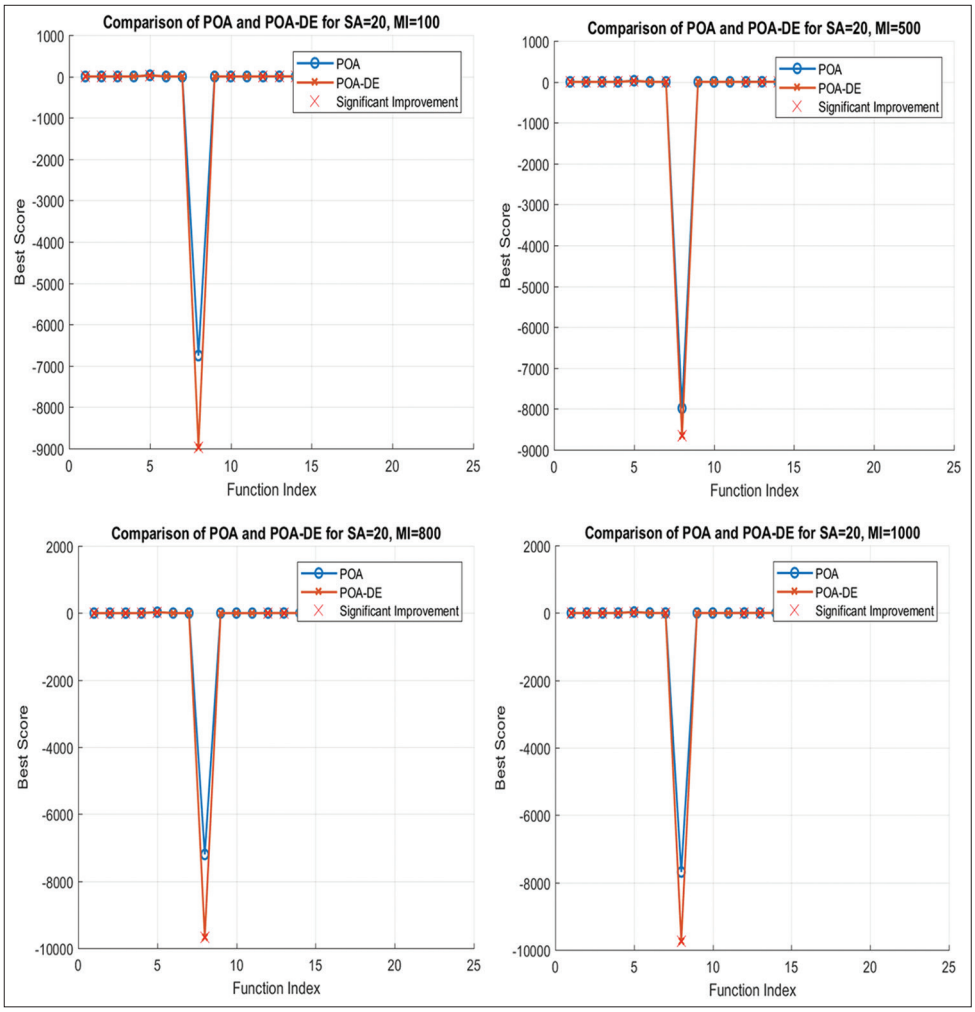


Fig. 1. Significant improvements in optimization performance: POA versus POA-DE for 20 SA. POA: Pelican optimization algorithm, DE: Differential evolution, SA: Search agent.

100, 500, 800, and 1000. In all the cases, as shown in Fig. 1, when SA is set at 20, POA-DE has superior performance to POA across the board, and the differences are enhanced with more iterations. The two regions circled in red in the figure represent such enhancements; POA-DE's superior optimization results over the POA mechanism are indicated here. The same observation can be made when comparing the results presented in Fig. 2, where the results for SA = 30 are shown, which points to the fact that, in most of the test function cases, as well as at higher iteration numbers, POA-DE outperforms POA. This is the reason why the specific substantial enhancements pointed out by me underscore the efficiency of the studied hybrid design with a moderately larger quantity of populace. Figs. 3 and 4 sustain this emphasis on tremendous enhancements for SA values of 50 and 80, respectively. As indicated by the red crosses, the number of significant improvements starts to appear more frequently as the iteration count increases. This implies that with a large population size, the hybrid algorithm has more ability to search and optimally solve complex functions. The hybrid algorithm outperforms all the other algorithms in each of the settings, and the largest SA of 80 yields significant improvements from the POA to the POA-DE, as seen in Fig. 4. These figures combined show that POA-DE has a substantial improvement over the original POA, although the number of significant improvements increases with population size and iteration count, effectively substantiating the use of the hybrid model in the different test functions.

6. REAL-WORLD APPLICATION

To evaluate the real-world effectiveness of the proposed POA-DE algorithm, it was applied to practical engineering

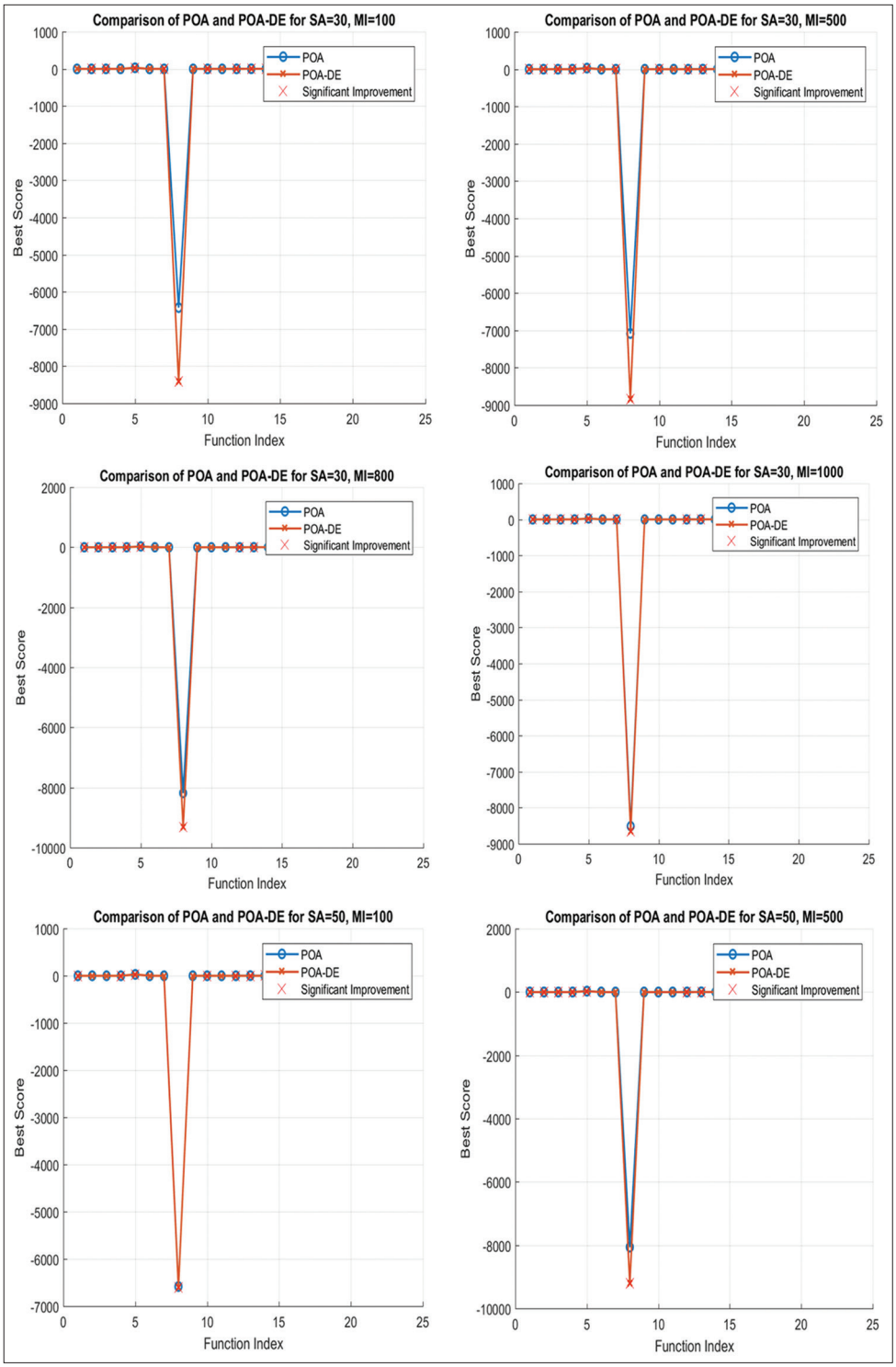


Fig. 2. Significant improvements in optimization performance: POA versus POA-DE for 30 SA. POA: Pelican optimization algorithm, DE: Differential evolution, SA: Search agent.

and healthcare problems. These applications show that POA-DE works better and finds solutions faster than the original POA and other similar methods, especially in complex situations with many variables and limitations.

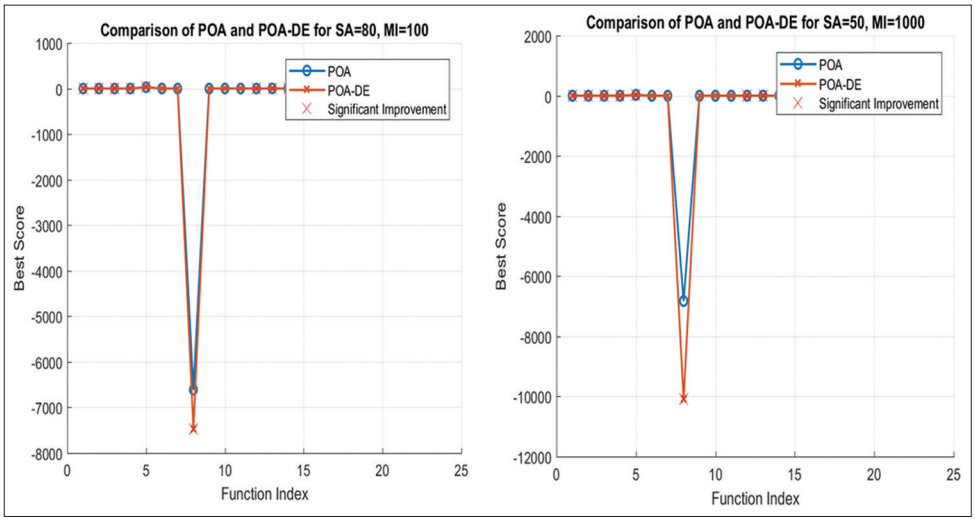


Fig. 3. Significant improvements in optimization performance: POA versus POA-DE for 50 SA. POA: Pelican optimization algorithm, DE: Differential evolution, SA: Search agent.

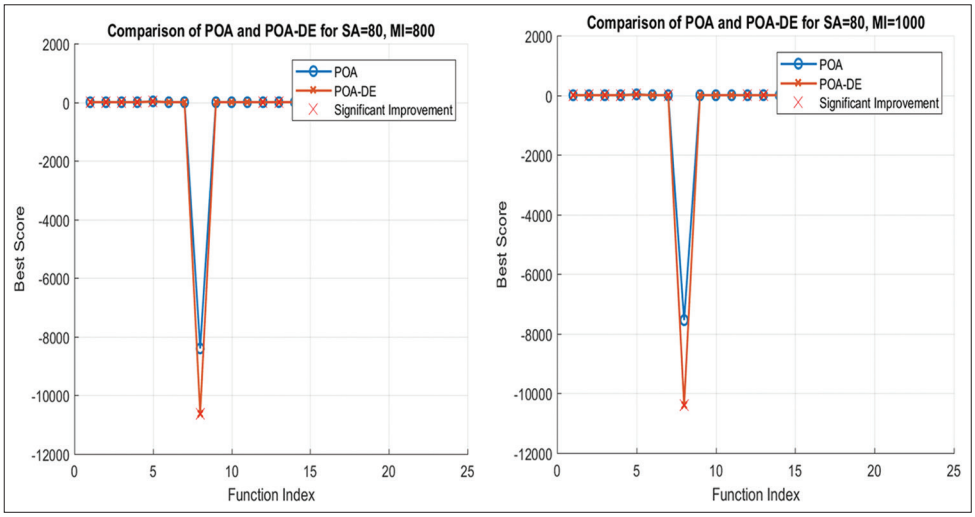


Fig. 4. Significant improvements in optimization performance: POA versus POA-DE for 80 SA. POA: Pelican optimization algorithm, DE: Differential evolution, SA: Search agent.

6.1. Medical Diagnosis–Feature Selection for Cancer Classification

POA-DE was employed to optimize feature selection for the Wisconsin breast cancer diagnostic dataset. This dataset involves diagnosing tumors as benign or malignant using 30 real-valued features. Feature selection was modeled as a binary optimization problem, aiming to maximize classification accuracy and minimize the number of selected features. A support vector machine served as the classifier, with 10-fold cross-validation to evaluate performance, as shown in Table 6.

The results show that POA-DE not only improved classification accuracy by ~1.7% over POA but also achieved it using fewer features, which reduces computational burden and enhances model interpretability. Its robust search capability in high-dimensional and noisy medical data makes it a promising candidate for clinical decision support systems.

6.2. Renewable Energy–Optimal Placement of Wind Turbines

POA-DE was applied to optimize the layout of wind turbines in a wind farm to maximize energy output while considering wake effect losses, turbine spacing constraints, and land area limitations. The wind flow was simulated based on prevailing direction and velocity using a simplified Jensen wake model, as shown in Table 7.

TABLE 6. Comparison of POA-DE and other algorithms for feature selection in breast cancer classification

Algorithm	Accuracy (%)	Selected features	Standard deviation
POA-DE	98.6	7	0.41
POA	96.9	9	0.57
GA	94.5	11	0.81
PSO	95.8	10	0.66

POA: Pelican optimization algorithm, DE: Differential evolution, PSO: Particle swarm optimization, GA: Genetic algorithm

TABLE 7: Performance comparison of POA-DE and other metaheuristics in wind farm layout optimization

Algorithm	Energy output (MWh/year)	Improvement (%)	Runtime (s)
POA-DE	14,275	+12.1	78.6
POA	12,735	___	84.2
DE	13,154	+3.3	92.4
PSO	13,079	+2.7	101.5

POA: Pelican optimization algorithm, DE: Differential evolution, PSO: Particle swarm optimization

Objective function:

- Maximize total annual energy production
- Penalize overlap or suboptimal spacing that increases wake losses.

The hybrid POA-DE produced layouts that generated up to 12.1% more energy compared to the original POA and also converged faster to near-optimal solutions. This application highlights POA-DE's ability to efficiently solve complex, non-linear, constrained engineering problems, contributing to the planning of sustainable renewable energy infrastructures.

7. CONCLUSION

This research presents a novel metaheuristic technique called POA-DE. This novel technique integrates advantageous aspects from both algorithms, leading to enhanced optimization performance in terms of solution quality and convergence as compared to the original POA. The results of the trials conducted on several benchmark functions demonstrate POA-DE's effectiveness in tackling complex optimization issues. The effectiveness of combining DE's exploitative capabilities with POA's exploratory qualities to enhance metaheuristic optimization approaches has been demonstrated. Future research will focus on enhancing and advancing numerous areas, including the following: initially, the algorithm's performance will be evaluated by subjecting it to a substantial collection of intricate real-world optimization problems. This will allow for the assessment of the technique's adaptability and reliability. Furthermore, diverse methodologies for configuring distinct parameters and implementing various types of adaptivity for the algorithm will be explored. One recommended area for future research is the integration of POA-DE with other metaheuristic approaches to provide more advanced and versatile optimization algorithms. Furthermore, it would be beneficial to do research on parallel and distributed computing approaches to effectively use them for optimizing large-scale algorithms and improving their performance. Future work will focus on several key directions to address current limitations. These include applying the algorithm to real-world optimization tasks to assess practical robustness, conducting parameter sensitivity analyses, exploring adaptive mechanisms, and extending the comparison with other advanced metaheuristics. In addition, incorporating statistical tests will ensure that performance improvements are significant and not due to randomness. These enhancements aim to further establish POA-DE as a reliable and scalable optimization framework.

REFERENCES

[1] N. A. Rashed, Y. H. Ali and T. A. Rashid. "Advancements in optimization: Critical analysis of evolutionary, swarm, and behavior-based algorithms". *Algorithms*, vol. 17, no. 9, p. 416, 2024.

[2] X. S. Yang and X. He. "Nature-inspired optimization algorithms in engineering: Overview and applications". *Studies in Computational Intelligence,* vol. 637, pp. 1-20, 2016.

[3] D. Das, A. S. Sadiq and S. Mirjalili. "Optimization methods: Deterministic versus stochastic". In: *Optimization Algorithms in Machine Learning*. Singapore: Springer, 2025.

[4] F. A. Hashim, E. H. Houssein, K. Hussain, M. S. Mabrouk and W. Al-Atabany. "Honey badger algorithm: New metaheuristic algorithm for solving optimization problems". *Mathematics and Computers in Simulation*, vol. 192, pp. 84-110, 2022.

[5] E. H. Houssein, M. K. Saeed, G. Hu and M. M. Al-Sayed. "Metaheuristics for solving global and engineering optimization problems: Review, applications, open issues and challenges". *Archives of Computational Methods in Engineering*, vol. 31. pp. 4485-4519, 2024.

[6] K. Rajwar, K. Deep and S. Das. "An exhaustive review of the metaheuristic algorithms for search and optimization: Taxonomy, applications, and open challenges". *Artificial Intelligence Review*, vol. 56, pp. 13187-13257, 2023.

[7] V. Tomar, M. Bansal and P. Singh. "Metaheuristic algorithms for optimization: A brief review". *Engineering Proceedings*, vol. 59, no. 1, p. 238, 2024.

[8] I. Vale, A. Barbosa, A. Peixoto and F. Fernandes. "Solving authentic problems through engineering design". *Open Education Studies*, vol. 5, no. 1, p. 20220185, 2023.

[9] X. Yu, W. Chen and X. Zhang. “An Artificial Bee Colony Algorithm for Solving Constrained Optimization Problems”. In: *2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*. IEEE, Xi’an, China, pp. 2663-2666, 2018.

[10] K. R. Khudaiberganovich. “The concept of mathematical models of economic problems”. *Miasto Przyszłości*, vol. 49, pp. 392-394, 2024.

[11] S. Alagarsamy, R. R. Subramanian, T. Shree, S. Kannan, M. Balasubramanian and V. Govindaraj. “Prediction of Lung Cancer Using Meta-heuristic Based Optimization Technique: Crow Search Technique”. In: *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*. IEEE, Greater Noida, India, pp. 186-191. 2021.

[12] A. Kaveh and Y. Vazirinia. “Construction site layout planning problem using metaheuristic algorithms: A comparative study”. *Iranian Journal of Science and Technology - Transactions of Civil Engineering*, vol. 43, no. 2, pp. 105-115, 2019.

[13] H. Salimi. “Stochastic fractal search: A powerful metaheuristic algorithm”. *Knowledge-Based Systems*, vol. 75, pp. 1-18, 2015.

[14] P. Trojovský and M. Dehghani. “Pelican optimization algorithm: A novel nature-inspired algorithm for engineering applications”. *Sensors (Basel)*, vol. 22, no. 3, p. 855, 2022.

[15] W. Tuerxun, C. Xu, M. Haderbieke, L. Guo and Z. Cheng. “A wind turbine fault classification model using broad learning system optimized by improved pelican optimization algorithm”. *Machines*, vol. 10, no. 5, p. 407, 2022.

[16] Y. Han, F. Zeng, L. Fu and F. Zheng. “GA-PSO algorithm for microseismic source location”. *Applied Sciences*, vol. 15, no. 4, p. 1841, 2025.

[17] L. Abualigah, A. Sheikhan, A. M. Ikotun, R. A. Zitar, A. R. Alsoud, I. Al-Shourbaji, A. G. Hussien and H. Jia. “Particle swarm optimization algorithm: Review and applications”. In: *Metaheuristic Optimization Algorithms. Optimizers, Analysis, and Applications*. Elsevier Science, Amsterdam, Netherlands, pp. 1-14, 2024.

[18] M. Ahmadipour, M. M. Othman, R. Bo, M. S. Javadi, H. M. Ridha and M. Alrifaey. “Optimal power flow using a hybridization algorithm of arithmetic optimization and aquila optimizer”. *Expert Systems with Applications*, vol. 235, p. 121212, 2024.

[19] S. Mirjalili and A. Lewis. “The whale optimization algorithm”. *Advances in Engineering Software*, vol. 95, pp. 51-67, 2016.

[20] J. Kennedy and R. Eberhart. “Particle swarm optimization”. In: *Proceedings of ICNN’95 - International Conference on Neural Networks*. Vol. 4. IEEE, Perth, Australia, pp. 1942-1948, 1995.

[21] R. V. Rao, V. J. Savsani and D. Vakharia. “Teaching-learning-based optimization: A novel method for constrained mechanical design optimization problems”. *Computer-Aided Design*, vol. 43, pp. 303-315, 2011.

[22] S. Mirjalili, S. M. Mirjalili and A. Lewis. “Grey wolf optimizer”. *Advances in Engineering Software*, vol. 69, pp. 46-61, 2014.

[23] K. Liu and Y. Wang, “A novel whale optimization algorithm based on population diversity strategy,” *IAENG International Journal of Computer Science*, vol. 52, no. 8, 2025.

[24] S. Kaur, L. K. Awasthi, A. L. Sangal and G. Dhiman. “Tunicate swarm algorithm: A new bio-inspired based metaheuristic paradigm for global optimization”. *Engineering Applications of Artificial Intelligence*, vol. 90, p. 103541, 2020.

[25] A. Faramarzi, M. Heidarinejad, S. Mirjalili and A. H. Gandomi. “Marine predators algorithm: A nature-inspired metaheuristic”. *Expert Systems with Applications*, vol. 152, p. 113377, 2020.

[26] D. E. Goldberg and J. H. Holland. “Genetic algorithms and machine learning”. *Machine Learning*, vol. 3, pp. 95-99, 1988.

[27] L. N. De Castro and J. I. Timmis. “Artificial immune systems as a novel soft computing paradigm”. *Soft Computing*, vol. 7, pp. 526-544, 2003.

[28] S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi. “Optimization by simulated annealing”. *Science*, vol. 220, pp. 671-680, 1983.

[29] E. Rashedi, H. Nezamabadi-Pour and S. Saryazdi. “GSA: A gravitational search algorithm”. *Information Sciences*, vol. 179, pp. 2232-2248, 2009.

[30] M. Dehghani, M. Mardaneh, J. M. Guerrero, O. Malik and V. Kumar. “Football game based optimization: An application to solve energy commitment problem”. *International Journal of Intelligent Engineering and Systems*, vol. 13, pp. 514-523, 2020.

[31] A. Kaveh and A. Zolghadr. “A novel meta-heuristic algorithm: Tug of war optimization”. *Iran University of Science and Technology*, vol. 6, pp. 469-492, 2016.

[32] A. Louchart, N. Tourment and J. Carrier. “The earliest known pelican reveals 30 million years of evolutionary stasis in beak morphology”. *Journal of Ornithology*, vol. 152, no. 1, pp. 15-20, 2011.

[33] J. G. T. Anderson and S. C. Waterbirds. “Foraging behavior of the American white Pelican (*Pelecanus erythrorhyncos*) in Western Nevada”. *Colonial Waterbirds*, vol. 14, no. 2, pp. 166-172, 1991.

[34] B. Zolghadr-Asli. “Differential evolution algorithm”. In: *Computational Intelligence-based Optimization Algorithms*. CRC Press, United States, 2023.

[35] S. Das and P. N. Suganthan. “Differential evolution: A survey of the state-of-the-art”. *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 1, pp. 4-31, 2011.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Intelligent System for Screening Epileptic Seizures in the Erbil Electroencephalogram Epilepsy Dataset Images Utilizing Cascaded Histogram of Oriented Gradients-Gray Level Co-occurrence Matrix Features

Hero Abdullah Mohammed[1]*, Salih Omer Haji[1], Raghad Zuhair Yousif[1,2]

[1]Department of Physics, College of Science, Salahaddin University-Erbil, Erbil, Kurdistan Region, Iraq, [2]Department of Computer Engineering, College of Engineering, Knowledge University, Erbil 44001, Iraq

ABSTRACT

Driven by strong demand on easy-to-use, accurate, and automated diagnosis tools to help people with epilepsy, especially in the Kurdistan region, where is a scarcity of specialized doctors in remote places, this study proposes a novel approach for epileptic seizure detection from electroencephalogram (EEG) signals using a statistical feature extraction method that being derived from a cascaded histogram of oriented gradients (HOG) and gray level co-occurrence matrix (GLCM) techniques for (117 normal) non-epileptic seizure and (117 abnormal) epileptic seizure diagnosed EEG signal images collected from Erbil teaching hospital. Four classification algorithms, namely support vector machine (SVM), K-Nearest Neighbors (KNN), decision tree, and discriminator with rigorous hyperparameter optimization using Bayesian techniques, were utilized to improve classification and three feature extraction approaches, cascaded HOG-GLCM, GLCM, based statistical features extraction, and HOG, were calculated. The proposed comprehensive simulation results revealed that the cascaded HOG-GLCM approach significantly outperforms single-feature methods. The SVM and KNN classifiers achieved exceptional performance with the cascaded features, both approximately reaching 98.57% accuracy, ensuring almost no epileptic events went undetected, which represents a substantial improvement over GLCM (best: 92.86% accuracy) and HOG approaches (best: 94.29% accuracy). The synergistic effect observed between gradient-based and texture-based features demonstrates how HOG captures directional patterns characteristic of seizure activity, while GLCM extracts spatial relationships within the signal. Neither feature type alone provides sufficient discriminative power, as evidenced by the 5–8% accuracy drop in single-feature approaches.

Index Terms: Histogram of Oriented Gradients, Machine Learning, Electroencephalogram, Epilepsy, Gray Level Co-occurrence Matrix

1. INTRODUCTION

Epilepsy represents one of the most prevalent neurological disorders worldwide, affecting approximately 51.7 million people globally as of 2021, with an age-standardized prevalence of 658/100,000 population [1]. Every year, the World Health Organization reports five million new cases [2]. The disorder disproportionately affects low- and middle-income nations, where over 80% of epileptics live and where treatment disparities may span 75% or more [3]. This great worldwide load emphasizes the urgent requirement for accurate, quick, and easily available diagnostic instruments. Epilepsy may arise from an anomaly in cerebral wiring, an imbalance of neurotransmitters, or a combination of these

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp115-128 | E-ISSN: 2521-4217 P-ISSN: 2521-4209

Copyright © 2025 Mohammed, *et al.* This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Hero Abdullah Mohammed, Department of Physics, College of Science, Salahaddin University-Erbil, Erbil, Kurdistan Region, Iraq. E-mail: hero.mohammed@su.edu.krd

Received: 24-05-2025 Accepted: 27-07-2025 Published: 09-09-2025

elements [4]. Neurons often create electrical signals that influence other neurons, glands, and muscles, resulting in human thoughts, emotions, and behaviors. Epilepsy is hard to diagnose since it has many different aspects. Epilepsy can happen when the wiring in the brai n is not right, when there are too many or too few neurotransmitters, or when these things happen together [4]. Neurons usually provide electrical signals that affect other neurons, glands, and muscles. This is how thoughts, feelings, and actions happen in people. Seizures can happen when these electrical patterns get messed up. They can cause strange feelings, emotions, behaviors, convulsions, muscle spasms, and loss of consciousness [5]. The problem is finding these unusual patterns in the complicated landscape of regular brain activity. The recent article published by Chen *et al.* in 2023 indicated that epilepsy is not only related to the seizure disorder but also results in unusual feelings, emotions, and behaviors, as well as convulsions, muscular spasms, and loss of consciousness that require management strategies [5].

The suggested cascaded histogram of oriented gradients-gray level co-occurrence matrix (HOG-GLCM) method solves these problems using the strengths of two well-known feature extraction methods that work well together. HOG is great at picking up gradient and edge information that shows the structural patterns of seizure activity. GLCM, on the other hand, is good at analyzing textural patterns and spatial correlations within electroencephalogram (EEG) signals in a more complex way. By combining these methods, we develop a feature extraction framework that is specifically designed to tell the difference between normal and epileptic EEG signals. This has a number of advantages over single-feature methods. The proposed cascaded approach has a number of benefits, including (1) better discriminative power by extracting both structural and textural features from EEG signals; (2) better performance across a range of signal conditions thanks to the multi-scale representation created by combining gradient-based and texture-based features; (3) better classification accuracy that is much better than using individual feature extraction methods; and (4) the ability to be scaled up for use in real-time clinical settings. However, there are some things to think about with this approach: (1) More complicated to compute than single-feature methods, so they need to be optimized for real-time use; (2) depend on high-quality EEG signal preprocessing to make sure feature extraction works; (3) need to carefully tune hyperparameter to get the best performance across different patient groups; and (4) need to be thoroughly validated across multiple datasets to show that they can be used in other situations.

2. RELATED WORK

In this section of our article, we will provide a comprehensive review of articles that are similar to the topic of our study. Several signal processing and feature extraction methods have been investigated for automatic epilepsy detection, such as time-domain, which was suggested by Fasil and Rajesh in 2019 [6], and also Chakrabarti *et al.* in 2012 used temporal properties of EEG to distinguish between normal and seizure cases [7]. Recently, researchers have used new techniques and methods to automatically detect the early stages of EEG. The study done by Shoeb in 2009 used machine learning based on support vector machine (SVM) to classify patient-specific data through the analysis of scalp EEG, which acquired 96% accuracy [8]. In 2020, Usman *et al.* used main component analysis for seizure prediction using the wavelet transform. Abnormal brain activity is seen to start before to a seizure, referred to as the predictable state. Twenty-three participants evaluated the models, resulting in a sensitivity of 93.1% for 84 seizures [9]. Investigated by Anuragi and Sisodia in 2020 are empirical wavelet transform-based automated alcoholism detection utilizing EEG signal characteristics. They presented an (EWT)-based automatic classification model for alcohol detection. With LS-SVM, they obtained 98% average accuracy and 98% area under the curve (AUC) value. Their research neglected important details on processing efficiency and real-time feasibility [10].

Ayman *et al.* (2023) conducted the identification of epileptic seizures by means of physical epileptic activity analysis instead of feature engineering; therefore, they classified epilepsy. K-Nearest Neighbors (KNN), (NB), (LR), (SGDC), (GB), decision tree (DT), and three deep learning models: Extreme learning machine (ELM), Long (LSTM), and (ANN) were seven separate machine learning techniques used for the binary categorizations of epileptic activity. With an accuracy of 100% and a 0.99 AUC, their proposed deep learning model, ELM, obtained the best results [11].

Farooq *et al.* (2023) conducted a review on the detection of epileptic seizures using machine learning, addressing its taxonomy, opportunities, and challenges. This research focused on identifying the most prevalent feature extraction techniques and classifiers used for the precise categorization of normal and epileptic seizures [12]. In 2023, Chandel *et al.* investigated the classification of epileptic EEG signals with a machine learning-based model. The differentiation between normal EEG and epileptic EEG has been conducted using three classification models: (RF), (DT), and (ET). Their conclusion indicates that the final classifier exhibits the

highest performance, with an accuracy of 99.63 for the suggested approach [13].

In 2025, Atlam *et al.* introduced a new way to improve the detection of epileptic seizures using SMOTE to balance the data and a combination of techniques called PCA and DWT to select important features. The suggested method seeks to make seizure detection systems more accurate and dependable by fixing data imbalance and pulling out important features from EEG signals. Performance improvements were seen, with the SVM classifier achieving an accuracy of 97.30% [14].

Object identification, particularly for human recognition in images, is achieved through the use of HOG. It employs the "HOG descriptor," a feature extraction method designed for human identification in an image [15]. Co-occurrence oriented gradient (CoHOG) histograms are a gradient-based technique used in many fields. The CoHOG method provides a consistent characterization of both statistical and differential properties of a texture [16].

Zhou *et al.* (2020) employed the HOG approach for feature extraction from raw Bayer pattern images [17]. Sarowar *et al.* (2019) utilized the (HOG) approach to extract features from images [18]. Bakheet and Al-Hamadi (2021) employed HOG features in conjunction with the SVM classifier to detect blinks and categorize power system equipment based on auditory data [19]. GLCM, introduced by Haralick *et al.* (1973), provides a statistical method for examining texture by analyzing spatial relationships between pixels [20]. GLCM has been widely applied in medical image analysis, offering valuable textural information that complements structural features. The integration of HOG and GLCM approaches has shown promise in various domains, as demonstrated by Kataoka *et al.* (2015) with their CoHOG [16]. This technique, as Thontadari and Prabhakar (2016) note, offers a consistent characterization of both statistical and differential attributes of texture [21].

The present study addresses these challenges by introducing a novel cascaded approach that combines HOG and GLCM techniques for EEG signal classification. This approach leverages the complementary strengths of both methods: HOG's ability to capture gradient and edge information and GLCM's capacity to analyze textural patterns and spatial relationships. By cascading these techniques, a feature extraction framework is specifically optimized for differentiating between normal and epileptic EEG signals.

Our research makes several significant contributions to the field:

- A novel cascaded HOG-GLCM feature extraction methodology has been introduced, which effectively captures both structural and textural characteristics of EEG signals, enhancing classification accuracy.
- Converting the EEG data from single dimension (signal data) to two dimensions (image data) which paved the road to use the techniques used in feature extraction of digital images in capturing the features of the EEG signal
- The establishment of a comprehensive dataset of epilepsy signal images, called the Erbil Epilepsy Dataset, includes EEG recordings sourced from governmental hospitals in Erbil, Kurdistan, thus fulfilling the demand for diverse and representative datasets in epilepsy research.
- To maximize classification performance, rigorous hyperparameter optimization (HPO) with Bayesian approaches was employed to fine-tune numerous classifiers.

The remainder of this paper is organized as follows: Section 3 details our methodology, including dataset collection, preprocessing, the proposed cascaded HOG-GLCM feature extraction approach, classification techniques, our experimental results, and comparative analysis. Section 4 discusses the implications of our findings, Section 5 computational complexity and real-time consideration, Section 6 limitations and constraints, and Section 7 concludes with a summary of contributions and directions for future research.

3. METHODOLOGY

The study focused on exploring four different methods for classifying EEG abnormalities. For each method, the system followed the essential steps of machine learning: Pre-processing the EEG signals collected, extracting relevant features, and performing classification. This approach allowed us to systematically analyze and categorize the EEG signals for accurate detection of seizure and non-seizure cases. The study focused on four different methods for classifying EEG abnormalities, following machine learning protocols for each method. This allowed for systematic analysis and categorization of EEG signals to accurately distinguish between seizure and non-seizure events. Fig. 1 below presents an example of two EEG image samples from the Erbil EEG Epilepsy Dataset, according to the categories under examination.

3.1. EEG Dataset Collection

Over 3 months from September to November 2024, the study collected 234 EEG cases from Erbil Teaching Hospital in

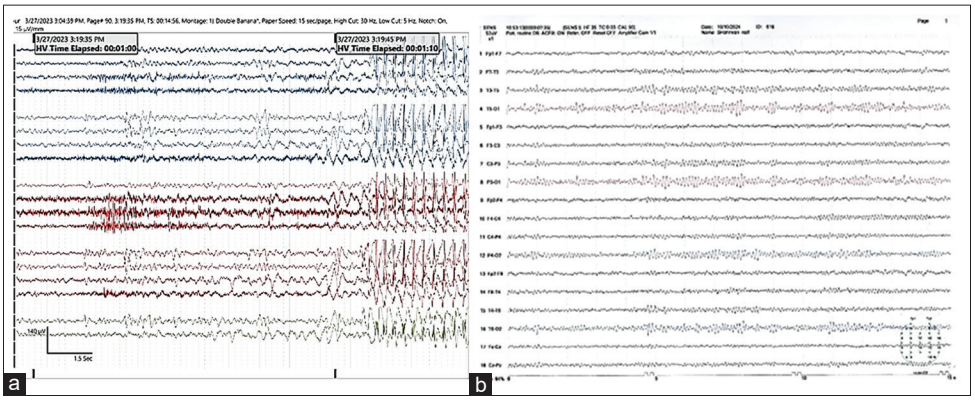


Fig. 1. Two samples of seizure and non-seizure electroencephalogram (EEG) brain signal images from Erbil EEG epilepsy dataset (a) seizure EEG brain signal image (b) non-seizure EEG brain signal image.

the Kurdistan Region of Iraq. The dataset consisted of 117 normal (non-seizure) and 117 abnormal (seizure) recordings, collected from male and female patients across diverse age groups. Seizure cases were confirmed by skilled neurologists, while normal cases were validated by the absence of seizure activity EEG data were gathered using NIHON KOHDEN-brand software Neuroworkbench version 05–01 with 32 electrodes arranged based on the "10–20 system" for placing the electrodes and gain sensitivity of 100 μV (50 Hz) ranging from a high frequency of 30Hz to a low frequency of 0.03 Hz, but only 18 active electrodes were used. Fig. 2 depicts the EEG machine used in this study. Part of the collected images, namely 50 normal and 50 abnormal, have been uploaded to GitHub with URL https://github.com/DrRaghad-75/ErbilEEG_dataset-/tree/main, aiming to add complete documentation and images in the near future.

Fig. 2. Electroencephalogram signal recording done by the corresponding author utilizing Nihon Kohden-brand machine.

3.2. Proposed Model

Fig. 3 below demonstrates the flow chart for the proposed system model. In this model, each EEG recording a patient has undergone (seizure and non-seizure) categories to help identify their seizure state. The entire process can be broken down into five key components: EEG Dataset Collection, pre-processing, feature extraction, classification, and, eventually, performance evaluation metrics calculation.

3.3. Preprocessing

The preprocessing phase of EEG signals encompasses numerous critical procedures for efficient feature extraction for analysis. Initially, signal-to-image conversion is employed to convert 1D EEG signals into 2D RGB and then into single-band (grayscale) images. The single-band image is cropped to eliminate the borders, focusing just on the regions of interest. The resultant images are subsequently scaled to a uniform resolution of 256 by 256 pixels, assuring consistency for subsequent analysis and modeling. In the denoising stage, histogram equalization followed by Median filter has been used to improve image quality and diminish noise. Ultimately, image binarization has been utilized to transform the processed photographs into a binary format, thereby streamlining the data and highlighting essential structures. This readies the dataset for dependable feature extraction and classification activities. Combining these four steps into a single conceptual equation is as shown below:

I_enhanced(x,y) = Rgb2gray(imbinarize(imresize(imcrop(medfilt2(histeq(I_old(x,y)), [M N])), [256 256])));

3.4. Feature Extraction Methodology

Imagine trying to identify patterns in brain waves that could signal an epileptic episode. The proposed new approach intends to do just that by merging two potent technologies in a particular manner. The proposed cascaded approach

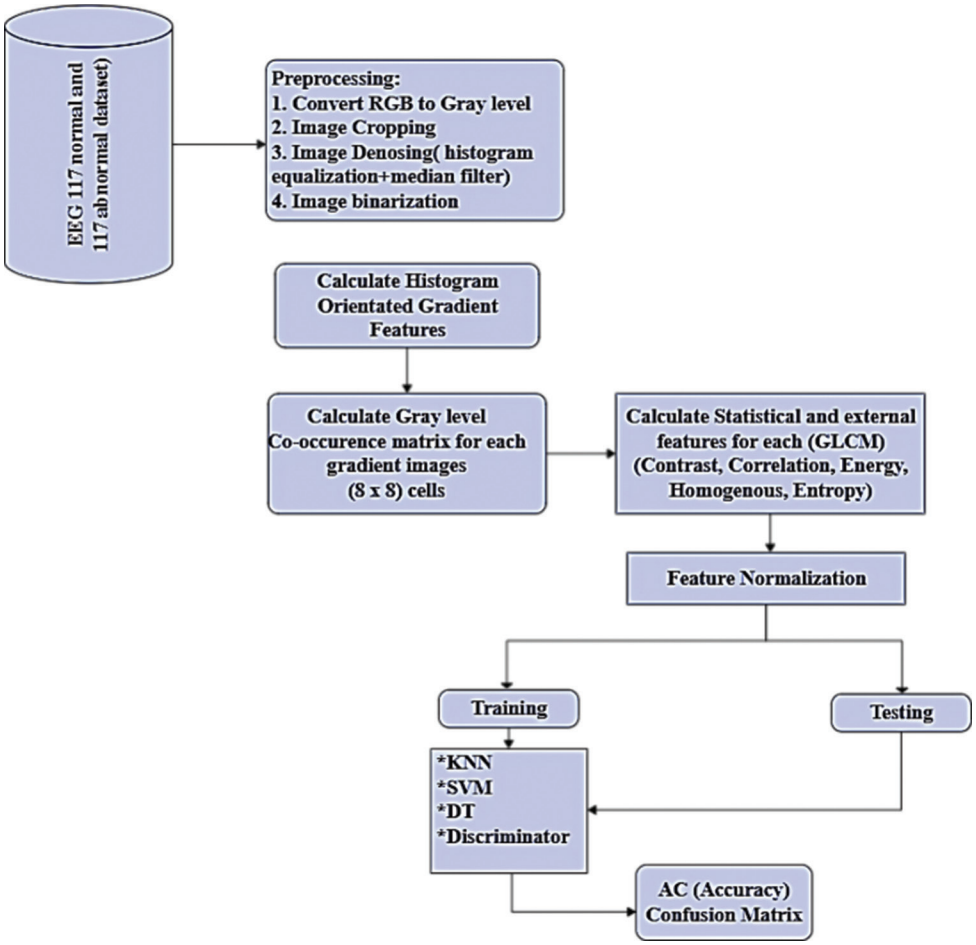


Fig. 3. The pattern of the model used to classify normal and epileptic seizures.

first captures the structural "shape" of brain signals using (HOG) then examines how these patterns spatially connect using (GLCM), and finally extracts relevant statistical features to facilitate the classifier's job to differentiate between normal brain activity from seizures. Recent work by Krishnan *et al.* (2024) showed that when time-based EEG signals are transformed into image-like representations, powerful computer vision techniques can be applied, which dramatically improve the ability to detect epilepsy [22]. This insight perfectly aligns with the proposed cascaded approach. Thus, the cascaded approach offers several key advantages that make it particularly well-suited for EEG signal classification: In the first place, it captures complementary information, think of HOG as an expert at identifying shapes and edges, while GLCM specializes in analyzing textures. By combining them, the best of both worlds can be captured - in another place provides multi-scale representation, hence, the proposed approach analyzes brain signals at different levels of detail - from local gradient patterns (HOG) to their spatial relationships (GLCM). This multi-layered view provides more comprehensive understanding of what's happening in the brain. In the third place, it handles signal variations robustly - Brain signals naturally vary in amplitude and frequency, making them challenging to analyze. Fortunately, HOG does not get thrown off by local changes in signal strength (similar to how it handles lighting changes in images), while GLCM captures consistent textural patterns regardless of absolute signal values. This robustness is particularly valuable when analyzing EEG data from different patients or recording sessions. Eventually, it enhances discriminative power - The statistical features have been derived from HOG-GLCM matrices provide highly distinctive attributes that effectively differentiate between normal and epileptic EEG patterns.

3.4.1. Cascaded HOG-GLCM approach

The HOG feature extraction process follows these systematic steps:

1. Gradient computation: For each pixel in the EEG image, the horizontal gradient and vertical gradient have been calculated as indicated below [19]:

$G_x(x, y) = f(x + 1, y) - f(x - 1, y)$ $G_y(x, y) = f(x, y + 1) - f(x, y)$ (1)

Where $f(x, y)$ represents the pixel intensity at position (x, y).

2. Gradient magnitude and orientation: From these gradients, were computed using [19]

Gradient magnitude: $\left|\nabla f(x, y)\right| = \sqrt{G_x(x, y)^2 + G_y(x, y)^2}$ (2)

Gradient orientation: $\theta(x, y) = \arctan\left(\frac{G_y(x, y)}{G_x(x, y)}\right)$

3. Cell histograms: Each EEG signal's image is divided into small spatial regions called "cells" (typically 8 × 8 pixels). For each cell, the histogram of gradient orientations is computed by: Quantizing the orientation angles into bins (typically 9 bins covering 0°–180° for unsigned gradients), then weighting each pixel's contribution by its gradient magnitude, and eventually accumulating these weighted votes into the orientation bins.
4. Block normalization: To account for variations in illumination and contrast, cells are grouped into larger spatial regions called "blocks" (typically 2 × 2 cells). The histograms within each block are concatenated and normalized using L2-norm depicted by equation 3 below [13]:

$$v_{normalized} = \frac{v}{\sqrt{\| v \|_2^2 + \epsilon}} \quad (3)$$

Where v is the unnormalized feature vector containing all histograms in a block, and ϵ is a small constant (typically 1e−5) to prevent division by zero.

5. Matrix creation: For a given HOG feature map with N normalized gray levels, $N \times N$ matrix is created where each element (i,j) represents the frequency of occurrence of pixel pairs with values i and j separated by a specific displacement vector (d_x, d_y).
6. Multiple direction analysis: To capture textural patterns in all directions, we compute GLCMs for four principal directions: 0° (horizontal), 45° (diagonal), 90° (vertical), and 135° (anti-diagonal), using displacement vectors $(d, 0)$, (d, d), $(0, d)$, and $(-d, d)$ respectively, where d is the distance parameter (typically $d = 1$).
7. Normalization: Each GLCM is normalized by dividing each element by the total number of pixel pairs, converting the frequency counts to probability values:

$$P(i, j) = \frac{C(i, j)}{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} C(i, j)} \quad (4)$$

Where $C(i, j)$ is the count in the original GLCM, and $P(i, j)$ is the normalized probability.

Poza *et al.* (2020) have validated the effectiveness of this multi-directional GLCM approach for analyzing complex patterns in biomedical signals [23].

8. Feature vector construction: The final HOG descriptor is formed by concatenating the normalized block histograms from all blocks in the detection window, creating a comprehensive feature vector that captures gradient information across the entire EEG image.

3.4.2. Statistical features extraction

This study has selected the following features. Contrast refers to the local variables or physical attributes quantified in the GLCM. Thus, variations or intensity differences are quantified in the GLCM [24].

$$Contrast = \sum_{i}^{M}\sum_{j}^{N}(i - j)^2 P(i, j) \quad (5)$$

As previously mentioned, $P(i, j)$ represents the normalized probability of every pair at an angle of 0°. Homogeneity quantifies the proximity of the element distribution in the GLCM to its diagonal. It measures how close the element distribution in the GLCM is to its diagonal [24]:

$$Homogeneity = \sum_{i}^{M}\sum_{j}^{N}\frac{P(i, j)}{1 + |i - j|} \quad (6)$$

Correlation denotes the linear dependence of gray level values inside the co-occurrence matrix.

$$Correlation = \sum_{i}\sum_{j} P(i, j)\frac{(i - \mu_x)(j - \mu_y)}{\sigma_x \sigma_y} \quad (7)$$

Where μ_x, μ_y, and σ_x, σ_y, the means and standard deviations are specified [10]:

$$\mu_x = \sum_i \sum_j iP(i,j), \mu_y = \sum_i \sum_j jP(i,j)$$

$$\sigma_x = \sqrt{\sum_i \sum_j (i-\mu_x)^2 P(i,j)}, \sigma_y = \sqrt{\sum_i \sum_j (j-\mu_y)^2 P(i,j)}$$

Entropy represents the unpredictability or level of disorder within an image generated by the GLCM pairs.

$$Entropy = -\sum_i \sum_j P(i,j) \ln P(i,j) \quad (8)$$

The energy measures how consistent the texture is, thus higher energy levels mean that the textures are more uniform and have fewer gray level transitions. Lower energy levels mean that the textures are more complex [24].

$$Energy = \sqrt{\sum_i^M \sum_j^N P(i,j)^2} \quad (9)$$

The five statistical features give a full picture of the textural aspects that come from the GLCM investigation. Each feature captures a separate aspect of the spatial relationships in the processed EEG images, making it a strong set of features for telling the difference between normal and epileptic patterns. The proposed system framework is presented by the algorithm below Algorithm 1:

ALGORITHM 1: Epileptic seizure classification using HOG-GLCM cascaded features

1: for all $P_i \in D$ train = {P_i|i = 1, ., I train} do
2: Convert 1D EEG signal of patient P_i to 2D grayscale image I(i) by scanning each patient EEG signals in to single binary medical image
3: Preprocess I(i) by cropping, resizing, denoising, and binarizing
4: End for
5: For all I(i) ∈ {I(1),.., I (itrain)} do
6: Compute gradients Gx(i), Gy(i) for each pixel in I(i)
7: Calculate gradient magnitude |∇f|(i) and orientation θ(i)
8: Divide I(i) into cells of size 8×8 pixels
9: Compute histogram Hc(i) of gradient orientations for each cell c
10: Group cells into blocks of size 2×2 cells with 50% overlap
11: Normalize histograms within each block to form HOG feature map M(i)
12: end for
13: for all M(i) ∈ { M(1), ., M(itrain) } do
14: Compute GLCM matrices C_θ(i) at angles θ ∈ {0°, 45°, 90°, 135°}
15: Extract statistical features:
F_θ(i) = [Contrast, Homogeneity, Energy, Correlation, Entropy] from each C_θ(i)
16: Average features across all angles: F(i) = (1/4) $\sum_\theta F_\theta$(i)
17: End for
18: Merge feature vectors F(1),.., F(itrain) into one matrix F
19: Train classifiers (SVM, KNN, DT, Discriminator) with F and labels L
20: For all $P_j \in D_{test}$ = {Pj|j = 1,.., j_{test}} do
21: Extract HOG-GLCM features F(j) using steps 2–16
22: Classify P_j using trained models and F(j)
23: Evaluate performance metrics (accuracy, precision, recall, F1-score)
24: End for

HOG-GLCM: Histogram of oriented gradients-gray level co-occurrence matrix

3.5. Experimental Settings and Implementation Details

The experimental conditions were carefully chosen to ensure fair comparisons between different feature extraction methods and classification algorithms. The study used a high-performance system with an Intel Core i7 processor, 8GB of RAM, and an NVIDIA GeForce RTX 3080 GPU for calculations. The dataset was divided into 70% for training and 30% for testing, ensuring fair representation of classes. A 10-fold stratified cross-validation was used for cross-validation to ensure accuracy and minimize random changes in data partitioning. Factors for feature extraction were improved through extensive testing. The cell size for HOG extraction was set to 8 × 8, the block size to 2 × 2, and the orientation bins to 9. The GLCM analysis set the displacement distance at 1 pixel and performed four directional analyses to gather information on texture. Bayesian optimization methods in MATLAB's Optimization Toolbox were used to optimize hyperparameters, with the negative cross-validation accuracy set as the objective function. The optimization process was allowed to run for 100 iterations to find the best or nearly best solutions.

3.6. Classifications and Performance Evaluation Metrics

Classification is a machine learning methodology wherein training data are utilized to construct models, which are subsequently employed to predict new data. The constructed model is assessed using testing data to determine algorithm performance. Pixel-based image classification systems examine the numerical characteristics of designated image feature vectors and categorize data accordingly. This work employs four distinct classification techniques: (DT), (SVM), (KNN), and discriminator (DR) to categorize epilepsy from image datasets. It is worth to mention that the created data set features vector has been dismasted with a ratio of 70% for training and 30% for testing with 10 k-folds. The performance of the proposed classification technique for elliptical seizure detection is evaluated using confusion matrix metrics, including true positives (TP), false positives (FP),

true negatives (TN), and false negatives (FN). These metrics enable the calculation of critical performance indicators such as follows [25]:

1. Sensitivity (Recall): Measures the proportion of correctly identified seizure events

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

2. Precision: Indicates the reliability of positive predictions

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

3. F1-Score: Balances precision and recall through the harmonic mean

$$F1 - Score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \tag{12}$$

4. Accuracy: Represents overall correct classifications

$$Accuracy = \frac{TN + TP}{TN + TP + FP + FN} \tag{13}$$

These metrics are widely used in medical diagnostics, particularly for epileptic seizure detection, where distinguishing abnormal EEG signals from normal ones is critical.

4. RESULTS AND DISCUSSION

All experiments and calculations were conducted in MATLAB version 2023b. Fig. 4a-e demonstrates the box plots for the five key statistical features, namely entropy, contrast, energy, correlation, and homogeneity, derived from all 234 EEG images scanned from the Erbil EEG Epilepsy Dataset. This leads to generate a feature vector of size 345 by 5, equally divided between the seizure and seizure images. The features plot depicted the remarkable separation between the previously mentioned two groups of classes (non-epileptic seizure and epileptic seizure). Consequently, Fig. 4a-e shows the box plots for the five most important statistical features that were taken from all 234 EEG images scanned from the Erbil EEG Epilepsy Dataset. These features are entropy, contrast, energy, correlation, and homogeneity. These box plots indicate important information about how well the derived features can mark the difference between normal and epileptic EEG patterns. Fig. 4's analysis shows a number of important things about the retrieved features. The entropy characteristic (Fig. 4a) makes it easy to tell the difference between normal and epileptic instances. Epileptic signals have higher entropy values, which means that the textural patterns are more complicated and random. This conclusion fits with what doctors already know: Seizures cause neurons to fire in a more random way than regular brain activity. The contrast characteristic (Fig. 4b) shows big variations between the two classes, with epileptic signals having higher contrast values. This means that seizure activity makes that the converted EEG images show more pronounced changes in intensity, which is what happens when the raw signals fluctuate quickly in amplitude. Energy values (Fig. 4c) are lower for cases of epilepsy than for normal cases, which means that the textural patterns are less consistent during seizures. This finding backs up the idea that seizures change the typical patterns of brain activity. Fig. 4d shows that there are various patterns between normal and epileptic cases. For example, epileptic signals have different spatial correlation features. This means that the spatial correlations that GLCM analysis finds are useful for telling things apart. Fig. 4e shows that the classes are clearly different from one other in terms of homogeneity. Normal EEG signals have higher homogeneity values, which means that the patterns are more uniform than the patterns that happen during a seizure. This analysis makes a feature vector that is 234 by 5, with half of the images being non-seizure and the other half being seizure images. The characteristics analysis showed a clear difference between the two groups of classes we talked about earlier (non-epileptic seizure and epileptic seizure).

Consequently, around 70 images are randomly selected, with 35 from each class – non-seizure and seizure. Various comparative scenarios for feature extraction are evaluated using different machine learning techniques, including HoG features alone, GLCM followed by statistical features alone, and eventually, the cascaded HOG-GLCM method followed by statistical features extraction. Table 1 presents a detailed assessment of HPO for the four employed classifiers, which reveals that SVM outperforms other classifiers, achieving the lowest objective function value of 0.02439. KNN shows the highest potential for improvement, with projected values between 0.012195 and 0.0088702. The DT exhibits full stability in its hyperparameters, indicating a local optimum. SVM performs best with standardization applied, while KNN yields better results without it. The Box Constraint parameter for SVM reveals significant variation between observed and calculated values, suggesting room for optimization. Table 1 also gives clear insight into the trade-off between model complexity and performance, assisting in model selection and optimization for epileptic seizures. A comprehensive evaluation of epileptic seizure detection using different feature extraction methods and classification algorithms has been achieved in this article which reveals significant insights into the optimal approach for EEG signal analysis.

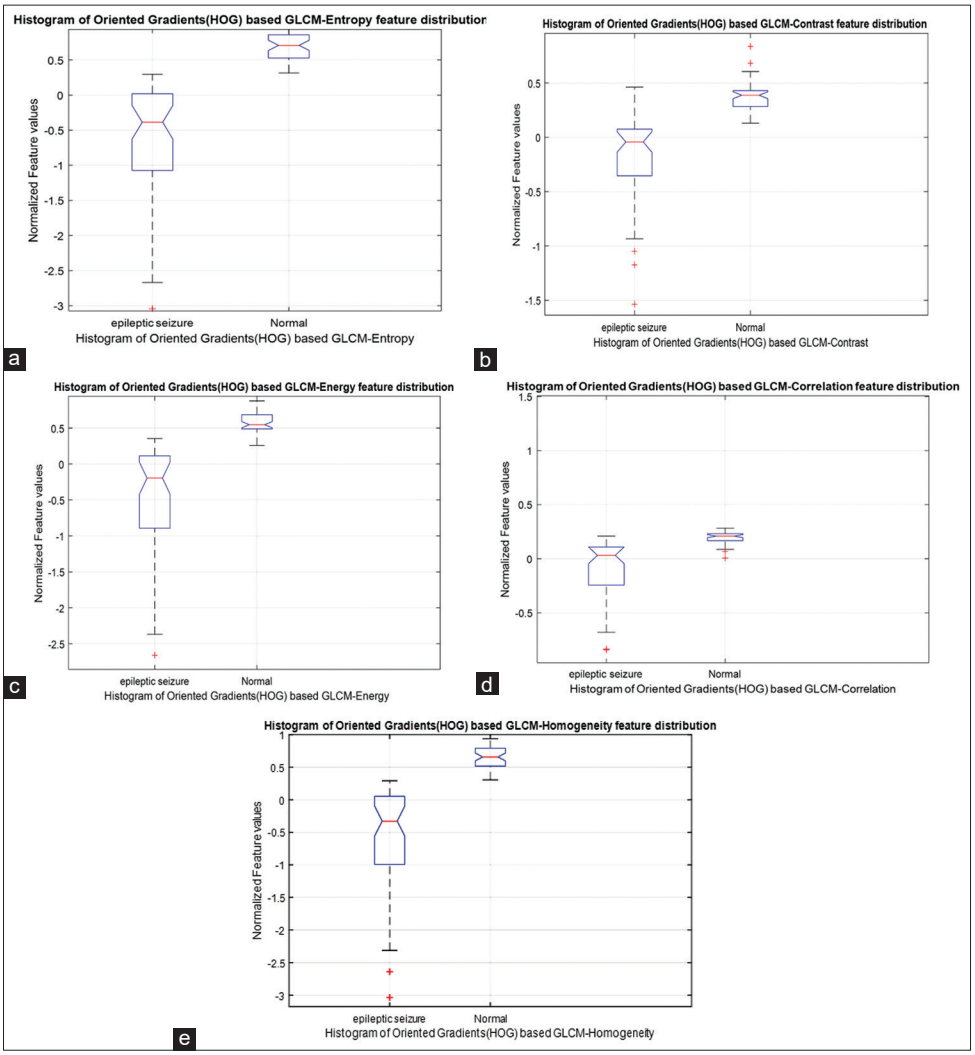


Fig. 4. (a-e) A box plot of the histogram of oriented gradients gray level co-occurrence matrix statistical features, the five features (Entropy, Contrast, Energy, Correlation, and Homogeneity) extracted from the Erbil electroencephalogram epilepsy dataset.

TABLE 1: Unified hyperparameter optimization table for all classifiers

Classifier	Hyperparameter	Observed value	Estimated value	Observed objective	Estimated objective
Discriminator	Delta	1.0337e−06	5.9012e−06	0.085366	0.091406
	Gamma	0.0013391	0.00015939		
SVM	Box constraint	60.163	548.88	0.02439	0.024324
	Kernel scale	0.0010041	0.0010293		
	Standardize	True	true		
Decision tree	Min leaf size	24	24	0.041667	0.041328
KNN	Num neighbors	5	5	0.012195	0.0088702
	Distance	Minkowski	Minkowski		
	Standardize	False	False		

SVM: Support vector machine, KNN: K-Nearest Neighbors

The confusion matrices and performance metrics clearly demonstrate the superiority of the cascaded HOG-GLCM approach over either feature extraction method used independently. As depicted in Fig. 5, which displays the confusion matrices for different classifiers in the case of cascaded features, the SVM classifier achieved near-perfect

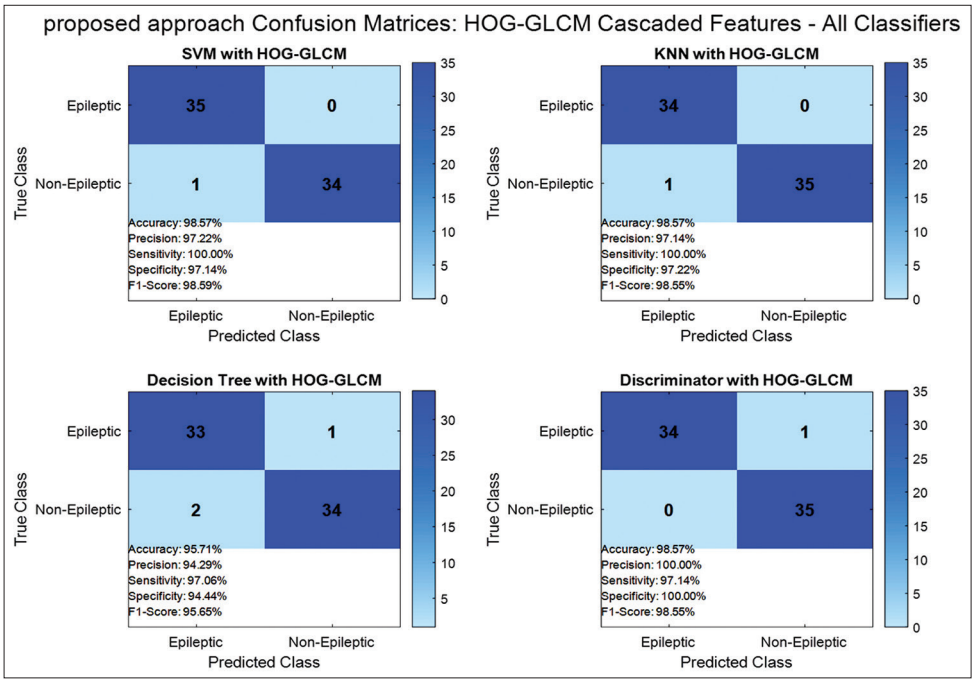


Fig. 5. Confusion matrices for different classifiers in the case of cascaded features (histogram of oriented gradients-gray level co-occurrence matrix).

performance with the recording an accuracy of 98.57%, precision of 97.22%, and most notably, a sensitivity of 100%. This perfect sensitivity indicates that the model successfully identified all epileptic cases without any false negatives – a critical factor in clinical applications where missing a seizure event could have serious consequences. The SVM's F1-score of 98.59% further confirms the balanced performance between precision and recall. Similarly, the KNN classifier with HOG-GLCM features matched the SVM's impressive performance, also achieving 98.57% accuracy and 100% sensitivity. The confusion matrix reveals that both SVM, KNN along with DR misclassified only a single case, resulting in a specificity of 97.14%, 97.22%, and 100%, respectively. This minimal error rate demonstrates the robustness of the cascaded feature approach when paired with these classifiers.

The DT classifier, while slightly less accurate at 95.71%, still performed admirably; its confusion matrix shows one false negative and two false positives, resulting in a sensitivity of 97.06% and specificity of 94.44%. The DR model achieved 98.57% accuracy with perfect precision and specificity (both 100%), misclassifying only one epileptic case as non-epileptic. This exceptionally high degree of sensitivity is especially important in clinical settings because it shows that the model correctly identified all cases of epilepsy without any false negatives. This is very important because missing a seizure could have serious consequences for patient safety and treatment plans.

As an alternative scenario for feature extraction to substantiate the superiority of the proposed feature extraction technique. Fig. 6 depicts the confusion matrices for different classifiers when using GLCM features followed by statistical feature extraction, all classifiers showed a noticeable decrease in performance compared to the cascaded approach. The SVM classifier achieved the highest accuracy among GLCM-only models at 92.86%, with balanced precision and sensitivity (both 91.43%). The confusion matrix reveals three false negatives and two false positives, indicating increased difficulty in class separation using only texture-based features. The KNN classifier with GLCM features performed slightly worse with 91.43% accuracy and identical precision, sensitivity, and F1-score values of 91.18%. The DT classifier struggled the most with GLCM features, achieving only 88.57% accuracy and 85.71% sensitivity, missing five epileptic cases. The DR model reached 90% accuracy with GLCM features, showing moderate performance with a sensitivity of 88.57%. These results suggest that while GLCM features extract significant textural information from EEG signal images, which are inadequate for effective seizure identification, especially when high sensitivity is essential.

Eventually, Fig. 7 demonstrated the confusion matrices for different classifiers when using HOG feature extraction, which performed better than GLCM features but still fell short of the cascaded approach. The SVM classifier

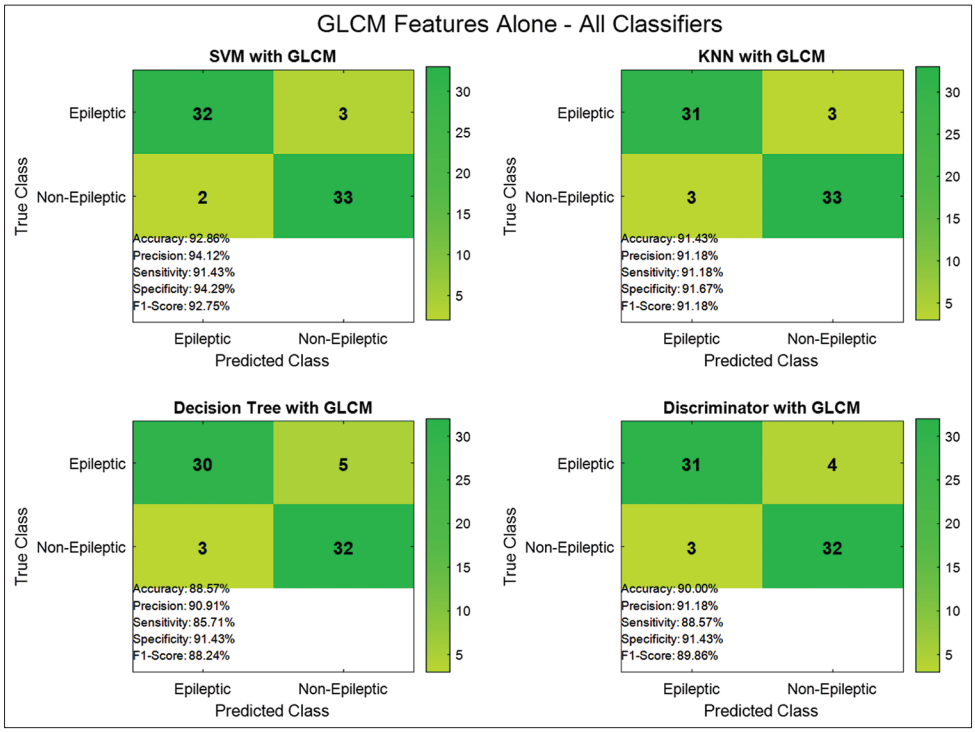


Fig. 6. Confusion matrices for different classifiers when using gray-level co-occurrence matrix features, followed by statistical feature extraction.

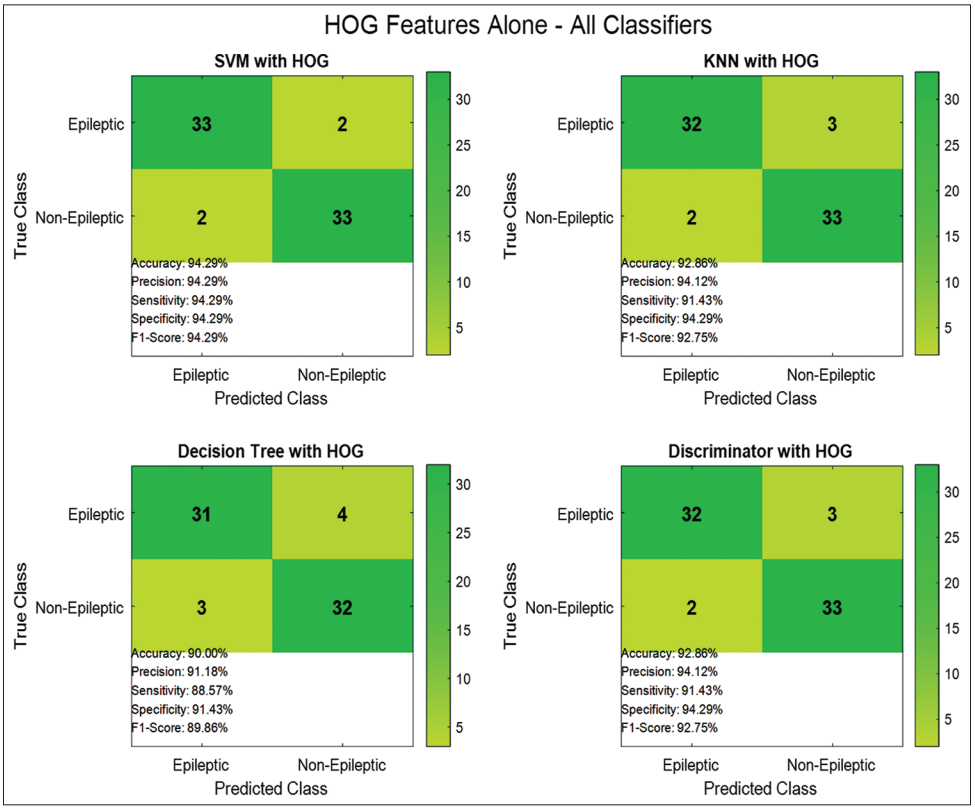


Fig. 7. Confusion matrices for different classifiers when using the histogram of oriented gradients feature extraction.

TABLE 2: The complete performance measures for all scenarios of feature extraction

Classifier	Feature set	Accuracy	Precision	Sensitivity	Specificity	F1-score
SVM	HOG-GLCM	98.571	97.222	100	97.142	98.591
SVM	GLCM	92.857	94.117	91.428	94.285	92.753
SVM	HOG	94.285	94.285	94.285	94.285	94.285
KNN	HOG-GLCM	98.571	97.142	100	97.222	98.550
KNN	GLCM	91.428	91.176	91.176	91.666	91.176
KNN	HOG	92.857	94.117	91.428	94.285	92.753
Decision tree	HOG-GLCM	95.714	94.285	97.058	94.443	95.652
Decision tree	GLCM	88.571	90.909	85.714	91.428	88.235
Decision tree	HOG	90	91.176	88.571	91.478	89.855
Discriminator	HOG-GLCM	98.571	100	97.142	100	98.550
Discriminator	GLCM	90	91.176	88.571	91.428	89.855
Discriminator	HOG	92.8572	94.117	91.428	94.284	92.753

SVM: Support vector machine, KNN: K-Nearest Neighbors,
HOG-GLCM: Histogram of oriented gradients-gray level co-occurrence matrix

with HOG features achieved 94.29% accuracy with balanced metrics across precision, sensitivity, specificity, and F1-score (all 94.29%). The confusion matrix shows two false positives and two false negatives, indicating a symmetrical error distribution. The KNN classifier with HOG features reached 92.86% accuracy with slightly lower sensitivity (91.43%) than precision (94.12%). The DT classifier achieved 90% accuracy, while the DR model performed slightly better at 92.86% accuracy with HOG features. These results indicate that gradient-based HOG features capture more discriminative information than GLCM features alone but still lack the comprehensive representation provided by the cascaded approach. The superior performance of the cascaded HOG-GLCM approach demonstrates a clear synergistic effect between gradient-based and texture-based features. This synergy can be attributed to the complementary nature of the information captured by each method. Table 2 reports all performance measures for all scenarios of feature extraction.

5. COMPUTATIONAL COMPLEXITY AND REAL-TIME CONSIDERATIONS

The HOG feature extraction part takes $O(N \times M \times B)$ time to run, where $N \times M$ is the size of the picture and B is the number of orientation bins. This means that each of the 256×256 pixel images with 9 orientation bins will need about 590,000 operations. The GLCM calculation is $O(L^2 \times N \times M)$, where L is the number of gray levels. With 256 gray levels, this takes about 16.7 million processes for each image. It takes an average of 0.15 s to extract features from one EEG image, but the time it takes to classify them varies by algorithm.

6. LIMITATIONS AND CONSTRAINTS

The proposed method for automatically detecting epilepsy has limitations due to the large dataset, difficulty in calculations, and the need for comparison and validation. The dataset consists of 234 EEG recordings from one institution, which may not fully represent real-world clinical populations. The cascaded HOG-GLCM method requires more processing power than simpler methods, making it less useful for real-time applications. The method also creates a high-dimensional feature space, making it difficult to use on limited resources or portable devices. Future research should focus on expanding datasets, validating across institutions, optimizing real-time, combining deep learning methods, and conducting clinical validation.

7. CONCLUSION

This study introduces a novel cascaded HOG-GLCM feature extraction methodology for epileptic seizure detection from EEG signal images. The comprehensive evaluation using the Erbil EEG Epilepsy Dataset demonstrates that this innovative approach substantially outperforms single-feature extraction methods, with a significant 5–8% improvement in accuracy. The synergistic integration of complementary feature extraction techniques – HOG capturing directional patterns characteristic of seizure activity and GLCM extracting spatial relationships – yields remarkable classification performance. SVM and KNN classifiers achieved 98.57% accuracy with perfect sensitivity (100%), ensuring no epileptic events went undetected. Most notably, the SVM classifier attained an exceptional F1-score of 98.59%, demonstrating balanced

precision (97.22%) and recall with minimal error rate. Our comparative analysis quantitatively validates the superiority of the cascaded approach over GLCM-only (92.86% accuracy) and HOG-only (94.29% accuracy) methods. The rigorous Bayesian HPO further enhanced classification performance across all tested algorithms. The establishment of the first comprehensive epilepsy dataset from Erbil, Kurdistan, comprising carefully collected and preprocessed EEG recordings, addresses the critical need for diverse and representative data resources in epilepsy research. Future work should focus on computational optimization for real-time applications and integration with deep learning techniques while maintaining the high accuracy and perfect sensitivity demonstrated in this study, ultimately contributing to improved epilepsy management and patient outcomes worldwide.

8. ACKNOWLEDGMENT

The authors would like to thank Salahaddin University–Erbil for supporting this article and also the Erbil teaching hospital, especially Dr. Tahsin Muhsin Qadr, head of the EEG department.

REFERENCES

[1] V. L. Feigin, T. Vos, B. S. Nair, S. I. Hay, Y. H. Abate, A. H. A. Abd Al Magied, S. A. E. A. Abdelkader, M. A. Abdollahifar, A. Abdullahi, R. G. Aboagye, L. G. Abreu, S. A. Rumeileh, H. Abualruz,…& C. J. L. Murray. "Global, regional, and national burden of epilepsy, 1990-2021: A systematic analysis for the global burden of disease study 2021". *The Lancet Public Health*, vol. 10, no. 3, pp. e203-e227, 2025.

[2] M. Leonardi, P. Martelletti, R. Burstein, A. Fornari, L. Grazzi, A. Guekht, R. B. Lipton, D. D. Mitsikostas, J. Olesen, M. O. Owolabi, E. Ruiz De la Torre, S. Sacco, T. J. Steiner, N. Surya, T. Takeshima, C. Tassorelli, S. J. Wang, T. Wijeratne, S. Yu and A. Raggi. "The world health organization intersectoral global action plan on epilepsy and other neurological disorders and the headache revolution: From headache burden to a global action plan for headache disorders". *The Journal of Headache and Pain*, vol. 25, no. 1, p. 4, 2024.

[3] F. Lado, S. Ahrens and E. Riker. "*Guidelines for Specialized Epilepsy Centers: Report of the National Association of Epilepsy Centers Guideline Panel*". vol. 386, 2024. https://lww.com/wnl/d [Last accessed on 2025 May 18].

[4] M. Mursalin, Y. Zhang, Y. Chen and N. V. Chawla. "Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier". *Neurocomputing*, vol. 241, pp. 204-214, 2017.

[5] Z. Chen, M. I. Maturana, A. N. Burkitt, M. J. Cook and D. B. Grayden. "Seizure forecasting by high-frequency activity (80-170 Hz) in long-term continuous intracranial EEG recordings". *Neurology*, vol. 99, no. 4, pp. e364-e375, 2022.

[6] O. K. Fasil and R. Rajesh. "Time-domain exponential energy for epileptic EEG signal classification". *Neuroscience Letters,* vol. 694, pp. 1-8, 2019.

[7] S. Chakrabarti, A. Swetapadma and P. K. Pattnaik. A review on epileptic seizure detection and prediction using soft computing techniques. In: "*Smart Techniques for a Smarter Planet Towards Smarter Algorithms"*. Springer, Berlin, pp. 37-51, 2019.

[8] A. H. Shoeb. *"Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment*". Massachusetts Institute of Technology, United States, 2009.

[9] S. M. Usman, S. Latif and A. Beg. "*Principle Components Analysis for Seizures Prediction using Wavelet Transform*". arXiv [Preprint], 2020.

[10] A. Anuragi and D. S. Sisodia. "Empirical wavelet transform based automated alcoholism detecting using EEG signal features". *Biomedical Signal Processing and Control*, vol. 57, p. 101777, 2020.

[11] U. Ayman, M. S. Zia, O. D. Okon, N. U. Rehman, T. Meraj, A. E. Ragab and H. T. Rauf. "Epileptic patient activity recognition system using extreme learning machine method". *Biomedicines*, vol. 11, no. 3, p. 816, 2023.

[12] M. S. Farooq, A. Zulfiqar and S. Riaz. "Epileptic seizure detection using machine learning: Taxonomy, opportunities, and challenges". *Diagnostics (Basel)*, vol. 13, no. 6, p. 1058, 2023.

[13] G. Chandel, S. K. Saini and A. Sharma. Epileptic eeg signal classification using machine learning based model. In: "*2023 International Conference on Disruptive Technologies (ICDT)"*. IEEE, United States, pp. 733-739, 2023.

[14] H. F. Atlam, G. E. Aderibigbe and M. S. Nadeem. "Effective epileptic seizure detection with hybrid feature selection and SMOTE-based data balancing using SVM classifier". *Applied Sciences*, vol. 15, no. 9, p. 4690, 2025.

[15] S. S. I. Umar. "*Accelerated Histogram of Oriented Gradients for Human Detection*". Universiti Teknologi Malaysia, Malaysia, 2016.

[16] H. Kataoka, K. Hashimoto, K. Iwata, Y. Satoh, N. Navab, S. Ilic and Y. Aoki. Extended co-occurrence hog with dense trajectories for fine-grained activity recognition. In: "*Computer Vision--ACCV 2014: 12th Asian Conference on Computer Vision". Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part V 12*". Springer, Berlin, pp. 336-349, 2015.

[17] W. Zhou, S. Gao, L. Zhang and X. Lou. Histogram of oriented gradients feature extraction from raw bayer pattern images. In: "*IEEE Transactions on Circuits and Systems II: Express Briefs*". vol. 67. IEEE, United States, pp. 946-950, 2020.

[18] M. G. Sarowar, M. A. Razzak and M. A. Al Fuad. HOG feature descriptor based PCA with SVM for efficient & accurate classification of objects in image. In: "*2019 IEEE 9th International Conference on Advanced Computing (IACC)*". IEEE, United States, pp. 171-175, 2019.

[19] S. Bakheet and A. Al-Hamadi. "A framework for instantaneous driver drowsiness detection based on improved HOG features and naïve bayesian classification". *Brain Sciences*, vol. 11, no. 2, p. 240, 2021.

[20] R. M. Haralick, K. Shanmugam and I. H. Dinstein. Textural features for image classification. In: "*IEEE Transactions on Systems, Man, and Cybernetics*". IEEE, United States, pp. 610-621, 1973.

[21] C. Thontadari and C. J. Prabhakar. "Scale space co-occurrence HOG features for word spotting in handwritten document images". *International Journal of Computer Vision and Image Processing (IJCVIP)*, vol. 6, no. 2, pp. 71-86, 2016.

[22] P. T. Krishnan, S. K. Erramchetty and B. C. Balusa. "Advanced

framework for epilepsy detection through image-based EEG signal analysis". *Frontiers in Human Neuroscience*, vol. 18, p. 1336157, 2024.

[23] J. Poza, C. Gómez, A. Bachiller and R. Hornero. "Spectral and nonlinear analyses of spontaneous magnetoencephalographic activity in Alzheimer' s disease". *Journal of Healthcare Engineering*, vol. 3, no. 2, pp. 299-322, 2012.

[24] S. O. Haji and R. Z. Yousif. "A novel run-length based wavelet features for screening thyroid nodule malignancy". *Brazilian Archives of Biology and Technology*, vol. 62, p. e19170821, 2019.

[25] Aayesha, M. B. Qureshi, M. Afzaal, M. S. Qureshi and M. Fayaz. "Machine learning-based EEG signals classification model for epileptic seizure detection". *Multimedia Tools and Applications*, vol. 80, no. 12, pp. 17849-17877, 2021.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Demographic and Clinical Profile of Patients Undergoing Colonoscopy in the Teaching Hospital for Gastroenterology and Hepatology in Sulaymaniyah City, Kurdistan/Iraq

Awat Jamal Hasan

University of Sulaimani, College of Nursing, Department of Community Health Nursing, Sulaimaniyah, Kurdistan Region, Iraq

ABSTRACT

Colonoscopy serves as the primary diagnostic modality, screening, and therapeutic instrument for a wide range of pathologies of the rectum, colon, and terminal ileum. The aim of this study is to evaluate the demographic, clinical, and histological parameters of the records of patients who underwent colonoscopy at the Teaching Hospital for Gastroenterology and Hepatology in Sulaymaniyah City. This is a retrospective cross-sectional study that included the hospital records of all the patients who had undergone colonoscopy from 2022 to 2024 in the Teaching Hospital for Gastroenterology and Hepatology in Sulaymaniyah City. The study involved 933 patients, with a mean age of 46.6 years. The diagnostic yield of colonoscopy was 62.43%, with colorectal polyps being seen 17.04% of the time and ulcer lesions being inconclusive 5.03% of the time. Internal hemorrhoids were 11.78%. The odds ratio for colorectal adenocarcinoma (CRC) was 9.86 for ages >50 years and 0.14 for females. Age ≥50 years was associated with a higher chance of adenocarcinoma compared to those <50 years, but the confidence interval is wide owing to limited adenocarcinoma incidence (n = 27). The study shows colonoscopy is most used for gastrointestinal bleeding, bowel habits, and chronic abdominal pain, with underutilization in developed countries and lower patient age in Kurdish-inhabited area.

Index Terms: Colonoscopy, Colorectal Cancer, Colorectal Polyps, Retrospective Study, Diagnostic Yield

1. INTRODUCTION

Lower gastrointestinal symptoms are commonly encountered in patients attending internal medicine and gastroenterology outpatient clinics. While some of these symptoms can be diagnosed clinically, others require endoscopic evaluation, particularly colonoscopy [1]. Colonoscopy is a widely used method for diagnosing and treating gastrointestinal diseases, remaining the "gold standard" for most large intestine diseases. With increasing cases, demand and availability of colonoscopy examinations are increasing. New methods for cleansing the intestine and administration patterns are emerging [2]. Colonoscopy is a crucial procedure for diagnosing, screening for, and treating gastrointestinal abnormalities in the rectum, colon, and terminal ileum. As the gold standard for evaluating lower gastrointestinal symptoms, it enables detection of colonic polyps, ulcers, bleeding, inflammation, and tumors. It is also considered the best modality for colon cancer screening and surveillance [3,10]. Colonoscopy is an endoscopic procedure that evaluates the entire colon and performs therapeutic procedures using a flexible tube and fiber-optic camera. It passes through the anus of patients after proper preparation. The

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp129-135

E-ISSN: 2521-4217

P-ISSN: 2521-4209

Copyright © 2025 Awat Jamal Hasan. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Dr. Awat Jamal Hasan University of Sulaimani, College of Nursing, Department of Community Health Nursing, Sulaimaniyah, Kurdistan Region, Iraq. E-mail: awat.hassan@univsul.edu.iq. Tel: +964 7702204005 | +964 7718322922

Received: 29-03-2025 Accepted: 18-07-2025 Published: 15-09-2025

procedure detects gastrointestinal abnormalities such as polyps, malignant masses, ulcerations, and bleeds. A biopsy confirms diagnosis, potentially diagnosing diseases such as inflammatory bowel diseases (IBD) and polyps. It also aids in therapeutics such as polypectomy, electrocautery, and radiofrequency ablation therapy for oncological patients. Regular colonoscopies are conducted in our institution for various indications such as abdominal pain, bleeding, melena, bowel habits changes, chronic diarrhea, anemia, weight loss, and malignancy suspicion [1]. Colonoscopy plays a crucial role in clinical practice, but there are limited studies on patients in Sulaymaniyah City, Kurdistan/Iraq. For clarity, we define key terms: colorectal cancer (CRC) refers to malignant neoplasms of the colon or rectum regardless of histology; adenocarcinoma specifically denotes the gland-forming epithelial malignancy constituting 90–95% of CRCs [3]. These terms will be used consistently throughout. This study aims to assess demographic characteristics, clinical presentations, colonoscopy observations, and histological findings of patients undergoing colonoscopy in Sulaymaniyah City for Kurdistan/Iraq, based on records from the Teaching Hospital for Gastroenterology and Hepatology [22].

2. MATERIALS AND METHODS

2.1. Study Design and Setting

This retrospective cross-sectional study included hospital records of all patients who underwent colonoscopy between January 1, 2022, and December 31, 2024, at the Gastroenterology Unit of the Teaching Hospital for Gastroenterology and Hepatology in Sulaymaniyah City. The study was approved by the Scientific and Ethics Committees of the College of Nursing, University of Sulaimani, Sulaimaniyah, Iraq. Final administrative approval for data collection was also obtained from the Sulaimani General Directorate of Health and the Sulaimani General Teaching Hospital.

2.2. Procedure of Colonoscopy

All patients adhered to a liquid diet on the day preceding colonoscopy. Each received oral administration of a 10 mg bisacodyl pill approximately 6–8 h before initiating overnight fasting. On the procedure day, patients consumed 2 L of bowel preparation solution consisting of 118 g polyethylene glycol 4000, 2.93 g sodium chloride, 1.48 g potassium chloride, 3.37 g sodium bicarbonate, and 11.36 g anhydrous sodium sulfate dissolved in 2 L of water. Patients were provided detailed instructions and guidance from trained nurses to ensure compliance with bowel preparation. Effective preparation was confirmed when patients experienced 8–10 episodes of loose stools.

Inadequate bowel preparation was noted when the patient failed to adhere to the specified instructions for bowel preparation or when the mucosa was not sufficiently visualized during colonoscopy. All colonoscopies were conducted utilizing Olympus CFQ145L, CF-Q165I, or CF-Q165L video colonoscopes. Tissue samples from visible lesions, when required, were collected and sent for histopathological analysis.

2.3. Study Participants and Selection Criteria

The study, conducted at a teaching hospital specializing in gastroenterology and hepatology, aimed to refine patient records for the analysis of colonoscopy procedures. From an initial 1020 patient records, 28 were excluded due to incomplete data, and 59 (5.7%) records were excluded because the patients had inadequate bowel preparation at the time of the colonoscopy. After exclusions for incomplete records and inadequate bowel preparation Figure 1, the final cohort comprised 933 patients.

2.4. Data Collection

The researcher reviewed patient entries in the physical colonoscopy register, retrieved relevant data, and entered it into a spreadsheet using IBM statistical package for social sciences (SPSS) (IBM SPSS Statistics for Windows, Version 27.0, IBM Corp). Indications labeled as "Other" in Table 1 encompassed chronic constipation, Crohn's disease, foreign bodies, rectovaginal fistulas, thickened bowel walls, and rare/unspecified indications (e.g., ischemic colitis). The study maintained participant confidentiality and anonymity, ensuring no personal or identifiable data were collected from the dedicated staff who maintained the register without further involvement.

2.5. Study Variables

The variables studied in this study included gender, age group, indications for colonoscopy: The reasons or symptoms prompting the colonoscopy procedure, observations made during the procedure, such as the presence of

TABLE 1: Indications for colonoscopy in the study participants (*n*=933)

Indication	Count	Percentage
Altered bowel habit	209	22.40
Rectal mass	88	9.43
Cancer screening or surveillance	73	7.82
Anemia	90	9.64
Lower gastrointestinal bleeding	246	26.36
Persistent or recurrent abdominal pain	207	22.18
Other*	20	2.14

Other* includes chronic constipation (*n*=6), Crohn's disease (*n*=7), foreign bodies (*n*=3), rectovaginal fistulas (*n*=2), thickened bowel walls (*n*=1), and unspecified indications (*n*=1).

polyps, abnormalities, or other significant findings, and histopathological findings.

2.6. Study Size

We considered the presence of adenocarcinomas the primary outcome. A standard formula for cross-sectional studies [4] was used: sample size $n = Z^2_{a/2} * P * (1–P)/d^2$, where $Z_{a/2}$ was set to 1.96 for a 95% level of confidence, P is the estimated proportion of prevalence of the main outcome, and d is the level of precision that was wanted, which was set to 0.02. We implemented a convenience sampling method to select participants. However, the study included all patients who met the study criteria, resulting in a final sample size of 933. Pregnant women were not part of the study population; no cases of pregnancy were recorded in the dataset.

2.7. Statistical Analysis

The study used continuous variables (mean, 95% confidence interval [CI], standard deviation [SD], range) and categorical variables (frequency and percentages). We used Pearson's two-test to evaluate the association between categorical variables, considering a $P < 0.05$ as statistically significant. The strength of associations was assessed using the odds ratio (OR). All statistical analyses were performed using the SPSS (SPSS Statistics for Windows, Version 27.0, IBM Corp).

3. RESULTS

3.1. Indications for Colonoscopy

This study included 933 patients. The mean age of the patients was 46.6 years (95% CI: 45.5, 47.7). The SD was 16.4 years, with ages ranging across the available dataset. Among the study participants, 458 (49.1%) were male, and 475 (50.9%) were female, yielding a male-to-female ratio of 0.96:1. Figure 2 shows the gender distribution of the study participants.

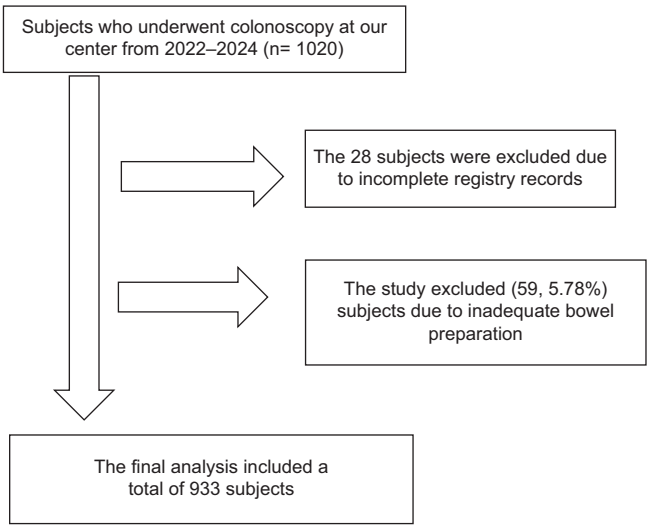


Fig. 1. The flowchart displays the participant's inclusion and exclusion from the cross-sectional study.

3.2. Indications for Colonoscopy

Common indications for colonoscopy included lower gastrointestinal bleeding (n = 246, 26.36%) and persistent or recurrent abdominal pain (n = 207, 22.18%). In addition, a change in bowel habits was also a common indication (n = 209, 22.40%). Other indications included cancer screening or surveillance (n = 73, 7.82%), anemia (n = 90, 9.64%), and rectal mass (n = 88, 9.43%). The study also found that some patients had abnormalities on abdominal imaging, such as bowel walls that were thicker (n = 1), foreign bodies (n = 3), rectovaginal fistulas (n = 2), chronic constipation (n = 6), and Crohn's disease (n = 7). Table 1 illustrates the reported indications.

3.3. Findings of Colonoscopy

Colonoscopy showed no abnormalities in 350 cases (37.51%). Therefore, the overall diagnostic yield of colonoscopy was 62.43%. Table 2 shows the diagnostic yield values according to the various indications for colonoscopy. The research determined the polyp detection rate (PDR) to be 17.04%, 110 cases of internal hemorrhoids (11.88%), 98 cases of colitis (10.5%), 47 cases of ulcer lesions (5.03%), 29 cases of IBD (3.1%), 28 cases of mass lesions (3%), 26 cases of proctitis (2.78%), and 47 cases of ulcer lesions (5.03%). Other findings were n = 86, 9.2%. Fig. 3 illustrates the reported findings.

3.4. Histopathological Findings

In the current study, 39.33% (n = 367) of the study subjects' cases were examined histopathologically. Of these 367 cases,

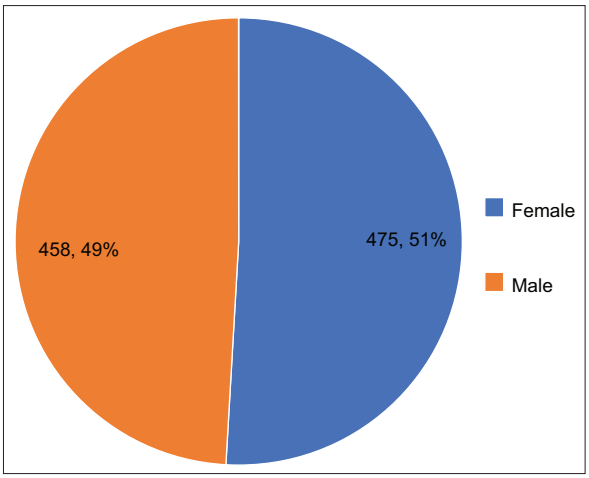


Fig. 2. Gender distribution of the study participants (n = 933).

78 did not reveal any histopathological abnormalities. Table 3 shows that adenocarcinoma (n = 27), neoplastic lesions (n = 20), hamartomatous polyps (n = 13), colonic lipomas (n = 12), colitis "Juvenile polyp is same" (n = 9), and ileitis (n = 6) were some of the most common abnormal findings. In 22 cases, histopathology records were unavailable for the histopathological examination.

TABLE 2: Diagnostic yield of colonoscopy for various indications

Indication	Significant finding present	Total no. cases	Diagnostic yield (%)
Altered bowel habits	98	209	46.88
Cancer screening or surveillance	48	73	65.75
Anemia	64	90	71.11
Lower gastrointestinal bleeding	169	246	68.99
Persistent or recurrent abdominal pain	123	207	59.42
Rectal mass	68	88	77.27

TABLE 3: Histopathological findings in patients who underwent biopsy (n=367)

Histopathological finding	Frequency (n)	Percentage
Hyperplastic polyp	82	22.34
No abnormality detected	78	21.25
Ulcerative colitis	41	10.89
Adenocarcinoma	27	7.35
Neoplastic lesions	20	5.44
Hamartomatous polyp	13	3.54
Colonic lipoma	12	3.26
Colitis	11	2.99
Inflammatory lesions	9	2.45
Juvenile polyp	9	2.45
Ileitis	6	1.63
Records not retrievable	22	5.99
Total	367	100

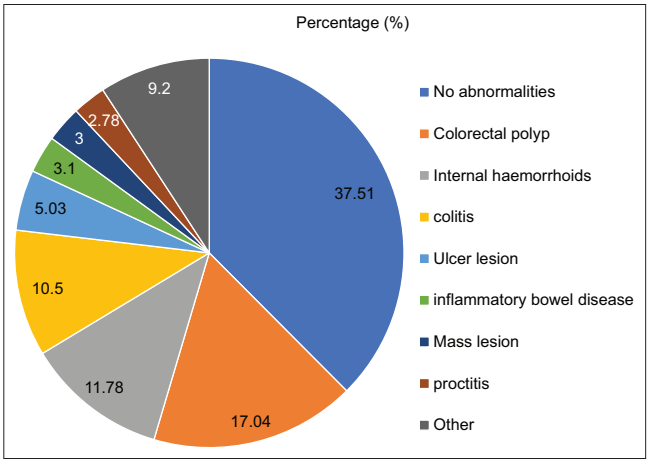


Fig. 3. The study participants (n = 933) underwent colonoscopy to analyze their findings.

3.5. Association between Demographic Parameters and the Presence of Colorectal Adenocarcinoma

Age ≥50 years is associated with a higher likelihood of adenocarcinoma compared to <50 years, but with a very wide CI due to the small number of cases.

The OR is lower in females compared to males for adenocarcinoma, though the wide CI implies statistical uncertainty.

1. Age (≥50 vs. <50)
   - OR: 9.86
   - 95% CI: (0.51, 191.47).
2. Sex (female vs. male)
   - OR: 0.14
   - 95% CI: (0.007, 2.66).

Correlation of demographic factors with the incidence of colorectal adenocarcinoma Pearson's χ^2 test of independence was performed, and it was seen that in patients who underwent colonoscopy, age 50 years and above was associated with the presence of colorectal adenocarcinoma (P = 0.017) with an OR of 9.86 (95% CI: 0.51, 191.47). Furthermore, female gender was not significantly associated with the presence of colorectal adenocarcinoma (P = 0.012, χ^2 test) with an OR of 0.14 (95% CI: 0.007, 2.66). The broad CI is due to statistical imprecision owing to the small number of cases. Results of Pearson's χ^2. test of independence are shown in Table 3.

The wide CIs for both age (95% CI: 0.51–191.47) and sex (95% CI: 0.007–2.66) reflect limited statistical power due to the small number of adenocarcinoma cases (n = 27). While the point estimates suggest strong associations (OR = 9.86 for age ≥50; OR = 0.14 for females), the imprecision cautions against definitive clinical interpretations. This limitation is contextualized further in the discussion.

4. DISCUSSIONS

The average age of the colonoscopy patients in this research was 46.6 years in Sulaymaniyah, Kurdistan/Iraq. This is comparable to the other similar studies done in Baghdad, which was 46.5 [5]. In contrast, a study from Southern Iraq reported an average age of 42.3 years [6]. However, the median age of colonoscopy patients was reported to be a

Table 4: Association between demographic parameters and presence of colorectal adenocarcinoma

Variables	Adenocarcinoma present, *n* (%)	Adenocarcinoma absent, *n* (%)	Total (*n*)	χ^2 statistic	Degrees of freedom	*P*-value
Age group						
≥50 years	3 (0.8)	386 (99.2)	389	5.67	1	0.017
<50 years	0 (0.0)	544 (100.0)	544			
Gender						
Female	0 (0.0)	475 (100.0)	475	6.31	1	0.012
Male	3 (0.7)	455 (99.3)	458			

mean age of 42.84 years in a study done in Erbil, Iraq [7]. In Kurdistan, widespread use of colonoscopy as a screening method for colorectal carcinoma in older individuals is yet to be practiced. This can partly explain why patients undergoing colonoscopy in Kurdistan/Iraq younger are compared to developed countries. Most of the patients who had undergone colonoscopy were women. The male-to-female ratio found was 0.96:1. It has been determined through research that lower socioeconomic status and lower education status among people can be the cause of the lower accessibility to healthcare among men and women during their stay in Kurdistan.

The 16-year age disparity between our cohort (46.6 years) and Western populations (e.g., 62.5 years in USA [8], [9], primarily reflects differential screening utilization. Where >60% of colonoscopies represent screening in developed nations [23], only 7.82% did in our study-consistent with Iraq's lack of organized CRC screening programs. Younger patients thus present predominantly with symptoms rather than prevention intent. Contextual barriers include: (1) absence of national screening guidelines; (2) limited public awareness; (3) resource constraints prioritizing symptomatic care; and (4) cultural reluctance toward preventive procedures-factors similarly observed in resource-limited settings [3], [19]. According to Fatakhova and Rajapakse (2024), the evolution of colon cancer screening and surveillance has necessitated a personal approach to identifying individual risks, optimal surveillance intervals, and the best options for dysplasia management and CRC [12]. The American College of Gastroenterology (ACG) strongly recommends the use of colonoscopy as a screening tool for CRC for individuals between 50 and 75 years [13]. In Iraq, however, according to Dhahi and Almansori, in 2024, the success rate of colonoscopy was relatively lower than the accepted rate because of patient-related factors. It was, however, higher than what was reported in previous studies. Due to the small sample size and the fact that the center was recently opened, we also found a low adenoma detection rate (ADR). The American Society for Gastrointestinal Endoscopy/ACG the Task Force on Quality in Endoscopy recommended that <15% of total colonoscopies performed should have poor bowel preparation [14].

Within this cohort, 5.78% of patients with colonoscopy had inadequate bowel preparation, a higher percentage than in Nepal. This discrepancy likely reflects contextual challenges in our setting, including patient-related factors (e.g., language barriers, low health literacy, and socioeconomic constraints affecting compliance), procedural hurdles (e.g., the complexity of ingesting 2 L of polyethylene glycol solution), and systemic limitations (e.g., lack of dedicated pre-procedure counseling teams and high staff workload). To address these issues, we recommend that trained staff provide simple, easy-to-understand instructions in the local language, frequent reminders, and pamphlets to reduce the percentage of patients with poor bowel preparation [3]. Future quality improvement initiatives could also incorporate phone call reminders and free preparation kits to mitigate socioeconomic barriers, as suggested in similar resource-limited settings [14].

The overall diagnostic yield in our study was 62.43%, with the highest yield in suspected cases of rectal mass (77.72%) followed by anemia (71.11%). The chance of finding a lesion on colonoscopy is maximum in patients with a rectal mass felt on rectal examination or proctoscopy. When we performed a colonoscopy in patients who presented with an altered bowel habit, the diagnostic yield was (46.88%). At our center, in all patients presenting with an altered bowel habit who are suspected of having irritable bowel syndrome (IBS), we perform a colonoscopy to rule out IBD. The ACG recommends testing for fecal calprotectin to rule out IBD in suspected IBS cases [15]. However, since any fecal calprotectin test was not available at our center, we use colonoscopy for the same. This could have been the cause for the low diagnostic yield in patients with an altered bowel habit.

ADR is defined as the percentage of patients who had one or more adenomas detected by colonoscopy [11, 14]. ADR has been widely applied as a measure of quality in colonoscopies.

The American Society for Gastrointestinal Endoscopy/ACG Task Force recommends ADR targets of over 25%, with a high correlation between PDR and ADR reported in several studies [14]. In our center, analysis revealed the PDR to be 17.04%. While CRC was historically less frequent in Asians than in Western populations [16]. Recent studies show Asian incidence rates are escalating and nearing Western levels. This convergence is primarily driven by lifestyle and dietary changes in Asian populations, including increased consumption of processed foods, reduced physical activity, and rising obesity rates associated with rapid urbanization [16]. The Iraq Cancer Registry showed that the incidence of colorectal carcinoma was 6.49% of total body malignancy (Directorate, 2019). Incidence significantly increased throughout the period 2000–2019 in all the age groups but more significantly among adults between 40 and 49 years. The least increase was seen in the age group of 20–29 years in 2019 [17]. Colon cancer incidence is increasing in Iraq, unlike in Western countries where it is decreasing due to proper colonoscopies. There are no studies done in this area, and screening and surveillance guidelines are not followed in the Nepalese healthcare system [18]. The 2024 Taiwan study reveals significant differences in ADR among age groups, suggesting that lowering the screening age may decrease overall ADR. However, the current standard of 25% ADR for screening colonoscopy may still be used. Our results suggest that even if CRC screening starts at age 40 or 45 years, stratified analysis by age/gender remains feasible, showing how common adenomas are in colonoscopy screening by age and gender. CRC is a growing threat all over the globe due to globalization, industrialization, and urbanization. Although it has a low incidence (<6.12/100,000 population), it has been increasing for the past two decades. High rates are found in Australia and Northern Europe; while neighboring countries such as Turkey, Iran, Saudi Arabia, and Jordan have higher rates. Epidemiological projections indicate concerning trends, and Iraq may reach Western IP by 2060 [19]. In the current study, patients who had undergone colonoscopy at the age of <50 were significantly associated with the presence of colorectal adenocarcinoma. This agrees with other studies done in the past in Iraq. CRC occurs in both old and young patients, and the symptoms are the same. Weakness and fatigue are observed in both groups. Most patients over 50 years undergo left colorectal adenocarcinoma [20]. Furthermore, our center found a strong correlation between male gender and the presence of colorectal adenocarcinomas. This is the same as global studies on colorectal cancer, which show higher prevalence in males than in females [21]. This difference may have been the result of differences in genetics, diet, and lifestyle [16]. The study shows colonoscopy is most used for gastrointestinal bleeding, altered bowel habits, and chronic abdominal pain, reflecting underutilization of screening compared to developed countries and a lower average patient age in this Kurdish-inhabited region. However, several limitations should be noted. First, the retrospective design and reliance on hospital records may have introduced selection bias, particularly due to excluded cases with incomplete data or inadequate bowel preparation. Second, the single-center nature limits generalizability to other regions in Iraq with differing demographics or healthcare practices. Third, the small number of adenocarcinoma cases ($n = 27$) resulted in wide CIs, reducing the precision of risk estimates. In addition, the study did not include follow-up data to assess long-term outcomes or adherence to surveillance guidelines. Finally, cultural and socioeconomic factors specific to Kurdistan/Iraq, such as healthcare accessibility and patient compliance, may further restrict the applicability of findings to other populations. Despite these limitations, this study highlights critical trends in colorectal pathology and underscores the need for expanded screening programs in Iraq.

5. CONCLUSION

This research demonstrated that the most common indications for colonoscopy include lower gastrointestinal bleeding, altered bowel habits, and chronic abdominal pain, with a lower frequency of screening colonoscopies in Kurdistan/Iraq, compared to developed countries. The age of patients undergoing colonoscopy in Kurdistan/Iraq is younger than that in developed nations, suggesting potential underutilization of screening in the region. The observed female predominance may be attributed to socioeconomic or cultural factors. The overall diagnostic yield was 62.43%, with the highest yields observed for rectal mass and anemia. The quality of bowel preparation in Iraq was found to be superior to international standards, although the ADR was lower than recommended levels. The incidence of CRC is increasing, particularly among adults aged 40–49 years, with younger patients and males being disproportionately affected.

REFERENCES

[1] K. Roka, K. C. Indu, S. M. Jha, R. C. Subedi and A. Adhikari. "Pattern of lower gastrointestinal diseases on colonoscopy and histopathological examination in a tertiary care center of Nepal". *Medical Journal of Shree Birendra Hospital*, vol. 21, no. 1, pp. 87-92, 2022.

[2] W. Latos, D. Aebisher, M. Latos, M. Krupka-Olek, K. Dynarowicz, E. Chodurek, G. Cieślar and A. Kawczyk-Krupka. "Colonoscopy:

Preparation and potential complications". *Diagnostics* (*Basel*), vol. 12, no. 3, pp. 747, 2022.

[3] P. Sharma, P. Sapkota, R. B. Gurung, P. Silwal, B. Yadav, N. Gupta, S. Pathak, S. Joshi and Y. Singh. "Demographic and clinical profile of patients undergoing colonoscopy at a tertiary care center in Nepal: A retrospective cross-sectional study". *Annals of Medicine and Surgery Journal* (*Lond*), vol. 86, no. 5, pp. 2633-2638, 2024.

[4] M. A. Bujang, N. Sa'at, T. M. Sidik and L. C. Joo. "Sample size guidelines for logistic regression from observational studies with large population: Emphasis on the accuracy between statistics and parameters based on real life clinical data". *Malaysian Journal of Medical Sciences*, vol. 25, no. 4, pp. 122-130, 2018.

[5] B. A. Abdulhassan, Q. A. Mahdi, M. T. Meeshal, M. H. Ali, S. M. A. Kadhim and A. Noori. "Descriptive study of patients referred for colonoscopy at gastro-enterology unit at Al-Imamain Al-Kadhemain medical city in Baghdad". *Journal of the Faculty of Medicine Baghdad*, vol. 66, no. 1, pp. 11-17, 2024.

[6] Z. K. Dhahi and L. "Almansori. Insight on colonoscopy findings in Southern Iraq : Retrospective study from a tertiary center". *Iraqi National Journal of Medicine*, vol. 6, no. 2, pp. 86-90, 2024.

[7] M. S. F. Khudhur and A. Al Dabbagh. "Colonoscopic findings in patients with bleeding per-rectum in colonoscopy center at rizgary teaching Hospital, Erbil, Iraq". *Anbar Medical Journal*, vol. 19, no. 1, pp. 42-47, 2023.

[8] S. Chaudhary, P. Chaudhary, N. Jaiswal and R. Chaurasia. "Colonoscopy: A two year experience from Western Nepal". *Journal of Universal College of Medical Sciences*, vol. 1, no3, pp. 28-32, 2013.

[9] S. Thomas-Gibson, C. Thapar, S. G. Shah and B. P. Saunders. "Colonoscopy at a combined district general hospital and specialist endoscopy unit: Lessons from 505 consecutive examinations". *Journal of the Royal Society of Medicine*, vol. 95, no. 4, pp. 194-197, 2002.

[10] S. Ashtari. "Overview of diagnostic and treatment colonoscopy function in gastrointestinal diseases". *Journal Liver Research Disorders Therapy*, vol. 2, no. 4, pp. 113-117, 2016.

[11] E. S. Boroff, M. Disbrow, M. D. Crowell and F. C. Ramirez. "Adenoma and polyp detection rates in colonoscopy according to indication". *Gastroenterology Research and Practice*, vol. 2017, pp. 7207595, 2017.

[12] K. Fatakhova and R. Rajapakse. "From random to precise: Updated colon cancer screening and surveillance for inflammatory bowel disease". *Translational Gastroenterology and Hepatology*, vol. 9, p. 27, 2024.

[13] A. Shaukat, C. J. Kahi, C. A. Burke, L. Rabeneck, B. G. Sauer and D. K. Rex. "ACG clinical guidelines: Colorectal cancer screening 2021". *The American Journal of Gastroenterology*, vol. 116, no. 3, pp. 458-479, 2021.

[14] D. K. Rex, P. S. Schoenfeld, J. Cohen, I. M. Pike, D. G. Adler, M. B. Fennerty, J. G. Lieb 2nd, W. G. Park, M. K. Rizk, M. S. Sawhney, N. J. Shaheen, S. Wani and D. S. Weinberg. "Quality indicators for colonoscopy". *Gastrointestinal Endoscopy*, vol. 81, no. 1, pp. 31-53, 2015.

[15] B. E. Lacy, M. Pimentel, D. M. Brenner, W. D. Chey, L. A. Keefer, M. D. Long and B. Moshiree. "ACG clinical guideline: Management of irritable bowel syndrome". *The American Journal of Gastroenterology*", vol. 116, no. 1, pp. 17-44, 2021.

[16] C. I. Pardamean, D. Sudigyo, A. Budiarto, B. Mahesworo, A. A. Hidayat, J. W. Baurley and B. Pardamean. "Changing colorectal cancer trends in Asians: Epidemiology and risk factors". *Oncology Reviews*, vol. 17, p. 10576, 2023.

[17] Iraqi Cancer Board. "*Iraqi Cancer Regist 2011*". Iraq, Iraqi Cancer Board, 2019, pp. 1-10.

[18] A. Kumar, S. B. Dhungana, R. K. Gupta, S. P. Sah and B. Khanal. "Clinico-pathological characteristics of obstructing colorectal cancer and its management outcomes at a tertiary referral center of Eastern Nepal". *BMC Gastroenterol*, vol. 22, no. 1, pp. 1-6, 2022.

[19] S. Ibrahem, H. Ahmed and S. Zangana. "Trends in colorectal cancer in Iraq over two decades: Incidence, mortality, topography and morphology". *Annals of Saudi Medicine*, vol. 42, no. 4, pp. 252-261.

[20] N. K. Dhahir and A. A. Noaman. "A comparative study of colorectal cancer based on patient's age". *Journal of the Faculty of Medicine Baghdad*, vol. 63, no. 2, pp. 70-73, 2021.

[21] M. C. S. Wong, J. Huang, J. L. W. Huang, T. W. Y. Pang, P. Choi, J. Wang, J. I. Chiang and J. Y. Jiang. "Global prevalence of colorectal neoplasia: A systematic review and meta-analysis". *Clinical Gastroenterology and Hepatology*, vol. 18, no. 3, pp. 553-561.e10, 2020.

[22] E. J. Kuipers, W. M. Grady, D. Lieberman, T. Seufferlein, J. J. Sung, P. G. Boelens, C. J. Van de Velde and T. Watanabe. "Colorectal cancer". *Nature Reviews Disease Primers*, vol. 1, p. 15065, 2015.

[23] D. A. Joseph, R. G. Meester, A. G. Zauber, D. L. Manninen, L. Winges, F. B. Dong, B. Peaker and M. Van Ballegooijen. "Colorectal cancer screening: Estimated future colonoscopy need and current volume and capacity". *Cancer*, vol. 122, no. 16. pp. 2479-2486, 2016.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Statistical Analysis of Realized Volatility of Bitcoin Price using Heterogeneous Autoregressive and Generalized Autoregressive Conditional Heteroskedasticity Models

Aras Jalal Mhamad Karim

Department of Statistic and Informatics - College of Administration & Economics-University of Sulaimani, Sulaimani, Kurdistan Region – Iraq

ABSTRACT

Bitcoin has recently gained extra attention in the financial industry and the blockchain community in general; it's considered the most popular form of technology. As a result, the purpose of the study is to predict the actual volatility of the bitcoin price using generalized autoregressive conditional heteroskedasticity (GARCH) and heterogeneous autoregressive (HAR). In this research, the researcher attempted to utilize the appropriate statistical methodology, such as GARCH and HAR models. GARCH models were created to address the issue of volatility aggregation, which is the tendency for prices to cluster together as large changes occur. With the GARCH model, we can represent the conditional heteroskedasticity and the fat tail of financial market data. The primary objective was to directly observe and predict the behavior of volatility in time series data. Overall, the model's architecture appears simple and is capable of reproducing the primary characteristics of financial information. The primary concept of this model is that investors with different time frames perceive and respond to different levels of volatility. Sample information about the price of the bitcoin cryptocurrency was distributed worldwide. It includes daily updates of the variable for the time period 31-Jun-17 to 31-Jan-22. The investigation has demonstrated that the HAR model is more effective at predicting variance for this period in comparison to GARCH (1, 1). The result shows that 1 **day** of previous year's variance estimates and jump estimates have a significant impact on the future variance (h = 1).

Index Terms: GARCH Model, HAR Model, Forecasting Volatility, Realized Variance, Time Series Analysis

1. INTRODUCTION

In recent years, bitcoin has been established as the world's leading cryptocurrency, taking the attention of consumers, businesses, and investors. On the other hand, the topic of forecasting volatility has attracted a lot of attention in the past decade from many academics and also financial professionals. It has been a subject of great discussion over the years and a lot of research has already been done. One of the most common approaches of modeling volatility indirectly is using ARCH or generalized autoregressive conditional heteroskedasticity (GARCH) models, but nowadays, with the realized measures, it has became possible to directly model volatility. One of the models that directly use the realized measures to forecast volatility is the heterogeneous autoregressive (HAR) model. The major idea of this model is that investors with different time horizons perceive and react to different types of volatility. It is a model that has a simple structure; it is easy to estimate and is able to replicate the main

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp136-147

E-ISSN: 2521-4217
P-ISSN: 2521-4209

Copyright © 2025 Karim AJM. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Aras Jalal Mhamad Karim, Department of Statistic and Informatics - College of Administration & Economics-University of Sulaimani, Sulaimani, Kurdistan Region – Iraq. E-mail aras.mhamad@univsul.edu.iq

Received: 14-04-2025 Accepted: 15-06-2025 Published: 18-09-2025

features of financial data, such as in Corsi's study in 2003 [1]. The HAR model is basically an additive cascade of realized volatilities, generated at different time horizons, that follows an autoregressive process. There are a lot of studies in the field of cryptocurrency that has been applied. According to Nathan Reiff's article [2], one of the first attempts to create a cryptocurrency comes from the Netherlands in the late 1980s. At the same time, an American cryptographer, David Chaum, introduced a different electronic cash, digiCash [3]. He developed a "blinding formula" to encrypt information to be passed among people. And then some companies applied these fundamentals in the 1990s. In 1998, Wei Dai developed an "anonymous, distributed electronic cash system", called b-Money [4]. This system was based on a digital pseudonym used to transfer currency through a decentralized network. The Bit Gold proposed by Szabo [5] introduced a proof-of-work system, which is used in some ways in bitcoin's mining network, many academic papers have focused on bitcoin. For Corsi's study in 2003 [1], bitcoin has a position in financial markets and project management between gold and the American dollar with clear advantages to risk averse investors. Zhu *et al.* [6] suggested that some factors, such as Consumer Price Index, Dow Jones Industry Average, and Fed Funds Rate, do have a long-run negative effect on bitcoin price, by applying the vector error correction model. Due to the importance of forecasting, many specialists have developed a variety of time series forecasting models. The framework of historical volatility suggested by Bollerslev [7] on how Realized Variance significantly outperforms GARCH-type models, and the long memory extension presented by Corsi [8] would indicate that this is a highly efficient model. Chung *et al.* [9] applied a HAR model to index options and compared it with IV, which proved significantly higher values of fit. Sea [10] used the HAR-RV model to test its performance on his data. He tested the HAR model against the simple autoregressive (AR) and GARCH (1,1) models. He concluded that the HAR model showed excellent in-sample forecasting performance against another models. According to Vortelinos [11], the HAR model produced the best accurate. The forecast against principal components will combining with neural networks, and GARCH models. From the above studies, the GARCH and HAR models have been widely utilized to enhance the accuracy of the prediction model, especially in the financial field. Therefore, using these models to analyze and forecast of bitcoin cryptocurrency price are suitable. Besides, a modern time series model exists and is applied to analyze bitcoin cryptocurrency, but the specifications and analyzing of bitcoin is always mysterious. For this reason, the present study deals with this problem by forecasting and analyzing the bitcoin cryptocurrency. Hence, the researcher tries to set up a theoretical model to analyze bitcoin cryptocurrency with the recent record of its data. Hence, the study contribution is building a volatility time series model to estimate and accurate forecast using these models, which are GARCH and HAR. Therefore, the objective of the study is to forecast the realized volatility of the bitcoin cryptocurrency price using GARCH and HAR.

Hence, the next section provides a brief overview of the framework, applying the GARCH and HAR models in the determination of the relationship between variables. In section 3, present the data and derive the time series models utilized in the analysis from the theoretical framework. The conclusions and further discussion of the study results are examined in section 4.

2. THEORETICAL FRAMEWORK

2.1. Bitcoin

Bitcoin is a peer-to-peer (P2P) electronic cash system introduced in the well-known paper of Nakamoto [12]. The P2P mechanism allows an ownership transfer from one party to another without a third-party intervention (financial institution). Payments can be made over the internet without any control or cost of a central authority for the 1st time. Individuals who want to own bitcoins can either run a program on their own computer that implements the bitcoin protocol or create an account on a website that runs bitcoin for its users. The bitcoins are saved in a file called a wallet, which the user may secure and backup. These programs connect to each other over the internet forming P2P networks, making the system resistant to a central attack by M. Crosby et al. [13].

For now, bitcoins are generated through a process of mining. Any member operates as a miner using their computer knowledge to maintain the network. Mining is a computationally process that requires miners to find a solution to a mathematical problem to create a new block into the blockchain. Miners resolve this issue using the proof-of-work concept. This algorithm involves recurrently difficult mathematical problems until getting to a solution. The first miner to find a solution broadcast it to the network to verify it. Once verified, the block is added to the blockchain. Every 10 min on average is found a new answer and a bitcoin is created. The bitcoin protocol is designed to generate a new bitcoin gradually. The difficulty of solving problems is adjusted every 2 weeks at the rate of six blocks per hour.

The size of the reward was initially 50 (genesis block) and it is halved every 4 years, this implies that the number of bitcoins in circulation will never exceed 21 million. Once the last bitcoin is generated, miners will instead be rewarded with transaction fees Lo [13].

2.2. GARCH Model

The GARCH model was first introduced by Bollerslev [7]. Back in these days, the concept of realized volatility modeling was not even introduced. At that period, the daily volatility was calculated as the squared daily return without taking into consideration any subintervals. The GARCH model is a conditional volatility model that allows the conditional variance to depend on the previous lags. It is based on the ARCH model by Engle [14], who used it to show that the conditional volatility is affected by volatility clustering. An autoregressive conditionally heteroskedastic (ARCH) model is a time series model with econometric applications that consider the variance of the current error term as a function of the variance of the error conditions of the previous time periods. One of the disadvantages of the ARCH model is that it responds slowly to large, unusual shocks. Thus, the need of an improvement of this model was crucial. Assuming an autoregressive moving average (ARMA) model for the error variance, then the model is a GARCH model. GARCH models were designed to deal with the problem of volatility clustering, which is the phenomenon where large changes in prices tend to cluster together, as A. J. M. Karim and N. M. Ahmed [15]; Botan et al. (2020) [16]; and R. F. Engle [14].

Before describing the GARCH model, the ARCH specification has to be introduced. The following return process has to be specified:

$$r_t = \mu_t + \varepsilon_t \quad with \quad \varepsilon_t = z_t\sqrt{\sigma_t} \tag{1}$$

Where, μ_t is a drift term that is explained by the structural model and z_t is an independent shock with zero mean and unit variance, signifying that ε_t is normally distributed $\varepsilon_t \sim Z(0, \sigma_t)$. The conditional variance in (1) can be transformed into a time-varying by specifying the ARCH (q) process:

$$\sigma_t = c + \sum_{i=1}^{q} a_i \varepsilon_{t-1}^2 \tag{2}$$

Where c is a constant and a_i is the coefficient for the past squared shocks (ε_t^2). Then the GARCH (p,q) model is derived by adding p lagged conditional variances, with orders p ≥ 1 and q ≥ 1:

$$\sigma_t = c + \sum_{i=1}^{q} a_i \varepsilon_{t-1}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j} \tag{3}$$

Where β_j is the coefficients for the past conditional variances, p is the past squared error terms, and q is the past estimated volatility terms. When q = 0, then the above equation (3) reduces to an autoregressive conditional heteroskedastic (ARCH) model. Given a distribution of ε_t in equation (1) and setting p = q =1, then the GARCH (1, 1) is derived:

$$\sigma_t = c + a_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1} \tag{4}$$

For which the condition c ≥ 0, $a_1 \geq 0$, and $\beta_1 \geq 0$ should stand for every positive value of σ_t. Since the GARCH model is non-linear, it cannot be estimated by an OLS regression like the HAR model. Thus, the Gaussian maximum likelihood (GMLE) method should be used for parameter estimation. When assuming normally distributed errors and starting from some parameter vector θ and a time series of size T ($y_1, y_2 \ldots y_T$), the GMLE method calculates the probability density for this specific sample by taking the product over all the marginal conditional probability densities of the observed data. In general, the GARCH model is using the returns to forecast volatility, and it depicts that today's return consists of yesterday's return plus some volatility part and this volatility is what we need. This model is also using a rolling regression method to forecast volatility, by moving one day ahead and leaving 1 day behind for every forecast, which means that the data window size remains stable.

2.3. Testing GARCH Effects (Test of Heteroscedasticity)

The availability of ARCH/GARCH effects may give serious model misspecification if they are ignored. Logically, ignoring ARCH effects will give the identification of ARMA models that are over-parameterized. In addition, as in heteroscedasticity, estimation assuming its absence will result in inappropriate standard errors of parameter estimates, which are typically smaller than what they should be. Therefore, it is important to check the presence of GARCH effects in time series modeling according to McLeod and Li [17]; Asraa *et al.* [18]; and Azhy *et al.* [19].

Two ways of testing GARCH effects are used. First is to check the Ljung-Box portmanteau Q statistics of a_t^2.

McLeod and Li show that the sample autocorrelations of a_t^2 have asymptotic variance n^{-1} and that portmanteau statistics calculated from their distribution is asymptotically Chi-square if a_t^2 are independent. Since the sample

autocorrelations of a are also pertinent to the identification of a GARCH model for a_t^2.

The second is the process of checking for conditional heteroscedasticity which is to utilize the Lagrange multiplier test of Engle. Think about the following model of regression for

a_t^2 on a_{t-j}^2 j=1,2,…,m

$$a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + ... + \alpha_m a_{t-m}^2 + v_t \text{, t=m+1…n} \quad (5)$$

Where y_t is the error term, m is a pre-determined positive integer, and n is the total number of data points in the series. Using the coefficient of determination from (5), Engle demonstrates that, under the null hypothesis H_0: $\alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$, the variance of nR^2 is approximately distributed according to a chi-square distribution with m degrees of freedom by Weiss [20]; Nakamoto [12].

2.4. Identification of a GARCH Model

If the Ljung-Box statistics and the LaGrange multiplier (LM) test are significant, then conditional heteroscedasticity of a_t^2 is present, and we need to identify an appropriate GARCH model for a_t^2. However, since the GARCH (1,1) model has been shown to be appropriate in many empirical studies, we may employ the GARCH (1,1) model at the beginning of the analysis. As the model is estimated, diagnostic checking procedures may be followed to see if the GARCH (1,1) model is okay, or if the orders of the GARCH model should be increased or decreased. Instead of using this trial-and-error approach, we may use the following procedure for the definition of a GARCH model for the $\{a_t^2\}$ series by Lon-Mu [21]; N. M. Ahmed and A. J. M. Karim [22], Dyhrberg [23].

2.5. Ljung-Box Q-Statistic

Adding to the visual inspection of the plotted autocorrelation, the Ljung-Box Q-Statistic is used for diagnostic checking by Box and Jenkins [24]. The Ljung-Box Q-Statistic is defined by equation (6)

$$Q^* = n(n+2)\sum_{j=1}^{K}(n-j)^{-1} r_j^2(a_t) \quad (6)$$

Where n is the number of observations, K is the largest degree of freedom used, and r_j is the sample association function at the jth degree of freedom of a relevant time series a_t, for example. The statistical r_j for is then calculated as:

$$r_j = \frac{\sum_{t=j+1}^{n}(a_t - \bar{a})(a_{t-j} - \bar{a})}{\sum_{t=1}^{n}(a_t - \bar{a})^2} \quad (7)$$

The Q-statistic was suggested for testing ARIMA and ARMA models; both the test statistics are determined by the calculation of the sample autocorrelation function for the residuals $\hat{\varepsilon_t}$ from those models. The similar test statistic based on different calculations using the autocorrelation function will be high benefit for small sample applicability, it is defined as Weiss [20]; M. S. Lo. [25].

$$Q^* = n(n+2)\sum_{j=1}^{K}\frac{r_j^2(e_t)}{(n-j)} \sim \chi_k^2 \quad (8)$$

where k is the number of lags considered in the test, and k is defined by:

$k = K - m$,

where K: Number of lags used in the test.

m: Number of parameters estimated in the mean and variance equations of the GARCH model, and r_j^* is:

$$r_j^* = \sum_{t=j+1}^{n}(\hat{a_t^2} - \bar{a})(\hat{a_{t-j}^2} - \bar{a}) / \sum_{t=1}^{n}(\hat{a_t^2} - \bar{a}) \quad (9)$$

2.6. Likelihood Function of GARCH Models

By defining $\alpha = [\alpha_0, \alpha_1, \ldots, \alpha_m, B_1, \ldots, B_r, \eta]$', the log likelihood functions of α may be derived under the Normality assumption of ε_t. If ε_t is assumed to follow a normal distribution. However, practically, there is substantial evidence showing that this assumption may not all the time be satisfactory by Lon-Mu [21].

For the GARCH (1.1) model, the joint density of the observations $a_1 \ldots a_T$ can be calculated as the product of the conditional densities, conditioning on the last observations from M. S. Lo. [25].

$$fa_1,\ldots,a_T\,(a_1,\ldots,a_1)$$

$$= \{\prod_{i=2}^{T} fa_i \,|a_1,\ldots a_{i-1}(a_i\,|a_1,\ldots a_{i-1})\} * fa_1(a_1) \quad (10)$$

Easy to say, the marginal population will decrease as for the ARIMA (1.1) model. For k = 2.,T, the probability of a_k, given the values $a_1 \ldots a_{(k-1)}$, is

$$fa_k|a_1,....,a_{k-1}(a_k|a_1,...,a_{k-1}) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_k^2}}\exp\left\{-\frac{a_k^2}{2\sigma_k^2}\right\} \tag{11}$$

Moreover, the conditional likelihood function given a_t and σ_t^2 is:

$$L(\alpha_0,\alpha_1,B_1) = fa_2,...a_T|a_1\sigma_1^2(a_2,...a_T|a_1\sigma_1^2)$$
$$= \prod_{i=1}^{T}\frac{1}{\sqrt{2\pi}\sqrt{\sigma_i^{*2}}}exp\left\{-\frac{a_i^2}{2\sigma_i^{*2}}\right\} \tag{12}$$

Where $\sigma_i^{*2} = \alpha_0 + \alpha_1 a_{i-1}^2 + B_1$ are obtained recursively. We substitute σ_t^2 by its expected value:

$$E(\sigma_t^2) = \frac{\alpha_0}{1-\alpha_1-B_1} \tag{13}$$

Using the logarithm and ignoring the constant term, we find that the log likelihood function is:

$$l(\alpha_0,\alpha_1,B_1|a,\sigma^2) = -\frac{1}{2}\left\{\sum_{i=2}^{T}\log\sigma_i^{*2} + \frac{a_i^2}{\sigma_i^{*2}}\right\} \tag{14}$$

Where $a = (a_1,....a_T)^-$ and $\sigma_t^2 = (\sigma_1^2,...\sigma_T^2)$.

2.7. Model Checking of GARCH (r,m)

For a GARCH model, the standardized errors $\varepsilon_t^\wedge = a_t^\wedge / \sigma_t^\wedge$ are independent and identically distributed random mistakes that are associated with either a standard normal or a non-normal distribution, such as the standardized student-t distribution. As a result, one can assess the effectiveness of a fitted GARCH model by inspecting the series $\{\varepsilon_t^\wedge\}$. Specifically, the sample autocorrelations and the Ljung-Box Q statistics of can be utilized to assess the effectiveness of the mean (primordial) equation and the sample autocorrelations and the Ljung-Box Q statistics of can be utilized to assess the validity of the volatility (secondary) equation A.J.M. Karim, and N.M. Ahmed [15]; Lon-Mu [21].

2.8. Forecasting the GARCH (1,1) Model

Forecasts of a GARCH model can be found using methods similar to those of an ARMA model. Consider the GARCH (1, 1) model in assume that the forecast origin is n. For a one-step-ahead forecast, we have:

$$\sigma_{n+1}^2 = \alpha_0 + \alpha_1 a_n^2 + B_1\sigma_n^2 \tag{15}$$

Where a_n^2 and σ_n^2 are known at t = n, therefore, the one-step ahead forecast is:

$$\sigma_n^2(1) = \alpha_0 + \alpha_1 a_n^2 + B_1\sigma_n^2 \tag{16}$$

For multi-step ahead forecasts, we use $a_t^2 = \sigma_t^2\varepsilon_t^2$:

$$\sigma_{t+1}^2 = \alpha_0 + (\alpha_1+B_1)\sigma_t^2 + \alpha_1\sigma_t^2(\varepsilon_t^2-1) \tag{17}$$

When t = n+1, the equation becomes:

$$\sigma_{n+2}^2 = \alpha_0 + (\alpha_1+B_1)\sigma_{n+1}^2 + \alpha_1\sigma_{n+1}^2(\varepsilon_{n+1}^2-1) \tag{18}$$

Since $E(\varepsilon\ (n+1)-1\ |F_n) = 0$, the two-step-ahead volatility forecast at the forecast origin n satisfies the equation:

$$\sigma_n^2(2) = \alpha_0 + (\alpha_1+B_1)\sigma_n^2(1) \tag{19}$$

In general, we have:

$$\sigma_n^2(\ell) = \alpha_0 + (\alpha_1+B_1)\sigma_n^2(\ell-1)\ ,\quad \ell>1 \tag{20}$$

This outcome is identical to the result of an ARMA (1, 1) model with an AR polynomial of degree 1 − (α_1+B_1) B. By repeatedly changing the values in (20), the forward forecast can be written as:

$$\sigma_n^2(\ell) = \frac{\alpha_0[1-(\alpha_1+B_1)^{\ell-1}]}{1-\alpha_1-B_1} + (\alpha_1+B_1)^{\ell-1}\sigma_n^2(1) \tag{21}$$

Therefore:

$$\sigma_n^2(\ell) \to \frac{\alpha_0}{1-\alpha_1-B_1} = \alpha_0^*\ ,\quad as\ \ \ell\to\infty \tag{22}$$

Provided that $\alpha_1 + B_1 < 1$

As a result, the multi-step-ahead predictions of volatility made by a GARCH (1,1) model match the unconditional variance of as the horizon for predictions increases to nothing if Var (a_t) is present Lon-Mu [21]; Cont [26]; and Nader *et al.* [27].

2.9. The HAR-RV Model

The idea of realized variance is based on these assumptions and Andersen *et al.* [28] provide us with an explanation in further detail on how efficient an estimator of volatility the realized volatility is, and moreover, how it outperforms

traditional GARCH-type models. Equation (23) presents the model for estimating a daily Realized Variance (RV), where r is high-frequency intraday log-returns as described in equation (24).

$$RV_t^{(d)} = \sqrt{\sum_{j=0}^{M-1} r_{t-j}^2} \tag{23}$$

where r is Log returns (or continuously compounded returns) are approximately equal to normal price returns, but hold significant benefits in simplicity in multi-period returns by Ruppert and Matteson [29]. Log returns are defined:

$$r = log\left(\frac{P_t}{P_{t-1}}\right), \qquad t = 1,\ldots,T \tag{24}$$

Based on the Heterogeneous Market Hypothesis (HMH), Corsi [8] proposes the HARRV as a model that will utilize three realized volatility components in an autoregressive manner, which all represent some time-dependent market component for the model. The following equations 25 and 26 consider the RV over the complementing horizons. They are quite simply the average of the daily RV, so for a weekly RV, we simply extend the model as following:

$$RV_t^{(w)} = \frac{1}{5}\left(RV_t^{(d)} + RV_{t-1d}^{(d)} + \ldots + RV_{t-4d}^{(d)}\right) \tag{25}$$

Moreover, the same definition for monthly volatility, but over 22 daily periods:

$$RV_t^{(m)} = \frac{1}{22}\left(RV_t^{(d)} + RV_{t-1d}^{(d)} + \ldots + RV_{t-21d}^{(d)}\right) \tag{26}$$

The added sum of these three volatilities can be regarded as an additive cascade of volatilities, each representing different components of market volatility. From this, it gets an almost long memory AR type of character (with lags one, five, and 22), but not strictly Corsi [8].

By expanding the expected values and utilizing straightforward recursive substitution, the volatility model will be given by a three-step cascade and has a form of something similar to three AR processes Corsi [8]):

$$\sigma_{t+1}^{(d)} = c + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \omega_{t+1d}^{(d)} \tag{27}$$

Now given (27), all variables are directly observable and available in the data set. The parameters will be able to be estimated through a simple Ordinary Least Squares estimation (OLS) using the Newey-West covariance matrix estimator. However, due to possible serial correlation, a Newey-West (NW) covariance correction will be applied, since the effects of covariance and autocorrelation must be considered in the estimation.

3. DATA ANALYSIS AND RESULTS

3.1. Data Description

In this paper, daily observations are used of the bitcoin price, the sample period is June 31, 2017–January 31, 2022, which obtained from the Kaggle website [30], the researcher use R-language to obtain results. Fig. 1 below shows the time series plot of the series during the sample period. Since 2017, the bitcoin price has become more volatile. On October 13th, 2017, bitcoin price breaks the $5,000 for the 1st time, on November 28th, 2017, the $10,000, and on December 18th, 2017, hits all-time high just below $20,000. The bitcoin price time series can be observed in Fig. 1, non-linear trend and non-stationarity are the first geometrical properties that are shown in Fig. 1.

To build an appropriate model, the series that are used in analysis must be stationary; therefore, it should check the unit-root structure of the data. Although the above graph gives a rough idea about the stationarity structure of the series, we have applied the Augmented Dickey-Fuller test to the series to test unit roots. Table 1 exhibits the results from the ADF test applied to levels, first differences of the series.

The ADF test results indicate that the variable is non-stationary by not rejecting the null hypothesis of unit-root at the level, but it is stationary after first differencing. Therefore, the researcher uses differenced series in its analysis. Fig. 2 below presents a time series plot of the differenced series.

Fig. 1. Time series plots of the variable.

TABLE I: Unit root results of the lag variable

Variable	Lags	Test value	*P*-value
Bitcoin Price	Level	−0.9695	0.7659
	First	−43.9029	0.0001

After achieving the stationarity condition of the series, we should fit the mean equation model as it is shown below:

From Table 2, *P*-value of the constant and lagged variable are less than the statistical significance (0.05), then one can say that the model is significant.

Fig. 3 shows that there is a prolonged period of high volatility from day 1 to the end of 2017, and also, there exists a prolonged period of low volatility from the end of 2018, to the beginning of 2019. This suggests that the residual or error term is conditionally heteroskedastic and it can be represented by the ARCH and GARCH models.

In the series, hypothesis of ARCH has an effect or not on the mean equation. For this purpose, a heteroskedasticity test ARCH have been used to test the following hypothesis below, and its results are shown in Table 3:

- H_0: There is no ARCH effect.
- H_1: There is an ARCH effect.

Table 3 presents *P*-value of the heteroskedasticity ARCH test, which is less than the statistical level (0.05), then the hypothesis can be rejected; in another word, there exists an ARCH effect in the series. The researcher achieved two main assumptions of using the GARCH model, which are the stationary of the series and the effect of ARCH in the mean equation model, and then GARCH model can be used to forecast the volatility. The GARCH (1,1) model has been run, the results of its fit are shown in Table 4 below:

From the above table, it is obvious that the estimators of the variance equation are significant, depending on *P*-value, which is less than 0.05. The residuals of the GARCH (1,1) model should be tested to find out that the model is suffered from serial correlation of residuals or not for the hypothesis below:

- H_0: There is no serial correlation of residuals.
- H_1: There is serial correlation of residuals.

From Table 5, *P*-value of the Q-statistic test for the 36 laggs are greater than the statistical level (0.05), then we cannot reject the hypothesis, that is mean there is no serial correlation of residuals.

TABLE 2: The fit of the mean equation model

Variable	Coefficient	Standard error	t-statistic	*P*
C	5.59E-07	3.09E-07	1.811371	0.0402
Original (-1)	0.993359	0.001206	823.5537	0.000

TABLE 3: The heteroskedasticity test ARCH

Test value	*P*-value
214.9139	0.000

TABLE 4: The fit of the GARCH (1,1) model

Variable	Coefficient	Standard error	z-statistic	Prob.
C	3.88E-07	1.86E-07	2.082135	0.0373
Original (-1)	0.99336	0.000632	1570.896	0.000
Variance equation				
C	6.35E-12	2.59E-13	24.52996	0.000
RESID (-1)^2	0.15	0.007405	20.2578	0.000
GARCH (-1)	0.6	0.012556	47.78632	0.000

R-squared: 0.9973

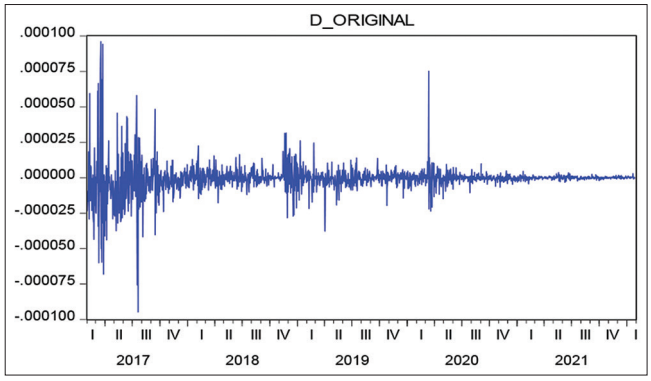


Fig. 2. Time series plot of the differenced variable.

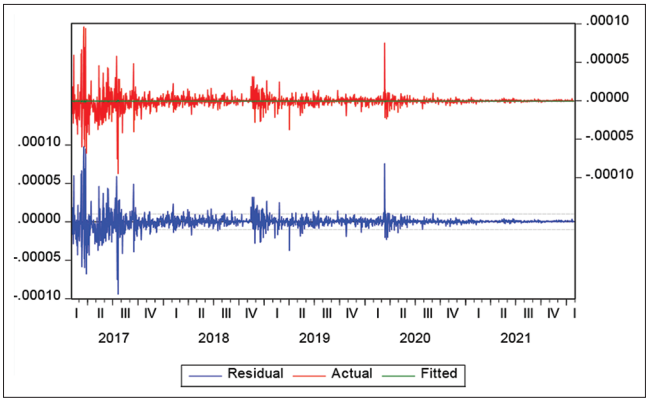


Fig. 3. The residuals of the mean equation.

Another test is the heteroskedasticity test to figure out that the postulated model is adequate or not for the following

TABLE 5: Testing of ACF and PACF

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	0.006	0.006	0.0659	0.797
		2	0.006	0.006	0.1309	0.937
		3	0.002	0.002	0.1417	0.986
		4	0.032	0.032	2.0313	0.730
		5	0.028	0.028	3.4952	0.624
		6	0.010	0.009	3.6653	0.722
		7	0.030	0.030	5.3633	0.616
		8	0.010	0.009	5.5533	0.697
		9	0.011	0.008	5.7559	0.764
		10	0.023	0.021	6.7321	0.750
		11	0.031	0.029	8.5472	0.664
		12	0.013	0.010	8.8353	0.717
		13	0.008	0.006	8.9489	0.777
		14	0.018	0.015	9.5403	0.795
		15	0.030	0.026	11.215	0.737
		16	0.022	0.019	12.139	0.734
		17	0.015	0.011	12.528	0.767
		18	0.011	0.008	12.770	0.805
		19	0.009	0.005	12.920	0.843
		20	0.009	0.005	13.085	0.874
		21	0.018	0.013	13.694	0.882
		22	0.007	0.002	13.788	0.909
		23	0.012	0.008	14.050	0.925
		24	0.014	0.010	14.414	0.937
		25	0.020	0.015	15.168	0.937
		26	0.022	0.017	16.076	0.934
		27	0.014	0.010	16.435	0.944
		28	0.028	0.023	17.839	0.930
		29	0.034	0.030	19.967	0.894
		30	0.006	0.001	20.040	0.915
		31	0.017	0.012	20.597	0.922
		32	0.003	-0.003	20.613	0.940
		33	0.011	0.004	20.823	0.951
		34	0.011	0.005	21.036	0.960
		35	0.035	0.029	23.286	0.935
		36	0.018	0.012	23.924	0.938

TABLE 6: The test of heteroskedasticity for the ARCH effect

Test-value	*P*-value
0.065713	0.7977

hypothesis below, and the results are shown in Table 6 below:

- H_0: ARCH has no effect
- H_1: ARCH has effect.

Table 6 shows the test for heteroskedasticity and it is clear that *P*-value of the test is greater than the statistical level (0.05) then cannot be reject the hypothesis, which means that there is no heteroskedasticity.

The final test is a normality test to figure out that the residual of the postulated model is normally distributed or not form the following hypothesis below, and the results are shown in Table 7 below:

Fig. 4. (a) Observed and forecasted realized volatility (RV), (b) forecasted RV only.

TABLE 7: A normality test for the residual of the model

Mean	Median	Max	Min	SD	Skew-ness	Kurt-osis	Jarque Bera	*P*-value
0.030	−0.017	16.36	−8.77	1.13	1.80	32.33	66410.98	0.000

TABLE 8: The forecasted values

Parameters	Estimate	Standard error	t value	Pr(>|t|)
beta0	0.00000051	0.00000039	1.298	0.195
beta1	0.966	0.031	31.015	0.000
beta2	0.042	0.039	1.083	0.279
beta3	0.014	0.019	0.707	0.479

Multiple R-squared: 0.9968
F-statistic: 186600 with *P*-value (0.000)

- H_0: The residual is normal.
- H_1: The residual is not normal.

Table 7 shows the test for residual normality and it is clear that *P*-value of the test is less than the statistical level (0.05) then one can be reject the hypothesis, which means that the residual are not normal and this is a good result because residual not normal, which indicate to good model. Moving now to the HAR-RV model, first, the heteroskedasticity test has to be done for the residuals. Using heteroskedasticity test, as shown in Table 3, that there is indeed heteroskedasticity was found. After conducting this test for the HAR-RV model, the realized volatility of yesterday (RV1) was significant at a level of 5%, the realized volatility of past week (RV5) was no again significant, and the realized volatility of past month (RV22) was no again significant at a level of 5%. All the coefficients were positive and the F-test for the model is very large, which is significant. The coefficient for RV1 is more than RV5 and RV22. Table 8 below provides an overview of the coefficients of the estimation.

TABLE 9: Comparison of postulated models

Model	AIC	BIC	MSE	Residual standard error	Log likelihood
GARCH	−21.1672	−21.15	1.011E-10	1.00E-05	19320.1
HAR*	−36475.5	−36448	6.1419E-13	9.829E-06	18242.7

TABLE 10: he forecasted values

Date	Forecast	Date	Forecast	Date	Forecast
November 03, 2021	61619.76	December 01, 2021	54641.91	January 01, 2022	44580.88
November 04, 2021	61222.82	December 02, 2021	54747.75	January 02, 2022	45812.86
November 05, 2021	60099.69	December 03, 2021	54107.37	January 03, 2022	45356.18
November 06, 2021	59618.95	December 04, 2021	50953.54	January 04, 2022	44687.40
November 07, 2021	59994.44	December 05, 2021	47245.07	January 05, 2022	44160.97
November 08, 2021	61809.40	December 06, 2021	47332.24	January 06, 2022	41972.35
November 09, 2021	65592.52	December 07, 2021	48573.66	January 07, 2022	41780.99
November 10, 2021	65107.30	December 08, 2021	48597.10	January 08, 2022	40388.64
November 11, 2021	63344.30	December 09, 2021	48412.76	January 09, 2022	40347.31
November 12, 2021	63388.27	December 10, 2021	45801.95	January 10, 2022	40470.99
November 13, 2021	62861.41	December 11, 2021	45644.80	January 11, 2022	40571.47
November 14, 2021	63062.32	December 12, 2021	47544.54	January 12, 2022	41413.94
November 15, 2021	64183.95	December 13, 2021	48167.93	January 13, 2022	42524.40
November 16, 2021	62065.81	December 14, 2021	45213.72	January 14, 2022	41454.08
November 17, 2021	58823.18	December 15, 2021	46687.91	January 15, 2022	41922.97
November 18, 2021	58644.59	December 16, 2021	47112.76	January 16, 2022	41910.83
November 19, 2021	55676.47	December 17, 2021	46145.70	January 17, 2022	41985.74
November 20, 2021	56342.73	December 18, 2021	44862.72	January 18, 2022	41329.44
November 21, 2021	58032.85	December 19, 2021	45324.90	January 19, 2022	41442.00
November 22, 2021	57217.83	December 20, 2021	45247.74	January 20, 2022	40879.03
November 23, 2021	54909.94	December 21, 2021	45455.27	January 21, 2022	39958.20
November 24, 2021	56151.42	December 22, 2021	47271.15	January 22, 2022	35918.61
November 25, 2021	55860.37	December 23, 2021	46954.62	January 23, 2022	34599.21
November 26, 2021	57195.02	December 24, 2021	48763.34	January 24, 2022	35582.92
November 27, 2021	52061.67	December 25, 2021	48714.33	January 25, 2022	35847.75
November 28, 2021	52269.52	December 26, 2021	48353.02	January 26, 2022	36130.89
November 29, 2021	54747.87	December 27, 2021	48699.05	January 27, 2022	36036.67
November 30, 2021	55152.15	December 28, 2021	48565.47	January 28, 2022	36406.57
		December 29, 2021	45845.78	January 29, 2022	36970.56
		December 30, 2021	44856.14	January 30, 2022	37487.20
		December 31, 2021	45372.75	January 31, 2022	37268.50

Table 8 reports the results of the HAR model, where the in-sample forecasting results show consistently that the lagged RV has a strong and persistent positive relationship with future realized variance log (RVt: t+h), especially in the first forecasting horizons h = 1. To detect the best model between GARCH and HAR models, one can use the criteria below such shown in Table 9.

Table 9 represents the comparison between models, according to all criteria; the HAR model is the best model than the GARCH model, which means that can be used HAR model to forecast the price of bitcoin cryptocurrency. In addition, the full-sample forecasting assumes the realized variance time series to be stable, so the researcher implements the rolling window method to allow the parameters to change over time, and then more reasonable comparisons can be obtained. The adaptive method mimics an investor who updates the forecasting model based on the most recent information. The window size T of the adaptive HAR models employed here is 90 days, that is, models are estimated using past 90-day samples, such as shown in Table 10. Moreover, the model is re-estimated every day. After the re-estimation of each day, the out-of-sample forecasts are performed in horizons h = 1; 5; 22, spontaneously. The parameters of the daily aggregated realized variances are evolving systematically, which justifies the adaptive forecasting method.

Table 10 reports about the forecast values. These forecasts are obtained by first estimating the parameters of the models on the full sample and then performing a series of static one-step-ahead forecasts. Fig. 4 reports the results for out-

of-sample forecasts of the realized volatility in which the model is re-estimated daily.

Fig. 4 illustrates that the upward and downward risk estimators serve complementary roles in forecasting over time. Specifically, the upward risk coefficients showed a gradual increase from 2017 until early 2021. After that point, they exhibited a stronger upward trend, indicating growing exposure to potential positive risks or gains in the forecasted variable.

4. CONCLUSION

As an emerging financial asset, bitcoin has been booming in recent years and has become a major alternative asset for many investors worldwide. This paper studies the realized volatility forecast of the bitcoin cryptocurrency price using GARCH and HAR, based on the data of a 5-year sample period from January 2017 to January 2022. The researcher employs the GARCH and HAR models to study the forecasting properties of bitcoin realized volatility. First of all, a full sample forecasting result reveals that the 1-day lagged realized variance estimators and jump estimators impact the future realized variance significantly across the forecasting horizons h = 1. Then, the researcher allows the forecasting model to be adaptive with a 90-day rolling window, and who finds that the signed jumps can be a significant predictor of the future realized variance of the longer horizon. The results show that the HAR model seems to successfully achieve the purpose of modeling the behavior of volatility in a very simple and parsimonious way. Moreover, in spite of the simplicity of its structure and estimation, the HAR model shows remarkably good forecasting performance. Based on the out-of-sample forecasting results for the long series of realized volatilities of bitcoin price, the HAR model steadily outperforms the short-memory models at all the time horizons considered (1 day, 1 week, and 2 weeks) and is comparable to the much more complicated and tedious to estimate GARCH model. The study has been found that the HAR model was able to better forecast volatility for this period against GARCH (1, 1).

A good future research suggestion would be that the extensions of these two models could be tested, such as a log HAR model and the 1,2 or 1,3 order of the GARCH model. The EGARCH and TGARCH are two extensions of the GARCH model that can be also tested for their forecasting performance with the HAR model. While high-frequency data enhances the accuracy of realized volatility estimates, it also introduces microstructure noise. HAR-RV may be more robust to such noise due to its aggregation over multiple time scales, but this can also smooth over meaningful short-term shifts.

REFERENCES

[1] F. Corsi. *A Simple long Memory Model of Realized Volatility*. *Manuscript*. University of Southern Switzerland, Switzerland, 2003.

[2] N. Reiff. *Were there Cryptocurrencies before Bitcoin*; 2019. Available from: https://www.investopedia.com/tech/were-there-cryptocurrencies-bitcoin [Last accessed on 2019 Aug 09].

[3] D. Chaum. Blind signatures for untraceable payments. In: D. Chaum, R. L. Rivest and A. T. Sherman, editors. *Advances in Cryptology: Proceedings of Crypto 82*. Springer, Germany, pp. 199-203, 1983.

[4] W. Dai. B-Money; 1998. Available from: https://www.weidai.com/bmoney.txt [Last accessed on 2019 Aug 10].

[5] N. Szabo. Bit Gold; 2008. Available from: https://unenumerated.blogspot.com/2005/12/bit-gold.html [Last accessed on 2019 Aug 10].

[6] Y. Zhu, D. Dickinson and J. Li. Analysis on the influence factors of bitcoin's price based on VEC model. *Financial Innovation*, vol. 3, p. 3, 2017.

[7] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, vol. 31, no. 3, pp. 307-327, 1986.

[8] F. Corsi. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, vol. 7, no. 2, pp. 174-196, 2009.

[9] F. Chung, E. Y. Sun and K. C. Shih. *Do HAR and MIDAS Models Outperform Implied Volatility Model? Evidence from Range-Based Realized Volatility. Manuscript*. Oversea Chinese Institute of Technology, 2008.

[10] P. Sea. Heterogeneous autoregressive model of the realized volatility: Evidence from Czech stock market. Advances in Finance and Accounting, pp. 32–37, 2013

[11] D. I. Vortelinos. Forecasting realized volatility: HAR against principal components combining, neural networks, and GARCH. *Research in International Business and Finance*, vol. 39, pp. 824-839, 2017.

[12] Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," version designed by K. Nordby, Nov. 3, 2020. [Online]. Available: http://www.KlausNordby.com/bitcoin

[13] M. Crosby, P. Nachiappan, P. Pattanayak, S. Verma and V. Kalyanaraman. Blockchain technology: Beyond bitcoin. *Applied Innovation Review*, vol. 7, no. 2, pp. 5-20, 2016.

[14] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, vol. 55, pp. 987-1007, 1982.

[15] A. J. M. Karim and N. M. Ahmed. Vector autoregressive integrating moving average (VARIMA) model of COVID-19 pandemic and oil price. *International Journal of Professional Business Review*, vol. 8, no. 1, p. e0988, 2023.

[16] B. K. Ahmed, S. A. Rahim, B. B. Maaroof and H. A. Taher. Comparison between ARIMA and Fourier ARIMA model to forecast the demand of electricity in Sulaimani Governorate. *Qalaai Zanist*

Journal, vol. 5, no. 3, pp. 908-940, 2020.

[17] A. I. McLeod and W. K. Li. Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis*, vol. 4, pp. 269-273, 1983.

[18] A. Asraa, W. T. Shasho, W. Rodeen and H. Tahir. Forecasting the impact of waste on environmental pollution. *International Journal of Sustainable Development and Science*, vol. 1, no. 1, pp. 1-12, 2018.

[19] A. A. Aziz, B. M. Shafeeq, R. A. Ahmed and H. A. Taher. Employing recurrent neural networks to forecast the dollar exchange rate in the parallel market of Iraq. *Tikrit Journal of Administrative and Economic Sciences*, vol. 19, no. 62, pp. 531-543, 2023.

[20] A. A. Weiss. ARMA models with ARCH errors. *Journal of Time Series Analysis*, vol. 5, pp. 129-143, 1984.

[21] L. Lon-Mu. *Time Series Models with Heteroscedasticity*. Vol. 1., Ch. 11. Springer, Germany, 2008.

[22] N. M. Ahmed and A. J. M. Karim. Multivariate time series analysis of COVID-19 pandemic and gold price by using error corrections model. *Tikrit Journal of Administration and Economics Sciences*, vol. 19, no. 64, pp. 674-693, 2023.

[23] H. Dyhrberg. *Bitcoin, Gold and the Dollar a GARCH Volatility Analysis. Working Paper Series, no. 15/20, UCD Centre for Economic Research*. University College Dublin, Ireland, 2015.

[24] E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 1976.

[25] M. S. Lo. *Generalized Autoregressive Conditional Heteroscedastic Time Series Models*. Simon Fraser University, Canada, pp. 16-42, 2003.

[26] R. Cont. Volatility clustering in financial markets: Empirical facts and agent-based models. In: *Long Memory in Economics*. Springer, Heidelberg, Berlin, pp. 289-309, 2007.

[27] R. A. Nader, A. J. Karim and M. M. Hussien. Using artificial neural networks and SPI measure techniques to forecast the risk of drought in Iraq and its impact on environment. *Journal of University of Human Development*, vol. 4, no. 2, pp. 69-77, 2018.

[28] T. G. Andersen, T. Bollerslev, F. Diebold and P. Labys. Modelling and forecasting realized volatility. *Econometrica*, vol. 71, no. 2, pp. 579-625, 2003.

[29] D. Ruppert and D. S. Matteson. *Statistics and Data Analysis for Financial Engineering*. *Springer Texts in Statistics*. 1st ed. Springer, Germany, 2010.

[30] P. Kottarathil. *Bitcoin Historical Dataset*. Kaggle; 2021. Available from: https://www.kaggle.com/datasets/prasoonkottarathil/btcinusd [Last accessed on 2024 Dec 10].

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Medical Waste Management and Treatment Techniques: Insights from Sulaymaniyah Governorate, Kurdistan Region- Iraq

Chrakhan Rahim Rashid[1], Aryan Hamid Fatah[2], Jamila Jalal Tahir[1]

[1]Department of General Sciences, College of Education, Charmo University, [2]Department of Pharmaceutical Chemistry, College of Science, Charmo University, Chamchamal, Sulaymaniyah, Kurdistan Region, Iraq

ABSTRACT

Inadequate medical waste management (MWM) has serious consequences for the environment and human health. Therefore, developing an optimal MWM system in healthcare facilities is of great importance. Unlike previous research, this study aimed to evaluate the MWM techniques practiced at private hospitals in Sulaymaniyah City. Qualitative and quantitative data collection and analytical methodologies were employed. Direct observation was conducted according to the guidelines established by the World Health Organization (WHO), along with semi-structured interviews for qualitative data collection. The quantitative data on waste generation rates for the present study were derived from the pre-existing records maintained by the General Directorate of Health in Sulaymaniyah. The statistical and graphing software GraphPad Prism version 9.5.1 was used to analyze the data. The analysis revealed that the quantity of the hospitals' monthly medical waste generation ranged between 274.0 ± 167.5 kg and 1212 ± 391.6 kg. Moreover, monthly medical waste generation per bed per hospital ranged between 20.20 ± 6.527 kg and 3.177 ± 0.8819 kg. It was also found that the daily generation rate ranges between 0.0576 and 0.827 kg/bed. Concerning management practice, all the studied hospitals have a separation system, a collection system, on-site temporary storage, and personal protective equipment. Meanwhile, 83% of the studied hospitals deliver training courses for their waste management staff. In addition, the three treatment techniques practiced are autoclave shredder systems, incineration, and deep burial.

Index Terms: Healthcare Waste, Private Hospitals, Public Health, Environmental Impact, Waste Management

1. INTRODUCTION

The term medical waste refers to waste produced by healthcare facilities, including public and private hospitals, clinical laboratories, healthcare sectors, and research centers. Several key factors influence waste generation, such as waste management strategies, the percentage of patients treated on a day-care basis, the classification of healthcare facilities, the medical specializations of hospitals, and the proportion of reusable materials used in healthcare. There are two main types of medical waste: Hazardous and non-hazardous, along with general medical waste. In developing countries, waste generation is categorized as follows: 80% is general waste, 15% is infectious waste, 3% is pharmaceutical and chemical waste, and 1% is sharp waste. The remaining <1% includes special wastes, such as used batteries and damaged thermometers. [1], [2]. The composition of medical waste encompasses municipal, chemical, metal, plastic, pharmaceutical, pathological, and laboratory wastes [3]. Anatomical waste includes animal and human waste as defined within the parameters of "biomedical

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp148-155

E-ISSN: 2521-4217
P-ISSN: 2521-4209

Copyright © 2025 Chrakhan Rahim Rashid. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Chrakhan Rahim Rashid, Department of General Science, College of Education, Charmo University, Chamchamal, Sulaymaniyah, Kurdistan Region, Iraq. E-mail: Chrakhan.rahim@chu.edu.iq

Received: 15-05-2025 Accepted: 02-08-2025 Published: 18-09-2025

material," except for bodily fluids and blood, nail clippings, hair, and removed teeth. Non-anatomical infectious waste refers to any waste that is recognized or clinically evaluated to pose a risk of contamination by microorganisms. Sharps and similar waste include any clinical item capable of causing a laceration or puncture, such as needles, blades, clinical glass, syringes, and similar objects. [4]. Pharmaceutical waste encompasses products and medicinal compounds that are no longer viable for patient treatment. Chemical waste pertains to refuse generated from diagnostic or experimental procedures or any other applications that possess genotoxic properties [5]–[7]**.** Radioactive waste comprises all materials that have been tainted with radioisotopes [8]. The most perilous dimension of medical waste arises from the management of such refuse, as waste handlers are susceptible to needle stick injuries from contaminated needles and sharp instruments, which may result in infections with the human immunodeficiency virus, as well as hepatitis B and C viruses, and this risk is most pronounced in healthcare institutions among medical personnel [9]. Internationally, regulatory agreements have been concluded in the planning of medical waste management (MWM). Some of those agreements are the Basel Convention on the Control of Transboundary Movements of Hazardous Wastes and Their Disposal [10], [11]. Moreover, within any country, national legislation constitutes the foundational framework for the enhancement of MWM practices. According to an investigation on medical waste by the World Health Organization (WHO) [12], in 22 developing countries, the percentage of healthcare institutions that failed to employ appropriate waste disposal techniques varied between 18% and 64%. A research investigation conducted at King Abdullah Hospital in Jordan elucidated that the present standards still fail to adequately safeguard the health and safety of healthcare practitioners both within public hospital settings and in external environments [13]. Moreover, a study conducted in Morocco indicated that the operational practices in the majority of surveyed hospitals did not adhere to the principles stipulated in Moroccan legislation [14]. In Iran, an examination performed at Tehran Hospital disclosed that the management of medical waste within healthcare institutions is grossly inadequate, with a conspicuous absence of appropriate health and environmental safety protocols [15]. In Turkey, the management of medical waste has been improved. According to a study by Arikan *et al.*, since 2010, medical waste has been sterilized and rarely incinerated [16]. Moreover, in healthcare centers, there is a manager and a management plan for MWM [17]. In Kurdistan of Iraq, Sulaymaniyah City, an investigation has been conducted that indicates a deficient understanding of MWM, accompanied by widespread hazardous practices among healthcare professionals (HCPs) concerning waste management and occupational health safety [18]. Inadequate management practices exert both direct and indirect effects on healthcare Staff, patients, waste pickers, and the overall hospital environment [19], [20]. The resistance to treatment and the pathogenicity of medical waste make it more dangerous and infectious. Water, air, and land can also be contaminated by medical waste. Through the incineration process of the improper treatment of medical waste, some harmful gases are emitted into the environment, which cause serious problems, such as skin diseases and cancer [21]. In addition, poor MWM leads to many other health and environmental issues, such as unpleasant smell, insect growth, and transmission of diseases through contaminated sharps [22], [23]. There is a huge desire for disposable medical products, which has increased the amount of waste generated and, then, more environmental and health problems [24]. The management of medical waste in an appropriate manner is of paramount significance to mitigate risks associated with both human health and environmental integrity. This study aimed to evaluate the MWM practices of selected private hospitals in Sulaymaniyah City by:

1. Quantifying the daily and monthly medical waste generation rates (in kg/bed/day).
2. Presenting the current MWM practices and techniques implemented by private hospitals.

In addition, the study offers recommendations for improvement based on the findings.

2. METHODOLOGY

2.1. Study Area Description

The present study was conducted in Sulaymaniyah, one of the four major urban cities of the Kurdistan Region of Iraq. The location of this city is approximately 370 km northeast of Baghdad, Iraq's capital, between latitudes from 35° 0' 0" N to 35° 45' 0" N and longitudes from 44° 50' 0" E to 45° 45' 0" E [25]. The city is located in a valley at the foot of Goizha Mountain in the north, stretching southward to the lower lands around the Tanjaro River, covering about 2400 km^2 with a population of approximately 963,390 residents in 2016 [25], [26].

2.2. Study Design and Data Collection Methods

A Mixed-method research approach that integrates both qualitative and quantitative data collection and analytical

methodologies has been utilized to achieve the goals of the present study. The study was limited to private hospitals that provided informed consent and agreed to participate in qualitative interviews.

Data were collected from six private hospitals located in Sulaymaniyah City (five general hospitals and one specialized hospital). In accordance with the principles of confidentiality, the names of the hospitals were designated as H1–H6 for the collection and utilization of information gathered from each hospital. Direct observation following the guidelines provided by the WHO for the safe management of wastes from health-care activities, along with semi-structured interviews, were used in this study. One interview has been conducted at each hospital. Respondents included the infection control officer, nursing superintendent, and the hospital's scientific research officer. Notes were taken during the interviews and after observing waste segregation, collection, treatment, and disposal practices. Quantitative data on the waste generation rate for the present study were based on the existing records from the General Directorate of Health in Sulaymaniyah.

2.3. Data Analysis Techniques

The results are presented as Mean ± standard deviation; an independent samples t-test was used to identify significant differences between the two groups. The data were analyzed using the statistical and graphing software GraphPad Prism version 9.5.1. Normality was assessed using the Kolmogorov-Smirnov test. Mann-Whitney analysis was used to investigate differences between non-parametric values. A $P < 0.05$ was considered statistically significant [27].

3. RESULTS AND DISCUSSION

3.1. Medical Waste Generation

The recorded data from January to June 2024 were analyzed to estimate the Monthly average of medical waste generation across the hospitals. The data indicated that the lowest quantity of medical waste generation among the included hospitals in this study for one month was 274.0 ± 167.5 kg, and the highest amount was 1212 ± 391.6 kg, as shown in Fig. 1 and Table 1.

The data also revealed that the maximum monthly medical waste generation for one bed was 20.20 ± 6.527 kg/hospital, while the minimum amount was 3.177 ± 0.8819 kg, as displayed in Fig. 2 and Table 2.

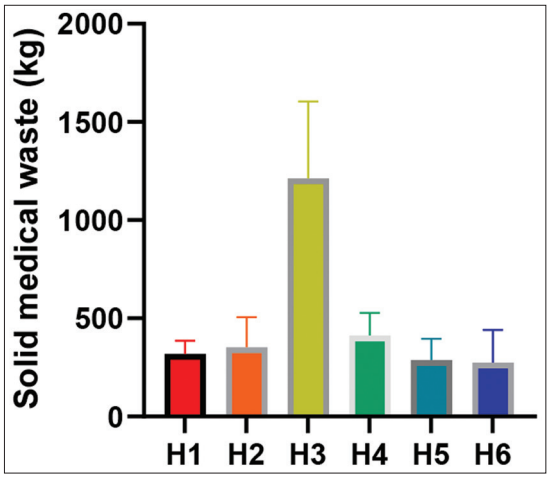


Fig. 1. Solid medical waste generation for hospitals/month.

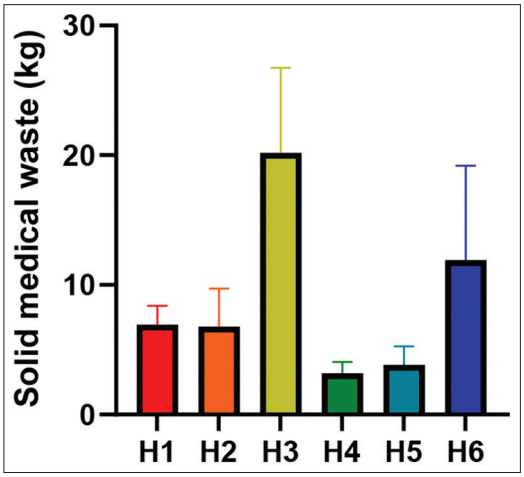


Fig. 2. Medical waste generation for a bed/hospital/month.

TABLE 1: Solid medical waste generation for hospitals/month

Hospital	Amount of solid medical waste (Mean±standard deviation)
H1	319.0±66.75
H2	353.0±152.4
H3	1212±391.6
H4	413.0±114.7
H5	288.0±107.1
H6	274.0±167.5

This indicates that the rate of medical waste generation in hospitals in this study was between 0.0576 and 0.827 kg/bed/Day. However, factors, such as hospital type, size, patient volume, and services offered may affect the variation in waste generation [28].

TABLE 2: Medical waste generation for a bed/ hospital/month

Hospital	Amount of solid medical waste (Mean±SD)
H1	6.935±1.451
H2	6.788±2.930
H3	20.20±6.527
H4	3.177±0.8819
H5	3.840±1.428
H6	11.91±7.284

According to studies, in some hospitals in Baghdad, Ramadi, and Tikrit, the rates of medical waste generation range from 0.43 kg/bed/day to 1.11 kg/bed/day, and they differ among areas and healthcare institutions in Iraq [29], [30]. Upon comparing the findings of this study with those of other cities in Iraq, it is apparent that the medical waste generation in the hospitals studied in Sulaymaniyah City was lower than the typical levels reported in hospitals from the middle and southern parts of Iraq. In other developing nations, such as Iran, Turkey, India, Thailand, and Bangladesh, hospital waste generation ranges from 0.11 to 3.9 kg/bed/day [31].

The findings further confirm that the medical waste generation in the present study is lower than what has been reported by the WHO (2014), which estimated the range of medical waste generation at 0.5–3.0 kg/bed/day in countries worldwide [32].

3.2. Waste Management Practices

Qualitative data on the recommended steps of MWM were collected and analyzed to assess MWM in the studied hospitals. All studied hospitals provide Personal Protective Equipment PPE for their waste management staff, as shown in Fig. 3 and Table 3.

3.2.1. Source Segregation and Collection

The results presented in Fig. 3 and Table 3 demonstrate a high proportion of waste separation in hospitals, as all 6 hospitals were doing source separation for waste, with a percentage of 100%.

Hospitals systematically separate different waste categories at the source, such as sharp, infectious, and general waste. Doctors, head nurses, pharmacists, and laboratorians working in the clinics and various hospital units carry out this process. This segregation is necessary to lower the risk of contamination and ensure that hazardous items are managed correctly. It also diminishes environmental and health risks associated with hospital waste and lowers waste management costs in private hospitals [28].

TABLE 3: Waste management practice in the studied hospitals

Waste management practices	Number of hospitals/practices	Occurrence %
Separation	6	100
Collection		
Four-bin system	4	66
Three-bin system	2	34
On-site temporary storage	6	100
Personal protective equipment	6	100
Training course	5	83
On-site waste treatment	0	0

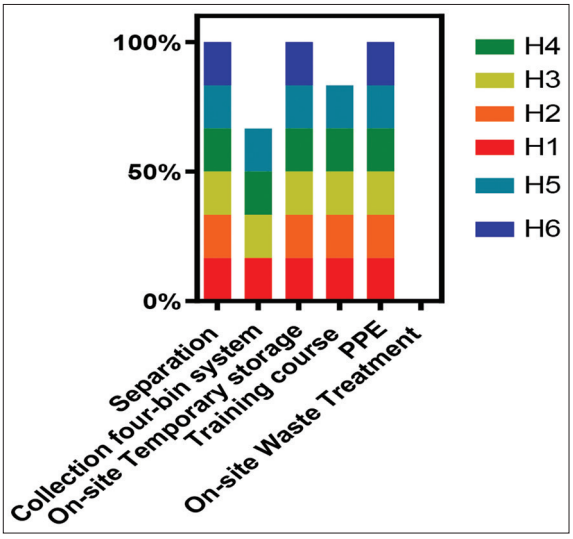


Fig. 3. Waste management practices in the studied hospitals.

The Color-coded bin system, following the WHO (2017) approach, was used for various waste categories to help with appropriate segregation in the studied hospitals (Fig. 4). The data revealed that four out of the six hospitals under study used the 4-bin collecting system, with a percentage of 66%according to the data in Table 3 and Fig. 3.

- Yellow bin with a yellow plastic bag in the bin for infectious, chemical, and pharmaceutical wastes.
- Disposable plastic or cardboard medical biohazard sharp containers for sharps (Fig. 5).
- Red bin with red plastic bag in the bin for anatomical waste. Particularly used in hospitals with a Maternity unit.
- Black bin with black plastic bag for general/non-hazardous waste.

By filling three-quarters of the bins' capacity, all kinds of waste are collected following WHO guidelines. This is to prevent spillage and kept covered to avoid casual access by individuals or disease vectors [33].

Fig. 4. Color-coded bins for collecting waste at the studied hospitals.

Fig. 5. Disposable medical biohazard sharp bins utilized in the studied hospitals.

Presented data from Table 3 and Fig. 3 also reveal that 83% of the studied hospitals conduct training courses on MWM for their medical and janitorial staff. The goal of training is to improve skills and increase awareness among healthcare workers. WHO (2017) has emphasized the necessity of intensive training courses, suitable personal protective equipment PPE, and more efficient resource use to improve the efficiency of MWM in healthcare facilities [33].

3.2.2. Storage and Transportation

On-site storage serves as a transitional phase before waste is processed or sent to other facilities. Efficient storage practices are essential for minimizing the risks associated with hazardous waste. Storage sites were located in the backyards and covered with plastic roofs to shield the waste from rain and direct sunlight. The areas were secured with concrete walls and metallic chain-link fences to minimize exposure to people and animals, as captured in Fig. 6. However, the storage areas in the studied hospitals were partially secured in terms of access to vectors that may cause the spread of diseases. According to WHO (2017) and UNEP (2021), the on-site storage space should be situated away from areas used for patient care, places where food is prepared, and areas where the public can enter. Only authorized staff and

Fig. 6. On-site storage area for temporarily storing medical waste.

waste collection vehicles should have access to these areas. To prevent leaks or mitigate contamination, storage facilities must include drainage systems, adequate ventilation, and impermeable, easy-to-clean flooring [33], [34].

Medical wastes are collected from the on-site storage areas 2–3 times/week. As stated in WHO guidelines (2017), the length of time medical waste can be kept on-site is determined by the waste type and the local climate. Without refrigeration, infectious waste should not be kept for longer than 48 h in a tropical climate or 72 h in a temperate climate. To avoid deterioration and odor release, extended storage times require cooling below 8°C [33].

Medical waste is typically kept in yellow-colored containers separate from regular waste in hospitals. Data in Table 3 and Fig. 3 reveal that none of the studied hospitals implement on-site treatment practice. This is due to the cost-effectiveness of treatment technologies and space limitations in certain hospital areas. To ensure the safe transportation, treatment, and final disposal of medical waste, the studied hospitals made a (Pay as you generate) contract with the Ministry of Health/ General Directorate of Health in Sulaymaniyah. Accordingly, private hospitals incur costs to the government based on the weight of the medical waste they generate. The hospital is billed depending on the measured waste, with different charges for different types of waste (hazardous, infectious, or anatomical). Penalties for underreporting or incorrect segregation could apply. This encourages hospitals to source-segregate wastes, which reduces the cost of waste disposal as well as the environmental footprint of hospital garbage [33], [35]. Waste is eventually transferred by the specialized Box Trucks for off-site treatment processes and final disposal.

3.2.3. Treatment Techniques

The effective treatment of medical waste is crucial for safeguarding public health, ensuring environmental safety,

and averting the transmission of infectious illnesses. Autoclave-shredder systems, incineration, and deep burial are the most popular techniques for medical waste handling in the studied area.

3.2.3.1. Autoclaving

Collected medical wastes are then treated using autoclave-shredder systems, which combine mechanical shredding and steam sterilizing [36]. By subjecting the waste to high-pressure steam at temperatures between 121°C and 134°C, autoclaving effectively destroys germs. This is followed by a shredding process, which increases the surface area and reduces the volume of the waste, ensuring thorough sterilization. According to M. A. Asad (Head of infection control, General Directorate of Health in Sulaymaniyah, personal Communication. Aug. 7, 2024), the General Directorate of Health in Sulaymaniyah placed two autoclave shredders in the Sulaymaniyah City Center, each capable of processing around 1000 kg of medical waste/day (Fig. 7).

The autoclave-shredding system is environmentally sustainable, as it produces no harmful emissions and is appropriate for processing sharps, contaminated plastics, textiles, and other biohazardous items. Although it has high operating costs due to high energy demand, and is not suitable for chemical and pharmaceutical waste [37].

3.2.3.2. Incineration

Medical waste is burned at high temperatures during incineration. It converts the waste's combustible components into gaseous byproducts, such as CO_2, H_2O, CO, and other gases (some of which are poisonous), while non-combustible components remain as a solid byproduct (ash) [36]. Incineration is carried out at temperatures ranging from 800 to 1200°C in the presence of sufficient air to destroy infections and organic wastes [38].

In the present study area, six incinerators are installed (Fig. 8). Two incinerators, which are located in a secure room with more advanced smoke treatment, are used for prohibited drug eradication as discussed by (N. M. Ahmed, incineration plant supervisor, General Directorate of Health in Sulaymaniyah, Personal Communication, Jan 16, 2025).

Incineration can release toxic emissions, such as furans and dioxins; therefore, sophisticated air pollution control systems are necessary to mitigate their environmental impact [38].

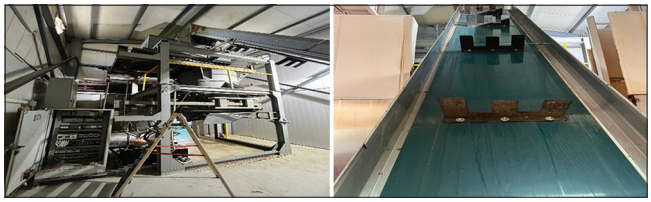


Fig. 7. Autoclave-shredder at the Sulaymaniyah teaching hospital in Sulaymaniyah city.

Fig. 8. Incinerators in the Tanjaro area.

In the present study area, the incinerators are located outside the city, precisely in the Tanjaro area near the dumping site. The leftover ashes are then buried in the specially designed concrete pits and covered to minimize the risk of exposure and contamination.

3.2.3.3. Deep burial

Deep burial is the disposal of garbage in defined sites, often several meters below the surface. This strategy can be especially beneficial in developing nations where resources and facilities for more advanced waste treatment are restricted [35]. Certain forms of waste are appropriate for deep burial. This comprises anatomical waste, microbiological trash, and other garbage that has been treated to lower its hazard level.

In the present research area, waste is deposited in an approximately 7–8-m pit in strata, with each stratum being encapsulated by a layer of soil. Upon filling the pit, it is enclosed with a substantial layer of soil, which aids in reducing odors and deterring scavengers, as discussed by Asad (2024). Although the deep burial method can serve as an effective waste disposal technique, inadequately managed burial sites may result in the leaching of hazardous compounds into groundwater and soil, along with the possible emission of greenhouse gases [39].

Despite the mentioned methods, other advanced approaches are practiced for hospital waste treatment in developed nations, including Microwave Disinfection, UV Disinfection, and Chemical disinfection using various chemical agents [40].

4. CONCLUSION

MWM begins at the point of generation and continues through the entire treatment process. In the present study, only private hospitals that provided informed consent and agreed to participate in qualitative interviews were included. The findings revealed that the maximum medical waste generation, 0.827 kg/bed/day, is lower than the WHO standard for medical waste and some other cities in the middle and southern parts of Iraq. The diversity in waste creation may be affected by hospital size, type, patient volume, and services offered.

The management procedure in Sulaymaniyah City's private hospitals starts with an infection control unit. The insights gathered from interviews at the participant hospitals underscore the critical importance of effective MWM in safeguarding both healthcare workers and patients. The data reveals that most hospitals conduct training courses on MWM for their medical and janitorial staff. Color-coded bins were applied to separate waste into various categories, and 66% of the studied hospitals used a 4-bin system. Each hospital has a designated area with specified features for collecting garbage, and none of them has a treatment unit. The General Directorate of Health in Sulaymaniyah handles the transportation, treatment, and final disposal phase. Medical waste is transported to various treatment facilities, such as autoclave-shredder units, incineration, and deep burial, depending on the type of waste, using specialized box trucks.

RECOMMENDATIONS

In light of the findings of this study, the following recommendations are suggested:

- Increase capacity and raise awareness: Public awareness campaigns and extensive training programs for HCPs can aid in enhancing MWM -related knowledge, attitudes, and behaviors.
- Invest in sustainable treatment and disposal technologies: The governorate ought to investigate and finance eco-friendly treatment and disposal techniques, such as microwave disinfection, autoclaving, or other technologies that have minimal negative effects on the environment.
- Continuous monitoring and control systems: Continuous supervision and control systems are essential to guarantee adherence to MWM laws and enhance procedures in all healthcare facilities.

ACKNOWLEDGMENT

We would like to express our sincere appreciation to the administration and infection control units staff at the General Directorate of Health in Sulaymaniyah, as well as to all private hospitals that contributed to this research, for their invaluable assistance and collaboration in providing the requisite data and insights.

REFERENCES

[1] A. Prüss, E. Giroult and P. Rushbrook. "*Safe Management of Wastes from Health-Care Activities*". World Health Organization, Geneva, Switzerland, 2014.

[2] Y. Y. Babanyara, D. B. Ibrahim, T. Garba, A. G. Bogoro and M.Y. Abubakar. "Poor medical waste management (MWM) practices and its risks to human health and the environment: A literature review". *World Academy of Science Engineering and Technology International Journal of Health and Medical Engineering*, vol. 7, no. 11, p. 757, 2013.

[3] S. Pradhan, S. Prasad, B. R. Chinmaya and S. Tandon. "Perception of biomedical waste management among dental health care personnel of various dental colleges in Delhi NCR, India: A knowledge, attitude, and practice study". *International. Journal of Scientific Study (IJSS)*, vol. 10, no. 2, pp. 208-215, 2016.

[4] S. F. Raphela. "Treatment and disposal of medical waste in rural and Urban clinics within Polokwane municipality of South Africa". *Journal for New Generation Sciences*, vol. 12, no. 2, pp. 97-109, 2014.

[5] R. R. A. M. Mato and M. E. Kaseva. "Critical review of industrial and medical waste practices in Dar es Salaam City". *Resources Conservation Recycling*, vol. 25, no. 3-4, pp. 271-287, 1999.

[6] M. M. A. El-Salam. "Hospital waste management in El-Beheira Governorate, Egypt". *Journal of Environmental Management*, vol. 91, no. 3, pp. 618-629, 2009.

[7] M. Attrah, A. Elmanadely, D. Akter and E. R. Rene. "A review on medical waste management: Treatment, recycling, and disposal options". *Environments*, vol. 9, no. 11, p. 146, 2022.

[8] Republic of South Africa. "*Nuclear Energy Act, 1999, Report 20759*". 1999.

[9] S. A. Tabish. "*Hospital Infection Control: Conceptual Framework Ecohealth: Management of Biomedical Waste*". Academa Publishers, West Bengal, pp. 139-145, 2005.

[10] M. S. Sundram. "Basel convention on transboundary movement of hazardous wastes: Total ban amendment". *Pace International Law Review*, vol. 9, no. 1, p. 1, 1997.

[11] International Committee of the Red Cross. "*Medical Waste Management"*. 2011.

[12] World Health Organization (WHO). "*Safe Healthcare Waste Management: Policy Paper*". WHO, Geneva, Switzerland, 2004.

[13] F. Abdulla, H. A. Qdais and A. Rabi. "Site investigation on medical waste management practices in Northern Jordan". *Waste Management*, vol. 31, pp. 450-458, 2008.

[14] A. Mbarki, B. Kabbachi, A. Ezaidi and M. Benssaou. "Medical waste management: A case study of the souss-massa-drâa region, Morocco". *Journal of Environmental Protection*, vol. 4, no. 9, pp. 914-919, 2013.

[15] F. Malekahmadi, M. Yunesian, K. Yaghmaeia and K. Nadafi. "Analysis of the healthcare waste management status in Tehran hospitals". *Journal of Environmental Health Science and Engineering*, vol. 12, no. 1, p. 116, 2014.

[16] O.A. Arikan, I. Ozturk, I. Demir and K. E. Maçin. "Medical Waste Management in Turkey". *2nd International Conference on Environmental Science, Belgrade- Serbia*, 2016.

[17] M. Akbolat, H. Saglam, C. Dede and O. Işık. "Medical waste management practices in Turkey: A case study in Sakarya". *Pakistan Journal Medical Scienc*e, vol. 27, pp. 892-895, 2011.

[18] S. M. Mohammed, N. Othman, A. H. Hussein and K. J. Rashid. "Knowledge, attitude and practice of health care workers in Sulaimani health facilities in relation to medical waste management". *Kurdistan Journal of Applied Research*, vol. 2, no. 2, pp. 143-150, 2017.

[19] M. A. Patwary, W. T. O'Hare, G. Street, K. M. Elahi, S. S. Hossain and M. H. Sarke. "Country report: Quantitative assessment of medical waste generation in the capital city of Bangladesh". *Waste Management,* vol. 29, no. 8, pp. 2392-2397, 2009.

[20] S. A. Tamplin, D. Davidson, B. Powis and Z. O'Leary. "Issues and options for the safe destruction and disposal of used injection materials". *Waste Management*, vol. 25, no. 6, pp. 655-665, 2005.

[21] S. V. Manyele. "Effects of improper hospital waste management on occupational health and safety". *African Newsletter on Occupational Health and Safety*, vol. 14, pp. 30-33, 2004.

[22] A. Coker and M. K. C. Sridhar. "Increase in healthcare facilities and rapid environmental degradation: A technological paradox in Nigeria's Urban centers". *African Journal of Environmental Science and Technology*, vol. 4, no.9, pp. 577-585, 2010.

[23] S. Yitayel, A. Tamrat and M. Adane. "Sharps injuries and exposure to blood and bloodstained body fluids involving medical waste handlers". *Waste Management and Research*, vol. 30, pp. 1299-1305, 2012.

[24] J. J. Edmund. "Characterization and analysis of medical solid waste in Osun State, Nigeria". *African Journal of Environmental Science and Technology*, vol.5, no.12, pp. 1027-1038, 2012.

[25] K. Muhamad and R. Alkaradaghi. "*Landfill Site Selection for Sulaimaniyah City, NE - Iraq"*. University of Sulaimani, Sulaymaniyah, Iraq, [Thesis], 2021.

[26] Kurdistan Region Statistics Office (KRSO). "*Displacement as Challenge and Opportunity: Urban Profile of Refugees, Internally Displaced Persons, and Host Community Sulaymaniyah Governorate and Garmian Administration, Ministry of Planning*". Kurdistan Region of Iraq, Erbil, 2016.

[27] T. W. MacFarland and J. M. Yates. "*Introduction to Nonparametric Statistics for the Biological Sciences Using R*". Springer International Publishing, Cham, pp. 103-132, 2016.

[28] A. Hosseinzadeh, M. A. M. Reshadi, M. Nazaripour and M. Rezaei. "Medical waste management in private hospitals in Tehran". *Journal of Advances in Environmental Health Research*, vol. 11, no. 3, pp. 142-146, 2023.

[29] S. K. Ali and D. T. Jasim. "Assessment of medical solid waste generation rates for teaching hospitals in Baghdad city". *In Association of Arab Universities Journal of Engineering Sciences*, vol. 25, no. 1, pp. 160-169, 2018.

[30] S. Thameel, S. F. Al-Chalabi, A. Mustafa and A. A. Mohsin. "An evaluation of health care waste generation and disposal at Ramadi teaching hospital in Iraq". *Iraqi Journal Civil Engineering*, vol. 16, no. 1, pp. 1-9, 2022.

[31] J. Torkashvand, H. Pasalari, A. Jonidi Jafari, M. Kermani, M. Farzadkia, J. Mehralipour and A. Ghalkhanbaz. "Dental solid waste management in Iran: A systematic review". *Journal Advances Environmental Health Research*, vol. 8, no. 4, pp. 297-303, 2020.

[32] World Health Organization (WHO). "*Safe Management of Wastes from Health-Care Activities"*. 2nd ed. World Health Organization, Geneva, 2014.

[33] World Health Organization (WHO). "*Safe Management of Wastes from Health-Care Activities: A Summary"*. Geneva, 2017.

[34] United Nations Environment Programme (UNEP). *Healthcare Waste Management: Global Guidelines*. UNEP, Nairobi, 2021.

[35] M. Ali, W. Wang, N. Chaudhry and Y. Geng. "Hospital waste management in developing countries: A mini review". *Waste Management and Research*, vol. 35, no. 6, pp. 581-592, 2017.

[36] L. F. Diaz, G. M. Savage and L. L. Eggerth. "Alternatives for the treatment and disposal of healthcare wastes in developing countries". *Waste Management*, vol. 25, no. 6, pp. 626-637, 2005.

[37] World Health Organization (WHO). "*Overview of Technologies for the Treatment of Infectious and Sharp Waste from Health Care Facilities"*. WHO Press, Geneva, 2019.

[38] P. C. Ojha, S. S. Satpathy, A. K. Ojha, L. B. Sukla and D. Pradhan. "Overcoming challenges due to enhanced biomedical waste generation during COVID-19 pandemic". *Science Total Environment*, vol. 832, p. 155072, 2022.

[39] H. Selman, H. Kubba, N. Al-Mukaram and R. Alkateeb. "Heavy metal pollution from hospital waste incinerators: A case study from Al-Muthanna province, Iraq". *IOP Conference Series Materials Science and Engineer*, vol. 1090, no. 1, p. 012036, 2021.

[40] S. Sahoo, W. Rathod, H. Vardikar, M. Biswal, S. Mohanty and S. K. Nayak. "Biomedical waste plastic: Bacteria, disinfection and recycling technologies-a comprehensive review". *International Journal of Environmental Science and Technology*, vol. 21, no. 1, pp. 1141-1158, 2024.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Regression Analysis of Soil Properties for Small Dam Bodies samples in Chamchamal and Qaradagh Districts, Sulaymaniyah Governorate

Abdalmajeed Mohammed Rahman* and Nawbahar Faraj Mustafa

Department of Water Resources Engineering, University of Sulaimani, Kirkuk Road, Sulaymaniyah, 46002, Iraq

ABSTRACT

Effective geotechnical engineering relies on accurately predicting soil parameters such as cohesion and the angle of internal friction, which are critical for ensuring the stability of structures such as small dams. Traditional laboratory testing can be prohibitively expensive and time-consuming, highlighting the need for efficient predictive models. This article aims to develop regression equations that estimate these parameters using easily obtainable soil properties in Chamchamal and Qaradagh Districts, Sulaymaniyah Governorate. Using soil water content, soil density, and plasticity index as key predictors, the one-way analysis of variance analysis achieved an R-squared value of approximately 0.87, with a root mean square error of 0.15 and a bias of about −1.2%, demonstrating high accuracy and robustness across different datasets. The analysis further revealed that increases in plasticity index significantly impacted the angle of internal friction (P = 0.014), while dry density showed a strong positive influence on cohesion. These findings underscore the role of soil parameters in estimating the soil compression index and demonstrate that a simplified, empirically derived model can offer practical insights for geotechnical applications. However, given the moderate correlation levels observed (R^2 = 0.38 for cohesion and R^2 = 0.61 for internal friction angle), the predictive capability of the models is limited. Therefore, the developed regression models should be regarded as preliminary tools, useful for initial assessments, but must be complemented by thorough field investigations and comprehensive engineering analyses to ensure the reliability and safety of dam structures.

Index Terms: Soil Properties, Regression Analysis, Small Dam Stability, Soil Behaviors, Soil Parameters Prediction

1. INTRODUCTION

Understanding the geotechnical properties of soil is a vital aspect of designing, constructing, and maintaining civil engineering structures, especially earth dams [1]. A solid soil foundation is critical when selecting a site for a dam, as it ensures stability and helps prevent erosion and potential failures [2].

Key parameters like shear strength (SS), which include cohesion and the angle of internal friction, play a significant role in various engineering projects. These parameters are especially important for determining the bearing capacity of foundations and assessing soil settlement [3], [4]. However, measuring these properties through traditional laboratory testing can be both time-consuming and costly, creating a need for more efficient assessment methods. This challenge has led engineers to explore predictive models that use regression techniques to estimate soil properties more easily [5].

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp156-165

E-ISSN: 2521-4217
P-ISSN: 2521-4209

Copyright © 2025 Rahman and Mustafa. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Abdalmajeed Mohammed Rahman, Department of Water Resources, Engineering, University of Sulaimani, Kirkuk Road, Sulaymaniyah, 46002, Iraq. E-mail: abdalmajeed.rahman@univsul.edu.iq

Received: 16-05-2025 Accepted: 24-08-2025 Published: 27-09-2025

In recent years, many geotechnical engineers have encountered difficulties when utilizing the actual soil available on-site. Often, this soil does not meet the necessary criteria for effective geotechnical design, leading to additional complications [6], [7]. Understanding and accurately predicting SS parameters through regression analysis empowers engineers to enhance safety and make informed decisions about the materials surrounding a dam [8].

By employing regression analysis, engineers can estimate important soil properties using data that is often easier to obtain than traditional laboratory tests. This not only saves time and money but also leads to more reliable and cost-effective designs [9]. Furthermore, regression techniques allow engineers to prioritize which tests to conduct, aiding in early fault detection and timely problem-solving. This ability to focus on critical tests becomes even more valuable when resources are limited, ensuring that major issues can be addressed promptly and efficiently [10].

Utilizing advanced predictive techniques like regression analysis to understand soil properties is crucial for optimizing engineering practices [11]. By enhancing the safety and sustainability of structures like earth dams, these innovations play an essential role in addressing the challenges faced in today's geotechnical engineering landscape [12]. With this knowledge and approach, engineers can better navigate the complexities of soil behavior, ultimately contributing to the success of their projects [13].

Previous studies have explored the application of regression analysis for predicting various soil properties. Sharma and Singh (2017) investigated regression-based models for predicting the unconfined compressive strength (UCS) of artificially structured soils, noting the limitations of traditional UCS determination methods [6]. They highlighted the time-consuming and costly nature of traditional UCS tests and proposed empirical equations using simple and multiple linear regression techniques. Mohammadi *et al.* (2020) used multivariate regression and artificial neural networks to predict SS parameters (C and ϕ), highlighting the impracticality of conducting numerous triaxial shear tests for large projects [14]. Hamaamin *et al.* explored indirect methods to estimate the soil compression index (Cc) using 177 undisturbed samples from cohesive soil in Sulaymaniyah Governorate, Iraq, employing Adaptive Neuro-Fuzzy Inference Systems (ANFIS) and regression techniques. The ANFIS model outperformed the regression approach, achieving an R^2 of 0.66 compared to 0.48, highlighting the effectiveness of machine learning for cost-effective Cc estimation [1].

A study by Nawaz *et al.* presented new machine learning models for estimating cohesion (c) and friction angle (ϕ) soil parameters using gene expression programming which is an evolutionary algorithm that combines the principles of genetic algorithms and genetic programming to create programs or mathematical expressions that solve specific problems. The models leveraged readily available soil attributes such as sand content, depth, specific gravity, liquid limit, plastic limit, and fine content, yielding a c-predictive model with an R^2 of 0.984 and RMSE of 1.13, and a ϕ-predictive model with R^2 of 0.927 and RMSE of 1.123. The proposed models showed significant accuracy improvements over existing models, offering a practical solution for efficient and sustainable geo-structural design [15].

The article by Khan and Wang focuses on enhancing slope stability through the development of correlations between the factor of safety (FS) and soil properties such as cohesion, friction, and unit weight. The authors find that using nailing techniques significantly improves FS, with values rising from unsafe (e.g., 1.091) to safe levels (e.g., 1.545) compared to stepped and natural slopes. Their comprehensive analysis, using various soil types, demonstrates that both nailing and stepping can effectively increase SS, making this research applicable for future slope stabilization projects [16].

These studies demonstrate the potential of regression analysis as a valuable tool in geotechnical engineering. The use of support variables in regression, as explored by Erbilen, can further enhance the accuracy of predictions, particularly when dealing with missing data [17].

1.1. Study Area

The study area encompasses fourteen small dams located within the Sulaymaniyah Governorate as shown in (Fig. 1). Name of the small dams and their locations in Sulaymaniyah Governorate are shown in (Table 1). Soil samples were collected from the dam bodies of these structures to conduct geotechnical investigations. The study was conducted within two key districts of Sulaymaniyah Governorate in the Kurdistan Region of Iraq – Chamchamal and Qaradagh – which together form the primary focus of this research. These two districts were selected due to their high concentration of unofficial small dams, making them particularly relevant for assessing sustainability and geotechnical stability. In total, fourteen small dams were analyzed, all of which are located within the boundaries of Chamchamal and Qaradagh. This region is predominantly characterized by homogeneous soil formations, mainly consisting of low-plasticity clayey and silty soils, with limited

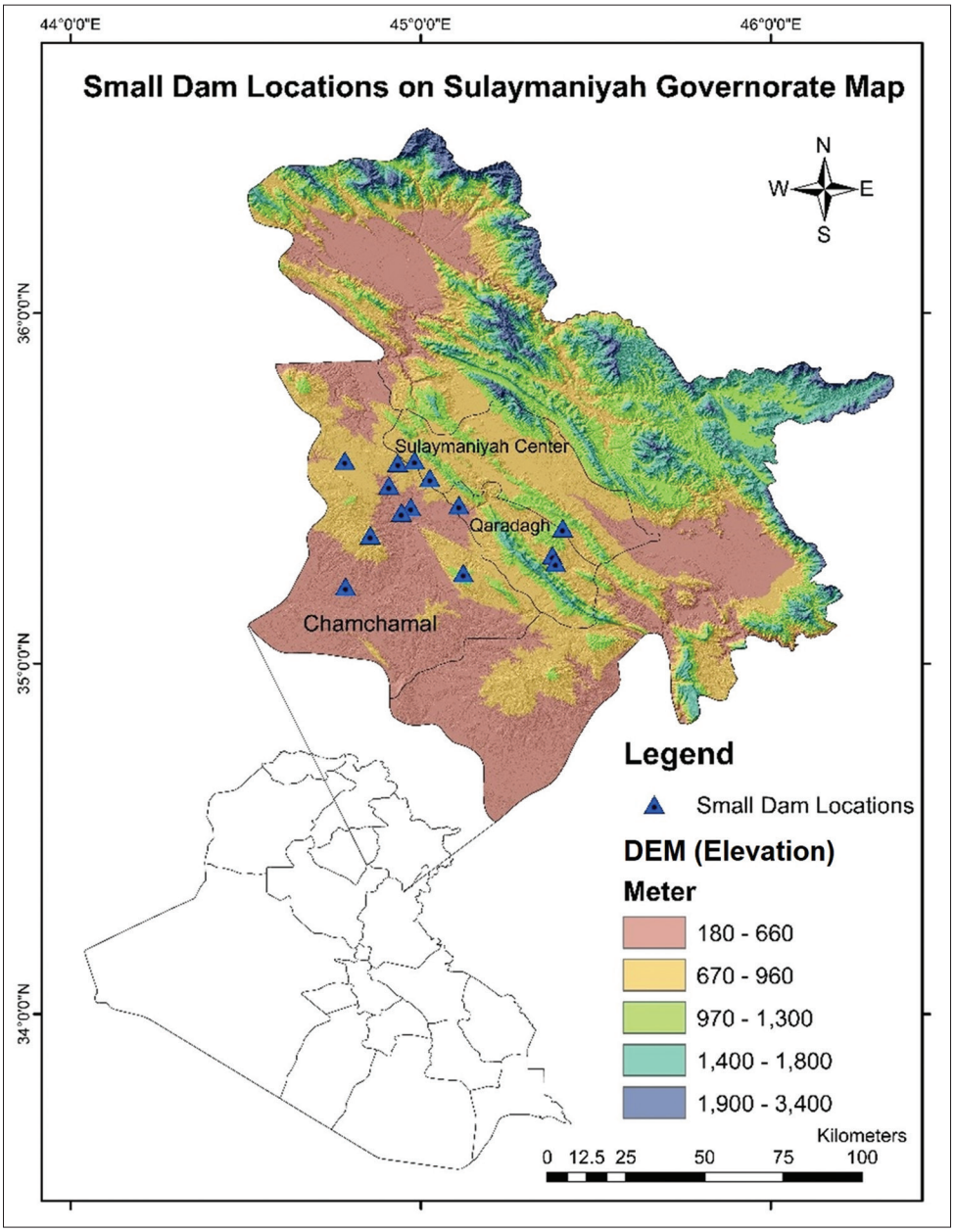


Fig. 1. Study area map and small dams location.

variations in stratigraphy across dam sites [18]. One of the critical reasons for choosing these specific dams is that most of them were constructed without standard engineering practices, particularly lacking structural elements such as cores, toe drains, or protective filters. Instead, the dam bodies were built uniformly from locally available soils, often compacted without rigorous quality control. This uniformity in construction, along with similar soil conditions, provides a consistent basis for comparative analysis. However, it also presents significant engineering concerns, especially regarding seepage control and slope stability. By narrowing the focus to this clearly defined geographic area and a relatively consistent set of soil conditions. These tests aim to assess the stability and safety of the dams, ensuring their structural integrity and reliability for ongoing use and future planning.

The objective of the article is to develop regression equations for predicting critical soil parameters, specifically cohesion and the angle of internal friction, using obtainable soil properties from small dam bodies in Sulaymaniyah Governorate, Iraq,

TABLE 1: Detailed information on the selected small dams

No.	Name of small dams	locations	Longitude	Latitude
1	Chollmak small dam	Chamchamal	35.53036	45.02549
2	Lakawa small dam	Chamchamal	35.57366	44.93386
3	Tazhga Small dam	Chamchamal	35.58107	44.98198
4	Zhalla small dam	Chamchamal	35.45302	45.10837
5	Goran small dam	Chamchamal	35.367244	44.855536
6	Alimansur small dam	Chamchamal	35.220325	44.784693
7	Qallachugha small dam	Chamchamal	35.5089	44.90872
8	Kuradawe small dam	Chamchamal	35.44752	44.97087
9	Kunakotr small dam	Chamchamal	35.4318	44.94447
10	Kallan small dam	Chamchamal	35.25878	45.12106
11	Hamza small dam	Chamchamal	35.57986	44.78268
12	Qaraman small dam	Qaradagh	35.38724	45.40487
13	Tavan small dam	Qaradagh	35.30972	45.37548
14	Chami Dewana small dam	Qaradagh	35.2887	45.38428

thereby providing a cost-effective alternative to traditional laboratory testing in geotechnical engineering.

2. METHODOLOGY

The methodology describes a systematic approach to soil sampling, laboratory testing, and statistical analysis using Minitab software to develop predictive models for critical geotechnical properties, specifically cohesion and the angle of internal friction, in small dam bodies located in Sulaymaniyah Governorate.

2.1. Soil Sampling

The best method for sampling soil in a dam body for testing analysis is the use of a systematic grid sampling approach. This involves dividing the dam area into a grid and collecting samples from designated points to ensure representative coverage of the site [19]. The soil sample for each small dam was collected at a depth of 0.5 m below the ground surface. This depth was chosen because it typically represents the active zone of the dam's embankment, where variations in moisture content, compaction, and stress conditions are most influential on the overall performance and stability of the structure [20]. Using an auger or tube sampler allows for the removal of undisturbed soil samples, which is crucial for accurate laboratory testing and analysis [21]. According to (ASTM D75), 14 soil samples were collected from small dam bodies to make soil properties tests in the laboratory [22], [23].

2.2. Laboratory Testing

The samples were subjected to a series of laboratory tests to determine their key geotechnical properties as illustrated in (Table 2). Water content was determined according to

TABLE 2: Conducted soil test details

No.	Soil parameters	Standards
1	W.C %	ASTM D2216
2	Dry density g/cm^3	ASTM D2937
3	Liquid limit %	ASTM D4318
4	Plastic limit %	ASTM D4318
5	Cohesion (C)	ASTM D3080
6	Angle of internal friction Ø	ASTM D3080
7	Fines (Silt+Clay) %	ASTM D6913

ASTM D2216 [17]. Particle size distribution was determined using sieve analysis (ASTM D6913) [17]. Atterberg limits, including liquid limit and plastic limit, were determined following ASTM D4318 [24]. The SS parameters, cohesion (c) and angle of internal friction (Ø), were measured using the direct shear test, as per ASTM D3080 [1].

2.3. Data Analysis

The data obtained from the laboratory tests were analyzed using Minitab software. Regression analysis was performed to develop predictive models for soil properties. Specifically, the analysis focused on the relationship between several independent variables to predict the angle of internal friction (ϕ) and cohesion (C).

Minitab is a powerful statistical software widely used for data analysis, including regression analysis, which is essential for understanding relationships between variables. Minitab employs methods like linear regression, which uses the least squares estimation technique to minimize the sum of the squares of the residuals (the differences between observed and predicted values) [25].

To perform regression analysis in Minitab, the data of soil parameters inputted, after selecting the regression option, and specifying the dependent variable (the one can be predicted) and independent variables (the predictors). Minitab then calculates coefficients, R-squared values, and *P*-values to assess the model's fit and significance. This process helps in identifying relationships and predicting missing data points effectively [26].

In Minitab, calculating a "Y line" generally refers to plotting a regression line or a fitted line in the context of statistical analysis, particularly in linear regression. The Y line (or fitted line) represents the predicted values of the response variable (Y) based on the predictor variable (s) (X) in a regression model. Ideally, the points should cluster around this line [27]. In a scatterplot, if the reference line is a diagonal line of best fit, it indicates that as the actual values increase, the predicted values also increase proportionately, demonstrating a good

fit of the model [28]. Deviations from this line suggest areas where the model might not be accurately predicting outcomes [29].

2.4. Assessment Metric

Table 3 is a completed version of the assessment metrics table based on the content of the article and includes an illustration of the terms used in the analysis.

3. RESULTS AND DISCUSSION

The regression analysis provides valuable insights into predicting critical soil properties essential for the structural integrity of small dam bodies [36]. By understanding how various soil parameters interact, engineers can make informed decisions that enhance the safety and effectiveness of engineering designs [37].

The regression analysis demonstrated that there is a relationship between dry density and the other measured soil properties. The resulting regression equation provides a basis for understanding how these variables interact. All data collected from the 14 soil samples, obtained through both laboratory and field tests, are presented in Table 4. These data reveal a considerable range in the values of key soil parameters, reflecting the inherent variability in the geotechnical characteristics of materials sourced from different dam bodies. Notably, the cohesion and angle of internal friction vary significantly among the samples, indicating differences in the strength and stability of the soils across the studied sites.

TABLE 3: Metrics used in the analysis

No.	Term	Meaning and how it was used in the article
1	F-value	A statistical test value used to assess the overall significance of the regression model [30]. Found in Minitab in the ANOVA or regression output under "F-Value". Example: "F-Value" in ANOVA table (e.g., 90.99).
2	*P*-value	Probability that the observed relationship occurred by chance [31]. Used to evaluate the significance of predictors, listed next to each F-value or *t*-value in the output; if $P<0.05$, the effect is typically considered significant.
3	R-squared (R^2)	Indicates the proportion of variance in the dependent variable explained by the model [32]. Shown in the Model Summary: "R-sq" (e.g., 0.86).
4	Adjusted R-squared	Corrected R-squared that accounts for the number of predictors [33]. Provided alongside R-squared in model summary (e.g., 85.02%).
5	Standardized effects (Pareto chart)	Visualizes the significance of each predictor variable [34]. Displayed as a Pareto chart in Minitab, ranked by magnitude. Terms crossing the significance line are considered influential.
6	Regression coefficient	Numerical value indicating the effect size of each predictor [35]. Listed in regression output (e.g., Coef for "Dry Density"=44.4). *t*-value and *P* value assess significance.
7	α	The significance level (α) represents the probability of rejecting the null hypothesis when it is actually true – that is, the risk of committing a Type I error. In this analysis, α was set at 0.05, meaning there is a 5% chance of concluding that significant differences exist among the group means when, in fact, no such differences are present.

ANOVA: Analysis of variance

TABLE 4: Soil parameter results for 14 small dam bodies

No.	W.C %	Dry density g/cm³	Liquid limit %	Plastic limit %	Plasticity index	Cohesion C (KPa)	Angle of internal friction Ø in degrees	Sand %	fines (Silt+clay) %
1	8.02	1.70	26.87	17.70	9.17	40.33	22.95	42.78	43.27
2	8.05	1.74	24.30	16.58	7.72	23.85	33.88	36.43	48.15
3	11.04	1.56	29.10	18.24	10.86	14.33	25.03	25.91	64.60
4	9.54	1.58	27.54	17.83	9.71	24.99	26.57	43.78	46.13
5	16.83	1.42	31.37	23.78	7.59	34.03	12.88	31.74	51.01
6	9.54	1.51	28.66	20.27	8.39	7.67	24.19	30.36	50.17
7	14.43	1.53	31.06	19.97	11.09	9.89	24.29	24.51	63.16
8	7.02	1.72	25.24	17.43	7.81	31.96	30.38	33.51	48.11
9	5.23	1.83	24.40	16.76	7.64	38.80	21.35	29.71	46.11
10	15.40	1.34	33.96	21.81	12.15	9.25	18.56	31.89	52.98
11	10.00	1.54	32.97	21.23	11.74	46.43	29.98	29.85	57.60
12	15.40	1.54	31.61	22.44	9.17	39.96	13.35	30.39	64.39
13	10.63	1.56	30.86	20.87	9.99	10.66	18.83	34.08	52.50
14	11.10	1.84	33.49	19.88	13.61	17.76	29.18	31.73	55.65

3.1. One-way Analysis of Variance (ANOVA) Analysis

The one-way ANOVA was conducted to evaluate whether there were significant differences in the means of the soil properties under investigation, including cohesion (C), angle of internal friction (Ø), dry density, liquid limit, plasticity index, fines %, W.C%, and plastic limit. The null hypothesis assumed that all means were equal, while the alternative hypothesis posited that not all means were equal, with a significance level set at $\alpha = 0.05$.

The results of the ANOVA analysis revealed a highly significant difference among the factors, as evidenced by the F-value of 90.99 and a $P = 0.000$. This indicates that the null hypothesis can be rejected, and it can be concluded that not all means are equal. The model summary further corroborated the strength of the analysis, with an R-squared value of 0.86, an adjusted R-squared value of 0.85, and a predictive R-squared value of 0.84. These metrics suggest that the model explains a large proportion of the variability in the data, indicating a strong overall fit.

The interval plot for cohesion (C) provided additional insights into the distribution of the data. The plot revealed that the mean cohesion value was 24.99 Ib/m^2, with a standard deviation of 13.55 Ib/m^2. The 95% confidence interval for the mean ranged from 21.73 Ib/m^2 to 28.26 Ib/m^2, indicating a moderate level of variability in the cohesion values across the soil samples.

3.2. Regression Analysis for Cohesion (C)

The regression analysis for cohesion (C) was conducted to identify the relationships between cohesion and several predictor variables, including dry density, liquid limit, plastic limit, W.C%, and fines %. The derived regression equation was:

$$\text{Cohesion} = -91 + 44.4\ \text{dry density} - 2.47\ \text{liquid limit} - 1.46\text{W.C} + 6.60\ \text{plastic limit} + 0.071\%\ \text{fines}.$$

The model summary for the cohesion regression indicated an R-squared value of 0.38, with an adjusted R-squared value of 0.20. These results suggest that the regression model has limited explanatory power, meaning that the predictors included in the model do not fully account for the observed variability in cohesion. This low R-squared value highlights the complexity of the relationships between cohesion and the predictor variables, which may not be fully captured by a linear regression model. Fig. 2 illustrates the scatter plot of soil cohesion (C) versus the predictor variable, along with the fitted regression line, demonstrating the correlation between the two and supporting the reliability of the derived regression model.

An examination of the coefficients revealed that dry density had the most significant positive effect on cohesion, as shown in (Fig. 3), with a coefficient of 44.4. This suggests that an increase in dry density leads to an increase in cohesion. Conversely, liquid limit and W.C% had negative coefficients (−2.47 and −1.46, respectively), indicating that increases in these variables are associated with decreases in cohesion. The plastic limit had a positive coefficient (6.60), suggesting that higher plastic limit values are associated with higher cohesion values. The % fines had a negligible positive effect on cohesion, with a coefficient of 0.071.

The ANOVA for the cohesion regression model revealed that the regression was not significant, with an F-value of 0.68 and a $P = 0.649$. This further supports the conclusion that the model has limited predictive power. The lack of significance in the regression model suggests that other factors such as permeability, compressibility, and specific gravity not

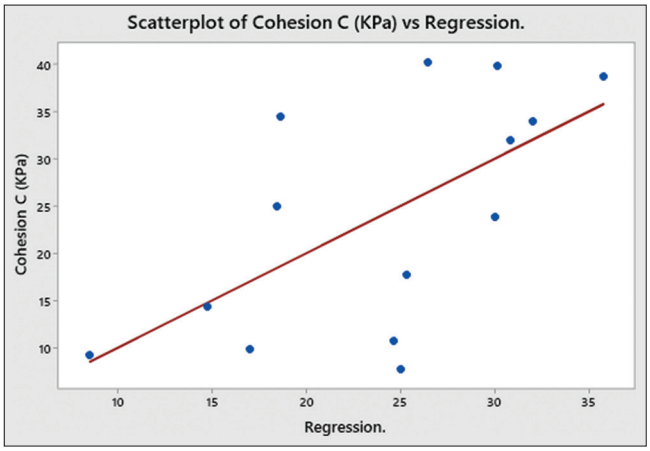


Fig. 2. Scatter plot of cohesion (C) versus predictor variable with regression line.

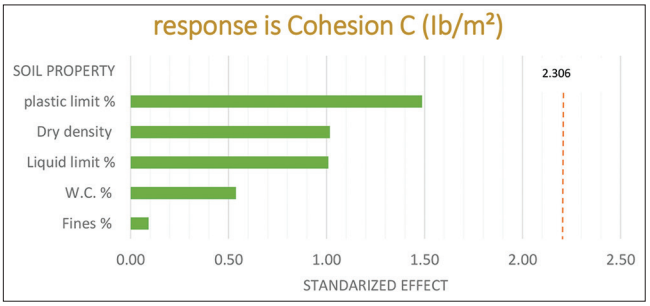


Fig. 3. Pareto chart of standardized effects for cohesion.

included in the analysis may play a more substantial role in determining cohesion.

3.3. Regression Analysis for Angle of Internal Friction (Ø)

The regression analysis for angle of internal friction (Ø) was conducted to explore the relationships between Ø and several predictor variables, including dry density, liquid limit, plastic limit, W.C%, and % fines. The derived regression equation was:

Angle of internal friction = 48.5–3.1 dry density + 1.216 liquid limit − 0.783 W.C − 2.58 plastic limit + 0.069% fines.

The model summary for the angle of internal friction regression indicated an R-squared value of 60.88%, with an adjusted R-squared value of 36.43% and a predictive R-squared value of 0.00%. While the R-squared value is higher than that of the cohesion regression, it still indicates that a significant portion of the variability in angle of internal friction is not explained by the predictors included in the model. Fig. 4 shows the scatter plot of the angle of internal friction (ϕ) against the predictor variable, including the regression line, which reflects the strength of the correlation and the suitability of the regression model for interpreting the data.

The coefficients for the predictors revealed that dry density had a negative effect on angle of internal friction, as shown in (Fig. 5), with a coefficient of −3.1, while liquid limit had a positive effect, with a coefficient of 1.216. W.C% and plastic limit had negative coefficients (−0.783 and −2.58, respectively), indicating that increases in these variables are associated with decreases in angle of internal friction. The % fines had a negligible positive effect on the angle of internal friction, with a coefficient of 0.069.

The ANOVA for the angle of internal friction regression model revealed that the regression was not significant, with an F-value of 2.49 and a $P = 0.12$. This suggests that the model has limited predictive power and that other factors not included in the analysis may influence the angle of internal friction (Table 5) and presents the complete regression analysis results for both cohesion (C) and the angle of internal friction (ϕ).

Regression models were used to estimate cohesion and the internal friction angle from simple soil properties – moisture, density, and plasticity. The results showed R^2 values around 0.38 for cohesion and about 0.61 for the friction angle. The predicted cohesion ranged from 10 to 20 kPa, and the internal

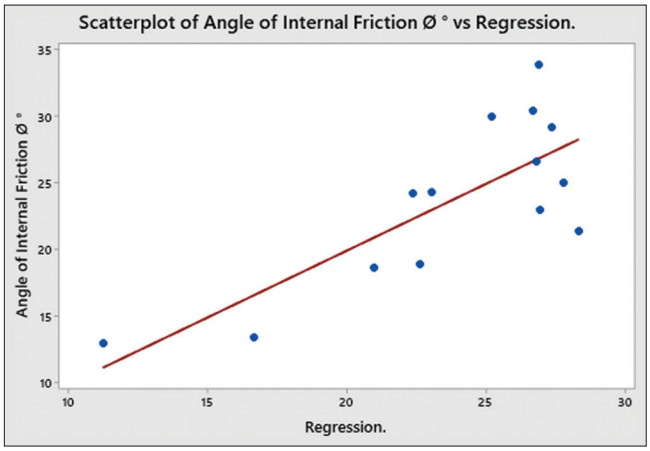


Fig. 4. Relationship between angle of internal friction (ϕ) and regression equation fit.

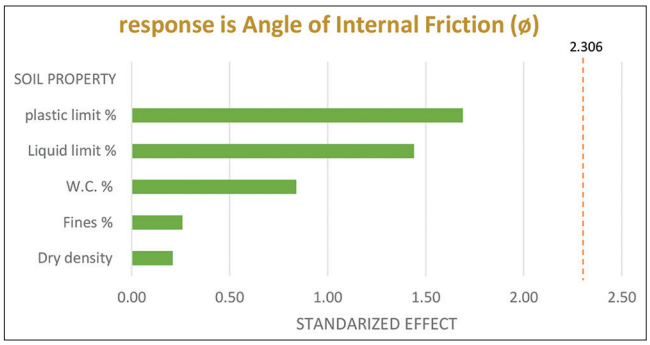


Fig. 5. Pareto chart of standardized effects for angle of internal friction.

TABLE 5: Metric results

No.	Term	For cohesion	For angle of internal friction
1	F-value	0.68	2.49
2	*P*-value	0.65	0.12
3	R-squared (R^2)	0.38	0.61
4	Adjusted R-squared	0.20	0.36

friction angles fell between 26° and 38°. These numbers are similar to what others have found, though some earlier studies like Khan and Wang (2021) reported slightly higher correlations, probably because of differences in the soils tested or data variability.

What makes this work stand out is that the parameters were directly estimated in the analysis based only on easy-to-measure soil properties, making it practical for quick assessments. While other studies show that these simple models can be very accurate (sometimes with R^2 values as high as 0.88), the results confirm that even with a bit more variability, the approach remains valid and useful, especially

in early planning or when quick decisions are needed.

Overall, these different studies all support the idea that basic soil properties are strongly related to SS parameters. The research adds a practical touch by providing simple equations that can be applied easily in the field or at early project stages, helping engineers and geotechnicians get reliable estimates without complex testing.

4. CONCLUSION

The analysis of the 14 soil samples provides a comprehensive understanding of the relationships between various soil properties, including cohesion (C), angle of internal friction (Ø), dry density, liquid limit, plasticity index, % fines, W.C%, and plastic limit %. The one-way ANOVA revealed significant differences among these soil properties, with a highly significant $P = 0.000$, indicating that the null hypothesis of equal means can be rejected. This highlights the inherent variability and complexity of soil behavior, which is critical for geotechnical engineering applications. The model summary further supported the strength of the analysis, with an R-squared value of 0.86, demonstrating a robust overall model fit.

The regression analysis for cohesion (C) and angle of internal friction (Ø) yielded mixed results. For cohesion, the regression equation identified several contributing factors, including dry density, liquid limit, W.C%, plastic limit %, and % fines. However, the low R-squared value of 0.38 and non-significant *P*-value (0.65) suggest that the model has limited explanatory power and predictive capability. This implies that the relationships between cohesion and the predictor variables are likely non-linear or influenced by factors not included in the analysis. Similarly, the regression model for angle of internal friction, with an R-squared value of 0.61, showed moderate explanatory power but also highlighted the complexity of the relationships between the predictors and the response variable.

The low R-squared values for both regression models underscore the limitations of linear regression in capturing the intricate interactions between soil properties. This raises important questions about the suitability of linear regression for predicting missing data or modeling soil behavior in general. The lack of significance in the regression models further emphasizes the need for alternative approaches, such as non-linear regression or machine learning techniques, to better account for the complexity of soil properties.

From a practical standpoint, the findings of this study have important implications for geotechnical engineering and soil management. Understanding the relationships between soil properties is essential for designing stable structures, predicting soil behavior under different conditions, and ensuring the safety and longevity of engineering projects. However, the limitations of the regression models highlight the need for caution when relying solely on statistical predictions for critical engineering decisions. Looking ahead, several steps can be taken to improve the accuracy and reliability of future analyses.

a. Expanding the dataset to include more soil samples would provide a more robust foundation for regression analysis
b. Exploring non-linear regression techniques or machine learning algorithms could better capture the complex relationships between soil properties
c. Incorporating additional variables, such as soil mineralogy, organic matter content, or environmental conditions, could enhance the explanatory power of the models.
d. Focusing on specific soil types or classifications could reduce variability and improve model fit
e. Validating statistical models with physical laboratory tests would provide a more comprehensive understanding of soil behavior and strengthen the practical applications of the findings.

While this study provides valuable insights into the relationships between soil properties, the limitations of the regression models highlight the need for a more nuanced approach to soil analysis. By addressing these limitations and refining future studies, researchers and engineers can develop more accurate and reliable models for predicting soil properties, ultimately leading to better-informed decisions in geotechnical engineering and soil management. This not only enhances the safety and efficiency of engineering projects but also contributes to the sustainable use of soil resources in various applications.

REFERENCES

[1] Y. A. Hamaamin, K. A. Rashed, Y. M. Ali and T. A. Abdalla. "Evaluation of ANFIS and regression techniques in estimating soil compression index for cohesive soils". *Journal of Engineering*, vol. 28, no. 10, pp. 28-41, 2022.

[2] N. F. Mustafa, S. F. Aziz, H. M. Ibrahim, K. Z. Abdulrahman, J. T. Abdalla and Y. A. Ahmad. "Double assessment of dam sites for sustainable hydrological management using GIS-fuzzy logic and ANFIS: Halabja water supply project case study". *Iranian Journal of Science and Technology Transactions of Civil Engineering*, vol. 49, pp. 2965-2983, 2024.

[3] Z. Yin, Y. Jin, J. S. Shen and P. Hicher. "Optimization techniques for identifying soil parameters in geotechnical engineering: Comparative study and enhancement". *International Journal for Numerical and Analytical Methods in Geomechanics*, vol. 42, no. 1, pp. 70-94, 2018.

[4] A. Krounis and F. Johansson. The Influence of Correlation between Cohesion and Friction Angle on the Probability of Failure for Sliding of Concrete Dams. In: "*3rd International Week on Risk Analysis, Dam Safety, Dam Security, and Critical Infrastructure Management, October 17-21 2011*". CRC Press, Valencia, Spain, pp. 75-80, 2012. Available from: https://www.diva/portal.org/smash/record.jsf?pid=diva2:490455 [Last accessed on 2025 Apr 16].

[5] J. G. Egbe, D. E. Ewa, S. E. Ubi, G. B. Ikwa and O. O. Tumenayo. "Application of multilinear regression analysis in modeling of soil properties for geotechnical civil engineering works in Calabar South". *Nigerian Journal of Technology*, vol. 36, no. 4, pp. 1059-1065, 2017.

[6] L. K. Sharma and T. N. Singh. "Regression-based models for the prediction of unconfined compressive strength of artificially structured soil". *Engineering with Computers*, vol. 34, no. 1, pp. 175-186, 2017.

[7] K. K. Phoon, W. A. Prakoso, Y. Wang and J. Ching. *Uncertainty Representation of Geotechnical Design Parameters*. In: "*Reliability of Geotechnical Structures in ISO2394*". CRC Press, pp. 67-106, 2016. Available from: https://www.taylorfrancis.com/chapters/edit/10.1201/9781315364179-11/uncertainty/representation/geotechnical/design/parameters/kok/kwang/phoon/widjojo/prakoso/yu/wang/jianye-ching [Last accessed on 2025 Apr 01].

[8] S. Pul, A. Ghaffari, E. Öztekin, M. Hüsem and S. Demir. "Experimental determination of cohesion and internal friction angle on conventional concretes". *ACI Materials Journal,* vol. 114, no. 3, pp. 407-416, 2017.

[9] M. M. Nawaz, S. R. Khan, R. Farooq, M. N. Nawaz, J. Khan, M. A. U. R. Tariq, R. F. Tufail, R. Farooq and A. W. M. Ng. "Development of a cost-based design model for spread footings in cohesive soils". *Sustainability*, vol. 14, no. 9, p. 5699, 2022.

[10] H. Do, S. Mirarab, L. Tahvildari and G. Rothermel. An empirical study of the effect of time constraints on the cost-benefits of regression testing. In: "*Proceedings of the 16th ACM SIGSOFT International Symposium on Foundations of Software Engineering*". ACM, Atlanta Georgia, pp. 71-82, 2008.

[11] P. Abbasimaedeh. "Soil liquefaction in seismic events: Pioneering predictive models using machine learning and advanced regression techniques". *Environmental Earth Sciences*, vol. 83, no. 7, p. 189, 2024.

[12] D. Basu, A. Misra and A. J. Puppala. "Sustainability and geotechnical engineering: Perspectives and review". *Canadian Geotechnical Journal*, vol. 52, no. 1, pp. 96-113, 2015.

[13] A. A. Firoozi and A. A. Firoozi. Soil-Structure Interaction: Understanding and Mitigating Challenges. In: "*Challenges in Foundation Engineering-Case Studies and Best Practices*". IntechOpen, 2023. Available from: https://www.intechopen.com/chapters/87761 [Last accessed on 2025 Apr 27].

[14] M. Mohammadi, S. M. F. Aghda, M. Talkhablou and A. Cheshomi. "Prediction of the shear strength parameters from easily-available soil properties by means of multivariate regression and artificial neural network methods". *Geomechanics and Geoengineering*, vol. 17, no. 2, pp. 442-454, 2020.

[15] M. N. Nawaz, B. Alshameri, Z. Maqsood and W. Hassan. "Predictive modelling of cohesion and friction angle of soil using gene expression programming: A step towards smart and sustainable construction". *Neural Computing and Applications*, vol. 36, no. 18, pp. 10545-10566, 2024.

[16] M. I. Khan and S. Wang. "Slope stability analysis to develop correlations between different soil parameters and factor of safety using regression analysis". *Polish Journal of Environmental Studies*, vol. 30, pp. 4021-4030, 2021.

[17] M. Erbilen. "*Imputing Missing Values using Support Variables with Application to Barley Grain Yield*". 2019. Available from: https://i/rep.emu.edu.tr/xmlui/handle/11129/5338 [Last accessed on 2025 Apr 01].

[18] K. A. Rashed. "Study of raw materials for manufacturing perforated clay bricks (Chamchamal, and Qaradagh regions in Sulaimani/Iraq)". *Sulaimani Journal for Engineering Sciences*, vol. 4, no. 4, pp. 50-66, 2017.

[19] M. R. Carter and E. G. Gregorich. "*Soil Sampling and Methods of Analysis*". CRC Press, 2007. Available from: https://www.taylorfrancis.com/books/mono/10.1201/9781420005271/soil/sampling/methods-analysis-gregorich-carter [Last accessed on 2025 Apr 14].

[20] D. Pennock, T. Yates and J. Braidek. Soil sampling designs. In: "*Soil Sampling and Methods of Analysis*". Canadian Society of Soil Science, Ontario, pp. 1-14, 2007.

[21] O. Fernández-Ugalde, A. Jones and R. G. Meuli. "Comparison of sampling with a spade and gouge auger for topsoil monitoring at the continental scale". *European Journal of Soil Science*, vol. 71, no. 2, pp. 137-150, 2020.

[22] T. Stark, S. Wilk and R. Swan. Sampling, Reconstituting, and Gradation Testing of Railroad Ballast In: "*Symposium on Railroad Ballast Testing and Properties*". ASTM International, pp. 135-143, 2018. Available from: https://asmedigitalcollection.asme.org/astm/ebooks/book/chapter/pdf/7398285/10_1520_stp160520170026.pdf [Last accessed on 2025 Apr 25].

[23] J. R. Boulding. "*Description and Sampling of Contaminated Soils: A Field Guide*". CRC Press, 2017. Available from: https://www.taylorfrancis.com/books/mono/10.1201/9781315140926/description/sampling-contaminated-soils-russell-boulding [Last accessed on 2025 Apr 27].

[24] A. R. Aquib, M. M. Uzzaman and M. F. Alam. "*Determining Engineering Properties of Fine-Grained Soils using Atterberg Limits*". Available from: https://www.researchgate.net/profile/ashkar/aquib/publication/372941330_determining_engineering_properties_of_fine/grained_soils_using_atterberg_limits/links/64d5f45bb684851d3d9e639e/determining/engineering/properties/of/fine/grained/soils/using/atterberg/limits.pdf [Last accessed on 2025 Apr 01].

[25] D. C. Montgomery and G. C. Runger. "*Applied Statistics and Probability for Engineers*". John Wiley and Sons, 2019. Available from: https://books.google.com/books?hl=en&lr=&id=nuo9eaaaqbaj&oi=fnd&pg=pa1&dq=montgomery,+d.+c.,+%26+runger,+g.+c.+(2010).+applied+statistics+and+probability+for+engineers.+wiley.&ots=jnbaw/rhip&sig=kwcr02cmpfzezjylw2pydnmdi2o [Last accessed on 2025 Apr 01].

[26] H. I. Okagbue, P. E. Oguntunde, E. C. Obasi and E. M. Akhmetshin. "Trends and usage pattern of SPSS and minitab software in scientific research". *Journal of Physics Conference Series*, vol. 1734, p. 012017, 2021.

[27] S. A. Lesik. "*Applied Statistical Inference with MINITAB®*". Chapman and Hall/CRC, 2018. Available from: https://www.taylorfrancis.com/books/mono/10.1201/9780429444951/applied/statistical/inference-minitab%c2%ae-second-edition-sally-lesik [Last accessed on 2025 Apr 16].

[28] R. S. Kenett and S. Zacks. "*Modern Industrial Statistics: With Applications in R, MINITAB, and JMP*". John Wiley and Sons,

2021. Available from: https://books.google.com/books?hl=en&lr=&id=iyeqeaaaqbaj&oi=fnd&pg=pr17&dq=minitab+y+line&ots=-cpvzlxfkx&sig=jqmw5a-6-rxp9ozynvftajokgia [Last accessed on 2025 Apr 27].

[29] W. C. Rinaman. Minitab. In: "*Revival: The Handbook of Software for Engineers and Scientists (1995)*". CRC Press, pp. 1028-1046, 2018. Available from: https://www.taylorfrancis.com/chapters/edit/10.1201/9780203710012/47/minitab/william/rinaman [Last accessed on 2025 Apr 16].

[30] E. Fragkiadaki, G. Evangelidis, N. Samaras and D. A. Dervos. "F-Value: Measuring an article's scientific impact". *Scientometrics*, vol. 86, no. 3, pp. 671-686, 2011.

[31] M. S. Thiese, B. Ronna and U. Ott. "P value interpretations and considerations". *Journal of Thoracic Disease*, vol. 8, no. 9, p. E928, 2016.

[32] J. Gao. "R-squared (R^2) - how much variation is explained". *Research Methods in Medicine and Health Sciences*, vol. 5, no. 4, pp. 104-109, 2024.

[33] J. Karch. "Improving on adjusted R-squared". *Collabra Psychology*, vol. 6, no. 1, p. 45, 2020.

[34] K. A. Hart, B. A. Steinfeldt and R. D. Braun. Formulation and applications of a probabilistic Pareto chart. In: *56th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*. American Institute of Aeronautics and Astronautics, Kissimmee, Florida, 2015.

[35] P. Nieminen. "Application of standardized regression coefficient in meta-analysis". *BioMedInformatics*, vol. 2, no. 3, pp. 434-458, 2022.

[36] J. Hur and M. C. Jung. "The effects of soil properties on the turbidity of catchment soils from the Yongdam dam basin in Korea". *Environmental Geochemistry and Health*, vol. 31, no. 3, pp. 365-377, 2009.

[37] R. B. J. Brinkgreve. Selection of soil models and parameters for geotechnical engineering application. In: "*Soil Constitutive Models*". American Society of Civil Engineers, Austin, Texas, United States, pp. 69-98, 2005.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

A Novel Class of Hybrid Conjugate Gradient Methods for Unconstrained Optimization with Applications to Image Denoising

Rega Jalal[1], Hawraz N. Jabbar[2]

[1]College of Science, Department of Mathematics, University of Sulaimani, Sulaymaniyah, 46001, Kurdistan Region of Iraq, [2] College of Science, Department of Mathematics, University of Kirkuk, Kirkuk, 36013, Iraq.

ABSTRACT

This paper presents a new hybrid conjugate gradient (CG) method for solving large-scale unconstrained optimization problems where classical CG algorithms may not perform well. This new method integrates four classical algorithms, namely, Liu and Storey, Fletcher and Reeves, Dai and Yuan, and Polak, Ribiere, and Polyak, using a convex combination of CGs and an inexact line search based on strong Wolfe conditions, and adheres to the Dai-Liao conjugate condition to enhance convergence properties. The theoretical study established the conditions for sufficient descent and global convergence of the hybrid CG method. Numerical studies demonstrated that the proposed method significantly reduced the number of iterations and computational time, achieving superior evaluation efficiency compared with other CG methods. In particular, the effectiveness of the proposed algorithm is validated for image impulse denoising, where it can recover high-quality images while preserving significant features, implying its practical applicability to real-world signal and image processing problems.

Index Terms: Conjugate Gradient Method, Strong Wolfe Condition, Global Convergence, Unconstrained Optimization, Convex Combination, Hybrid Method, Image Processing

1. INTRODUCTION

Conjugate gradient (CG) methods are a category of unconstrained optimization algorithms that are distinguished by their low memory requirements and robust local and global convergence characteristics [1]. The main objective of the CG method was to solve a linear system, Hestenes and Stiefel [2]. Later, the method was modified to address an unconstrained optimization problem because solving a linear system is equivalent to minimizing a positive definite quadratic function [3].

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp166-183

E-ISSN: 2521-4217

P-ISSN: 2521-4209

Copyright © 2025 Jalal R, Jabbar HN. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

CG methods have significantly contributed to solving large-scale non-linear optimization problems. The CG method is often used to solve the following unconstrained optimization problem:

$$\min_{x\in\mathbb{R}^n} f(x) \tag{1}$$

where $f:\mathbb{R}^n \to \mathbb{R}$ is a smooth non-linear function with an available gradient.

The iterative formula of the CG method is expressed as follows,

$$x_0 \in \mathbb{R}^n$$

$$x_{(k+1)} = x_k + a_k d_k, k = 0,1,\ldots n \tag{2}$$

and

Corresponding author's e-mail: Rega Jalal, College of Science, Department of Mathematics, University of Sulaimani, Sulaymaniyah, 46001, Kurdistan Region of Iraq. E-mail: rega.awrahman@univisul.edu.iq

Received: 03-07-2025 **Accepted:** 16-08-2025 Published: 27-09-2025

$$d_k = \begin{cases} -g_k, & if\ k = 0, \\ -g_k + \beta_k d_{k-1}, & if\ k \geq 1. \end{cases} \quad (3)$$

In this context, x_0 is the initial point, and x_k represents the present iteration. The gradient of f at point x_k is given by $g_k = \nabla f(x_k)$. The search direction is denoted by d_k, and the conjugate (update) parameter. $\beta_k \in \mathbb{R}$ determines various iterations of the CG method. In addition, the step size, which can be achieved using numerous line search methods, is represented by a positive value $a_k > 0$.

In this study, we use an inexact line search called the strong Wolfe line search, which has garnered significant focus in convergence analysis and the execution of CG methods and is characterized by the following requirements [4,5].

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k \quad (4)$$

$$\left| g_{k+1}^T d_k \right| \leq -\sigma g_k^T d_k \quad (5)$$

where δ, σ are scalars and satisfy $0 < \delta \leq \sigma < 0.5$.

Assume that $\|.\|$ is *Euclideannorm* and $y_k = g_{k+1} - g_k$. If f is a strongly convex quadratic, then theoretically, all six options for the update parameter in Table 1 are equal to an exact line search [12]. However, they exhibit distinct behavior for general non-quadratic functions with inexact line search. Various CG methods correspond to different values based on the scalar parameter β_k. The methods can be classified into two groups.

TABLE 1: Choices of β_k in standard conjugate gradient methods.

Formula	Description
$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}$	(1952) Hestenes and Stiefel linear (CG) [2]
$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}$	(1964) Fletcher and Reeves non-linear (CG) [6]
$\beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2}$	(1969) Polak and Ribiere [7], Polyak [8]
$\beta_k^{CD} = \frac{\|g_{k+1}{}^2\|}{-g_k^T d_k}$	(1987) Conjugate Descent (CD) by Fletcher [9]
$\beta_k^{LS} = \frac{g_{k+1}^T y_k}{-d_k^T g_k}$	(1991) Liu and Storey [10]
$\beta_k^{DY} = \frac{\|g_{k+1}\|^2}{y_k^T d_k}$	(1999) Dai and Yuan [11]

In the first group, all CG parameters have the same numerator $g_{k+1}^T y_k$. The methods have been proposed by Hestenes and Stiefel (HS), Polak, Ribiere, and Polyak (PRP), and Liu and Storey (LS) using the following CG parameters [1].

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{y_k^T d_k}, \beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2}, \beta_k^{LS} = \frac{g_{k+1}^T y_k}{-d_k^T g_k}.$$

In the second group, $\|g_{k+1}\|^2$ is the common numerator of all the CG parameters. The following CG parameters were proposed by Fletcher and Reeves (FR), Fletcher (conjugate descent or CD), and Dai and Yuan (DY), and include the following CG parameters [1].

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \beta_k^{CD} = \frac{\|g_{k+1}\|^2}{-g_k^T d_k}, \beta_k^{DY} = \frac{\|g_{k+1}\|^2}{y_k^T d_k}$$

Each CG method has distinct advantages and disadvantages. The second group of FR, CD, and DY methods with $\|g_{k+1}\|^2$ in the numerator has significantly worse numerical performance than the first group of HS, PRP, and LS algorithms [13]. The poor numerical performance of the second group can be attributed to the jamming phenomenon, whereby the algorithm may execute several brief iterations without achieving much advancement toward the minimum [13]. Nonetheless, the first group HS, PRP, and LS with $g_{k+1}^T y_k$ in the numerator inherently has an automated approximation restart mechanism that prevents jamming. Specifically, when the step α_k is small, the component y_k in the numerator approaches zero. Consequently, β_k decreases, and the new search direction d_{k+1} approximates the steepest descent direction. However, Powell [14] developed a three-dimensional counterexample demonstrating that the PRP and HS methods may cycle indefinitely without a solution, indicating their lack of global convergence under specific conditions.

The FR, CD, and DY methods generally exhibit strong global convergence features [15], although their computational performance may be limited [13]. Simultaneously, the HS, PRP, and LS methods can fail to consistently converge, even though they often exhibit superior computing efficiency [13].

In Hager and Zhang [1] asserts that a comprehensive review of several CG approaches was presented.

The most important category of CG algorithms is hybrid CG algorithms. The goal of this study was to improve the computing speed and maintain robust global convergence by combining four types of CG methods. Hybrid CG methods include switching a parameter in the second group to the first group when iterations are stuck [16]. Recently, some hybrid CG methods have been proposed. Andrei proposed the following hybrid CG method: $\beta_k^{hyb} = (1-\theta_k)\beta_k^{HS} + \theta_k\beta_k^{DY}$ [17]. Djordjević proposed following hybrid CG methods: $\beta_k^{\text{hyb}} = (1-\theta_k)\beta_k^{\text{PRP}} + \theta_k\beta_k^{\text{FR}}, \beta_k^{\text{hyb}} = (1-\theta_k)\beta_k^{\text{LS}} + \theta_k\beta_k^{\text{FR}}$ [18,19]. Babando *et al.* proposed this hybrid method: $\beta_k^{hyb} = (1-\theta_k)\beta_k^{PRP} + \theta_k\beta_k^{DY}$ [3]. Liu and Li introduced a new hybrid approach that combines the two methods: $\beta_k^{hyb} = (1-\theta_k)\beta_k^{\text{LS}} + \theta_k\beta_k^{\text{DY}}$ [20]. Hanachi *et al.* presented a hybrid approach that is a combination of three CG methods, as described in [21] and [22].

$$\beta_k^{\text{New}} = \gamma_k\beta_k^{\text{FR}} + \eta_k\beta_k^{\text{DY}} + (1-\gamma_k-\eta_k)\beta_k^{\text{PRP}},$$
$$\beta_k^{\text{New}} = \delta_k\beta_k^{\text{LS}} + \gamma_k\beta_k^{\text{FR}} + (1-\delta_k-\gamma_k)\beta_k^{\text{PRP}}.$$

Inspired by the research of Hanachi *et al.* [21] and [22], we propose a new hybrid CG method that combines the LS, FR, DY, and PRP methods to solve problem (1). This study represents the first research effort to propose and analyze a hybrid CG method that utilizes a convex combination of four classical CG methods. To the best of our knowledge, no previous study has explored such a combination for developing of CG algorithms. The remainder of this paper is organized as follows: In section 2, we introduce the hybrid CG approach and demonstrate how we obtained the parameters ζ_k,ξ_k and ω_k using various methods. Under moderate circumstances, we show that the chosen method, when used with a strong Wolfe line search, provides directions that satisfy the sufficient descent condition. The algorithm is described in section 3. In section 4, we examine the convergence properties of the proposed method. Through comprehensive numerical comparisons with several existing hybrid methods in the literature using 170 distinct test problems, demonstrating the improved efficiency of the proposed method in section 5, was also validated for an application to image impulse denoising in section 6. Section 7 concludes with a brief summary.

2. HYBRID CG ALGORITHM

In this section, to enhance the performance of the CG updating parameter proposed by Hestenes and Stiefel [2], Fletcher and Reeves [6], Dai and Yuan [11], Polak and Ribiere [7] and Polyak [8], CG methods, we employed their convex combination, in which the parameter β_k in the proposed method, referred to as β_k^{hRH}, i.e.,

$$\beta_k^{\text{hRH}} = \zeta_k\beta_k^{\text{LS}} + \xi_k\beta_k^{\text{FR}} + \omega_k\beta_k^{\text{DY}} + (1-\zeta_k-\xi_k-\omega_k)\beta_k^{\text{PRP}} \tag{6}$$

The parameters ζ_k,ξ_k,ω_k in equation (6) that meet the condition $0\leq\zeta_k$, ξ_k, $\omega_k\leq 1$ are referred to as hybridization parameters. These parameters are obtained in a specific manner, which will be discussed in detail later and the search direction d_k is computed from

$$d_k = \begin{cases} -g_k, & \text{if } k=0, \\ -g_k + \beta_k^{hRH} d_{k-1}, & \text{if } k\geq 1. \end{cases} \tag{7}$$

To determine the step length a_k, we utilize the strong Wolfe line search (4), (5).

There are 15 separate cases depending on the parameter values where $0\leq\zeta_k$, ξ_k, $\omega_k\leq 1$ and $\zeta_k+\xi_k+\omega_k\leq 1$.

Case 1: If $\zeta_k=1$, $\xi_k=0$, $\omega_k=0$, then $\beta_k^{\text{hRH}} = \beta_k^{\text{LS}}$

Case 2: If $\zeta_k=0$, $\xi_k=0$, $\omega_k=1$, then $\beta_k^{\text{hRH}} = \beta_k^{\text{FR}}$

Case 3: If $\zeta_k=0$, $\xi_k=0$, $\omega_k=1$, then $\beta_k^{\text{hRH}} = \beta_k^{\text{DY}}$

Case 4: If $\zeta_k=0$, $\xi_k=0$, $\omega_k=0$, then $\beta_k^{\text{hRH}} = \beta_k^{\text{PRP}}$

Case 5: If $0<\zeta_k<1$, $\xi_k=0$, $\omega_k=0$, then $\beta_k^{\text{hRH}} = \zeta_k\beta_k^{\text{LS}} + (1-\zeta_k)\beta_k^{\text{PRP}}$

Case 6: If $\zeta_k=0$, $0<\xi_k<1$, $\omega_k=0$, then $\beta_k^{\text{hRH}} = \xi_k\beta_k^{\text{FR}} + (1-\xi_k)\beta_k^{\text{PRP}}$

Case 7: If $\zeta_k=0$, $\xi_k=0$, $0<\omega_k<1$, then $\beta_k^{\text{hRH}} = \omega_k\beta_k^{\text{DY}} + (1-\omega_k)\beta_k^{\text{PRP}}$

Case 8: If $0<\zeta_k<1$, $0<\xi_k<1$, $\omega_k=0$ and $\zeta_k+\xi_k=1$, then $\beta_k^{\text{hRH}} = \zeta_k\beta_k^{\text{LS}} + (1-\zeta_k)\beta_k^{\text{FR}}$

Case 9: If $0<\zeta_k<1$, $0<\xi_k<1$, $\omega_k=0$ and $0<\zeta_k+\xi_k<1$, then
$\beta_k^{\mathrm{hRH}} = \zeta_k\beta_k^{\mathrm{LS}} + \xi_k\beta_k^{\mathrm{FR}} + \left(1-\zeta_k-\xi_k\right)\beta_k^{\mathrm{PRP}}$

Case 10: If $0<\zeta_k<1$, $\xi_k=0$, $0<\omega_k<1$ and $\zeta_k+\omega_k=1$, then
$\beta_k^{\mathrm{hRH}} = \zeta_k\beta_k^{\mathrm{LS}} + \left(1-\zeta_k\right)\beta_k^{\mathrm{DY}}$

Case 11: If $0<\zeta_k<1$, $\xi_k=0$, $0<\omega_k<1$ and $0<\zeta_k+\omega_k<1$, then
$\beta_k^{\mathrm{hRH}} = \zeta_k\beta_k^{\mathrm{LS}} + \omega_k\beta_k^{\mathrm{DY}} + \left(1-\zeta_k-\omega_k\right)\beta_k^{\mathrm{PRP}}$

Case 12: If $\zeta_k=0$, $0<\xi_k<1$, $0<\omega_k<1$ and $<\xi_k+\omega_k=1$, then
$\beta_k^{\mathrm{hRH}} = \xi_k\beta_k^{\mathrm{FR}} + \left(1-\xi_k\right)\beta_k^{\mathrm{DY}}$

Case 13: If $\zeta_k=0$, $0<\xi_k<1$, $0<\omega_k<1$ and $0<\xi_k+\omega_k<1$, then
$\beta_k^{\mathrm{hRH}} = \xi_k\beta_k^{\mathrm{FR}} + \omega_k\beta_k^{\mathrm{DY}} + \left(1-\xi_k-\omega_k\right)\beta_k^{\mathrm{PRP}}$

Case 14: If $0<\zeta_k<1$, $0<\xi_k<1$, $0<\omega_k<1$ and $\zeta_k+\xi_k+\omega_k=1$, then
$\beta_k^{\mathrm{hRH}} = \zeta_k\beta_k^{\mathrm{LS}} + \xi_k\beta_k^{\mathrm{FR}} + \omega_k\beta_k^{\mathrm{DY}}$

Case 15: If $0<\zeta_k<1$, $0<\xi_k<1$, $0<\omega_k<1$ and $0<\zeta_k+\xi_k+\omega_k<1$, then
$\beta_k^{\mathrm{hRH}} = \zeta_k\beta_k^{\mathrm{LS}} + \xi_k\beta_k^{\mathrm{FR}} + \omega_k\beta_k^{\mathrm{DY}} + \left(1-\zeta_k-\xi_k-\omega_k\right)\beta_k^{\mathrm{PRP}}$

It is clear from equations (6) and (7) that

$$d_k = \begin{cases} -\mathrm{g}_k, & \text{if } k=0 \\ -\mathrm{g}_k + \zeta_k \dfrac{\mathrm{g}_k^T \mathrm{y}_{k-1}}{-\mathrm{d}_{k-1}^T g_{\mathrm{k-1}}}\mathrm{d}_{k-1} \\ +\xi_k \dfrac{\|\mathrm{g}_k\|^2}{\|\mathrm{g}_{k-1}\|^2}\mathrm{d}_{k-1} + \omega_k \dfrac{\|\mathrm{g}_k\|^2}{\mathrm{y}_{k-1}^T \mathrm{d}_{k-1}}\mathrm{d}_{k-1} \\ +\left(1-\zeta_k-\xi_k-\omega_k\right)\dfrac{\mathrm{g}_k^T \mathrm{y}_{k-1}}{\|\mathrm{g}_{k-1}\|^2}\mathrm{d}_{k-1}, & \text{if } k\geq 1 \end{cases} \quad (8)$$

For both numerical computation and convergence analysis, the traditional conjugacy condition is crucial in the CG method. This condition is based on an exact line search to obtain the step size a_k; however, in practical computations, where an inexact line search is commonly utilized, the condition may not be strictly satisfied.

$$\mathrm{d}_{k+1}^T \mathrm{y}_k = 0 \quad (9)$$

In 1978, Perry [25] generalized the conjugacy condition and incorporated ideas from quasi-Newton methods. In particular, Perry supplemented the conjugacy condition by adding modifications to the standard conjugacy condition. The secant equation was incorporated into this formulation. Perry's extension enables the CG method to be brought in line with the quasi-Newton direction, is flexible, and is more robust under an exact line search.

$$\mathrm{d}_{k+1}^T \mathrm{y}_k = -\mathrm{g}_{k+1}^T \mathrm{s}_k \quad (10)$$

However, it is still contingent on the exact line search and is therefore not particularly effective in certain numerical optimization procedures. In addition, inexact line searches rather than exact line searches are used to determine the step size a_k.

Therefore, in 2001, Dai and Liao [26] suggested a less stringent conjugacy condition called the D-L conjugacy condition:

$$\mathrm{d}_{k+1}^T \mathrm{y}_k = -t\mathrm{g}_{k+1}^T \mathrm{s}_k,\ t \geq 0 \quad (11)$$

Where $\mathrm{s}_k = \mathrm{x}_{k+1} - \mathrm{x}_k$, and t is a non-negative modulating parameter. The selection of an optimal value for t remains an active area of research, with several potential choices proposed in [27-35]. Notably, for $t=0$, equation (11) simplifies to equation (9), while for $t=1$, it corresponds to equation (10). To select parameter t, this study employs $t = \dfrac{\mathrm{s}_k^T \mathrm{y}_k}{\mathrm{s}_k^2} + \dfrac{\mathrm{y}_k}{\mathrm{s}_k}$ as specified in [30]. The parameters ζ_k, ξ_k, and ω_k are chosen by applying the D-L conjugacy condition, as given in equation (11). Taking the inner product of equation (3) with the vector y_k^T and calculating β_k using equation (6) yields

$$\mathrm{d}_{k+1}^T \mathrm{y}_k = -\mathrm{g}_{k+1}^T \mathrm{y}_k + \beta_k^{hRH}\left(\mathrm{d}_k^T \mathrm{y}_k\right)$$

By enforcing the D-L condition $\mathrm{d}_{k+1}^T \mathrm{y}_k = -t\mathrm{g}_{k+1}^T \mathrm{s}_k$, we obtain

$$-t\mathrm{g}_{k+1}^T \mathrm{s}_k = -\mathrm{g}_{k+1}^T \mathrm{y}_k + \begin{pmatrix} \zeta_k\beta_k^{LS} + \xi_k\beta_k^{FR} + \omega_k\beta_k^{DY} \\ +\left(1-\zeta_k-\xi_k-\omega_k\right)\beta_k^{PRP} \end{pmatrix}\left(\mathrm{d}_k^T \mathrm{y}_k\right)$$

To simplify the presentation, we define the following terms:

$$A_{LS} = \beta_k^{LS}\left(\mathrm{d}_k^T \mathrm{y}_k\right) = \frac{\mathrm{g}_{k+1}^T \mathrm{y}_k}{-\mathrm{g}_k^T \mathrm{d}_k}\left(\mathrm{d}_k^T \mathrm{y}_k\right) \quad (12)$$

$$A_{FR} = \beta_k^{FR}\left(d_k^T y_k\right) = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}\left(d_k^T y_k\right) \quad (13)$$

$$A_{DY} = \beta_k^{DY}\left(d_k^T y_k\right) = \frac{\|g_{k+1}\|^2}{d_k^T y_k}\left(d_k^T y_k\right) = \|g_{k+1}\|^2 \quad (14)$$

$$A_{PRP} = \beta_k^{PRP}\left(d_k^T y_k\right) = \frac{g_{k+1}^T y_k}{\|g_k\|^2}\left(d_k^T y_k\right) \quad (15)$$

Rearranging the equation to solve for the hybridization parameters yields

$$-tg_{k+1}^T s_k + g_{k+1}^T y_k - A_{PRP} = \zeta_k\left(A_{LS} - A_{PRP}\right) + \xi_k\left(A_{FR} - A_{PRP}\right) + \omega_k\left(A_{DY} - A_{PRP}\right)$$

This is an underdetermined linear equation in ζ_k, ξ_k, ω_k. For our algorithm, we propose a strategy in which we compute one parameter based on the values of the other two (which can be set to defaults or previous values). This yields the following update rule:

$$\zeta_k = \frac{\left(-tg_{k+1}^T s_k + g_{k+1}^T y_k - A_{PRP}\right) - \xi_k\left(A_{FR} - A_{PRP}\right) - \omega_k\left(A_{DY} - A_{PRP}\right)}{A_{LS} - A_{PRP}} \quad (16)$$

$$\xi_k = \frac{\left(-tg_{k+1}^T s_k + g_{k+1}^T y_k - A_{PRP}\right) - \zeta_k\left(A_{LS} - A_{PRP}\right) - \omega_k\left(A_{DY} - A_{PRP}\right)}{A_{FR} - A_{PRP}} \quad (17)$$

$$\omega_k = \frac{\left(-tg_{k+1}^T s_k + g_{k+1}^T y_k - A_{PRP}\right) - \zeta_k\left(A_{LS} - A_{PRP}\right) - \xi_k\left(A_{FR} - A_{PRP}\right)}{A_{DY} - A_{PRP}} \quad (18)$$

Note that while these formulas appear simpler, substituting the full expressions for A_{LS}, A_{FR}, A_{DY}, and A_{PRP} would recover more detailed forms of the equations. However, this formulation makes the underlying structure of the linear system explicit. In our algorithm, care is taken to handle cases in which any denominator is close to zero.

The parameters ζ_k, ξ_k, and ω_k. may lie beyond the interval [0, 1]. The following rule must be applied to obtain a valid convex combination in equation (6): If $\zeta_k, \xi_k, \omega_k \leq 0$, then set $\zeta_k, \xi_k, \omega_k = 0$ in (6). If $\zeta_k, \xi_k, \omega_k \geq 1$, then we assign $\zeta_k, \xi_k, \omega_k = 1$ in (6). If $\zeta_k + \xi_k + \omega_k \geq 1$, then we change $\zeta_k, \xi_k, \omega_k = 1$ in (6).

Algorithm 1: hybrid Rega and Hawraz Algorithm

Require: Initial point $x_0 \in R^n$, parameters δ, σ with $0 < \delta \leq \sigma < 0.5$, tolerance: $\epsilon = 10^{-6}$

Initialization: Set $k = 0$, compute $f(x_0)$, $g_0 = \nabla f(x_0)$. If $||g_0|| < \epsilon$, stop.

Set: $d_0 = -g_0$.

while $||g_k|| \geq \epsilon$ do

Line Search: Find step size $a_k > 0$ satisfying strong Wolfe conditions (4) and (5).

Update variables:

$x_{k+1} = x_k + a_k d_k$; $g_{\{k+1\}} = \nabla f(x_{k+1})$

$s_k = x_{k+1} - x_k$; $y_k = g_{\{k+1\}} - g_k$.

Compute hybridization parameters:

Compute t: Set $t = \frac{s_k^T y_k}{\|s_k\|^2} + \frac{\|y_k\|}{||s_k||}$

Compute $A_{LS}, A_{FR}, A_{DY}, A_{PRP}$ using equations (12)-(15).

Compute initial values for ζ_k, ξ_k, ω_k using equations (16)-(18).

If any denominator is near zero, the corresponding parameter is set to zero.

Enforce convex combination constraints: Clamp each parameter to the interval [0, 1]:

$\zeta_k \leftarrow \max(0, \min(1, \zeta_k))$

$\xi_k \leftarrow \max(0, \min(1, \xi_k))$

$\omega_k \leftarrow \max(0, \min(1, \omega_k))$

if $\zeta_k + \xi_k + \omega_k > 1$ then

Normalize the sum to 1:

$$Sum \leftarrow \zeta_k + \xi_k + \omega_k;\ \zeta_k \leftarrow \frac{\zeta_k}{Sum};\ \xi_k \leftarrow \frac{\xi_k}{Sum};\ \omega_k \leftarrow \frac{\omega_k}{Sum}.$$

end if

Compute conjugate parameter $\boldsymbol{\beta_k^{hRH}}$

Calculate

$$\beta_k^{hRH} = \zeta_k \beta_k^{LS} + \xi_k \beta_k^{FR} + \omega_k \beta_k^{DY} + \left(1-\zeta_k-\xi_k-\omega_k\right)\beta_k^{PRP}$$

Update search direction: Set $d_{k+1} = -g_{k+1} + \beta_k^{hRH} d_k$

Restart condition (Powell):

if $\left|g_{k+1}^T g_k\right| \geq 0.2\left\|g_{k+1}\right\|^2$ then

$d_{k+1} \leftarrow -g_{k+1}$

end if

Iteration Update: Set $k \leftarrow k+1$

end while

3. SUFFICIENT DESCENT CONDITION

In this section, we demonstrate the sufficient descent of the proposed method, which is crucial for examining global convergence. These fundamental assumptions regarding the objective function are required to demonstrate the global convergence of the hRH method.

Assumption 1. The level set $\mathcal{L} = \left\{x \in \mathbb{R}^n \middle| f(x) \leq f\left(x_0\right)\right\}$ is bounded, i.e., ∃ a constant $B>0$ such that

$$\|x\| \leq B, \forall x \in L \tag{19}$$

Assumption 2. In a neighborhood N of L, the function f is continuously differentiable, and its gradient $\nabla f(x)$ is Lipschitz continuous, that is, ∃ a constant $0<L<\infty$ such that:

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|, \forall x, y \in N \tag{20}$$

Under Assumptions (1) and (2) on f, ∃ a constant c such that

$$\|\nabla f(x)\| \leq \Gamma, \forall x \in L \tag{21}$$

Theorem 3.1. If the relations (6) and (7) are valid, then

$$\mathrm{d}_{k+1}^{\mathrm{hRH}} = \zeta_k \mathrm{d}_{k+1}^{\mathrm{LS}} + \xi_k \mathrm{d}_{k+1}^{\mathrm{FR}} + \omega_k \mathrm{d}_{k+1}^{\mathrm{DY}} + \left(1-\zeta_k-\xi_k-\omega_k\right)\mathrm{d}_{k+1}^{\mathrm{PRP}}.$$

Proof. Now, from (7), we have

$$\mathrm{d}_{k+1}^{\mathrm{hRH}} = -\mathrm{g}_{k+1} + \beta_k^{\mathrm{hRH}} \mathrm{d}_k.$$

Given in the equation (6), the final form is given by

$$\mathrm{d}_{k+1}^{\mathrm{hRH}} = -\mathrm{g}_{k+1} + \begin{pmatrix} \zeta_k \beta_k^{\mathrm{LS}} + \xi_k \beta_k^{\mathrm{FR}} + \omega_k \beta_k^{\mathrm{DY}} \\ +\left(1-\zeta_k-\xi_k-\omega_k\right)\beta_k^{\mathrm{PRP}} \end{pmatrix} \mathrm{d}_k$$

can write

$$\begin{aligned} \mathrm{d}_{k+1}^{\mathrm{hRH}} = & -\left(1-\zeta_k+\zeta_k-\xi_k+\xi_k-\omega_k+\omega_k\right)\mathrm{g}_{k+1} \\ & +\left(\zeta_k \beta_k^{\mathrm{LS}} + \xi_k \beta_k^{\mathrm{FR}} + \omega_k \beta_k^{\mathrm{DY}} + \left(1-\zeta_k-\xi_k-\omega_k\right)\beta_k^{\mathrm{PRP}}\right)\mathrm{d}_k \\ \mathrm{d}_{k+1}^{\mathrm{hRH}} = & -\left(\zeta_k \mathrm{g}_{k+1} + \xi_k \mathrm{g}_{k+1} + \omega_k \mathrm{g}_{k+1} + \left(1-\zeta_k-\xi_k-\omega_k\right)\mathrm{g}_{k+1}\right) \\ & +\left(\zeta_k \beta_k^{\mathrm{LS}} + \xi_k \beta_k^{\mathrm{FR}} + \omega_k \beta_k^{\mathrm{DY}} + \left(1-\zeta_k-\xi_k-\omega_k\right)\beta_k^{\mathrm{PRP}}\right)\mathrm{d}_k \end{aligned}$$

It follows that

$$\begin{aligned} \mathbf{d}_{k+1}^{\mathrm{hRH}} = & \zeta_k\left(-\mathrm{g}_{k+1} + \beta_k^{\mathrm{LS}}\mathrm{d}_k\right) + \xi_k\left(-\mathrm{g}_{k+1} + \beta_k^{\mathrm{FR}}\mathrm{d}_k\right) + \\ & \omega_k\left(-\mathrm{g}_{k+1} + \beta_k^{\mathrm{DY}}\mathrm{d}_k\right) + \left(1-\zeta_k-\xi_k-\omega_k\right)\left(-\mathrm{g}_{k+1} + \beta_k^{\mathrm{PRP}}\mathrm{d}_k\right) \end{aligned}$$

Here, we obtain

$$\mathrm{d}_{k+1}^{\mathrm{hRH}} = \zeta_k \mathrm{d}_{k+1}^{\mathrm{LS}} + \xi_k \mathrm{d}_{k+1}^{\mathrm{FR}} + \omega_k \mathrm{d}_{k+1}^{\mathrm{DY}} + \left(1-\zeta_k-\xi_k-\omega_k\right)\mathrm{d}_{k+1}^{\mathrm{PRP}} \tag{22}$$

Thus, the theorem has been proven.

Theorem 3.2. Let the sequences $\{\mathrm{g}_k\}$ and $\{\mathrm{d}_k\}$ be generated by the hRH method. The search direction d_k satisfies the sufficient descent condition.

$$\mathrm{g}_{k+1}^T \mathrm{d}_{k+1} \leq -c\left\|\mathrm{g}_{k+1}\right\|^2, \forall k \geq 0, where\ c > 0 \tag{23}$$

Proof. According to the hRH method, when the Powell restart criterion is satisfied, that is, $\left|\mathrm{g}_{k+1}^T \mathrm{g}_k\right| \geq 0.2\left\|\mathrm{g}_{k+1}\right\|^2$ it follows that $\mathrm{d}_{k+1} = -\mathrm{g}_{k+1}$ and (23) is also valid. We proceeded with the assumption that the Powell restart criterion was invalid. Next, we obtain

$$\left|\mathrm{g}_{k+1}^T \mathrm{g}_k\right| \leq 0.2\left\|\mathrm{g}_{k+1}\right\|^2 \tag{24}$$

We demonstrate the proof by mathematical induction for $k=0$, $d_0 = -g_0$ so $g_0^T d_0 = -\left\|g_0\right\|^2$ We found that the sufficient descent condition is satisfied for $k=0$. We now assume that (23) is hold for k, thus

$$g_k^T d_k \le -c\|g_k\|^2 \tag{25}$$

We now show that this is true for k+1. From the strong Wolfe condition (5),

$$\le \frac{\Gamma LD}{c\,r^2} + \frac{\Gamma^2}{r^2} + \frac{\Gamma^2}{c\,(1-\sigma)\,r^2} + \frac{LD}{r^2} = M\,. \tag{26}$$

Multiplying (22) by g_{k+1}^T from the left yields

$$g_{k+1}^T d_{k+1} = \zeta_k g_{k+1}^T d_{k+1}^{LS} + \xi_k g_{k+1}^T d_{k+1}^{FR} + \omega_k g_{k+1}^T d_{k+1}^{DY} + (1-\zeta_k-\xi_k-\omega_k) g_{k+1}^T d_{k+1}^{PRP} \tag{27}$$

It is necessary to prove 15 cases depending on the parameter values:

Case 1: If ζ_k=1, ξ_k=0 and ω_k=0, then the relation (27) becomes

$$g_{k+1}^T d_{k+1} = g_{k+1}^T d_{k+1}^{LS}.$$

The search direction for the LS method is given by

$$d_{k+1}^{LS} = -g_{k+1} + \beta_k^{LS} d_k \tag{28}$$

Taking the inner product of (28) with g_{k+1}^T

$$g_{k+1}^T d_{k+1}^{LS} = g_{k+1}^T(-g_{k+1}) + \beta_k^{LS} g_{k+1}^T d_k = -\|g_{k+1}\|^2 + \left(\frac{g_{k+1}^T y_k}{-d_k^T g_k}\right) g_{k+1}^T d_k$$

Using the definition $y_k = g_{k+1} - g_k$, we can expand the term $g_{k+1}^T y_k$:

$$g_{k+1}^T d_{k+1}^{LS} = -\|g_{k+1}\|^2 + \left(\frac{g_{k+1}^T(g_{k+1}-g_k)}{-d_k^T g_k}\right) g_{k+1}^T d_k = -\|g_{k+1}\|^2 + \left(\frac{\|g_{k+1}\|^2 - g_{k+1}^T g_k}{-d_k^T g_k}\right) g_{k+1}^T d_k = -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2 g_{k+1}^T d_k}{-d_k^T g_k} - \frac{(g_{k+1}^T g_k)(g_{k+1}^T d_k)}{-d_k^T g_k}$$

By taking the absolute value and applying the triangle inequality, we obtain

$$g_{k+1}^T d_{k+1}^{LS} \le -\|g_{k+1}\|^2 + \left|\frac{g_{k+1}^T d_k}{-d_k^T g_k}\right| \|g_{k+1}\|^2 + \left|\frac{g_{k+1}^T d_k}{-d_k^T g_k}\right| |g_{k+1}^T g_k|.$$

From the second strong Wolfe condition (5) and d_k is a descent direction (25), which gives

$$\left|\frac{g_{k+1}^T d_k}{-d_k^T g_k}\right| \le \sigma \tag{29}$$

Substituting (29) into our main expression, we obtain

$$g_{k+1}^T d_{k+1}^{LS} \le -\|g_{k+1}\|^2 + \sigma\|g_{k+1}\|^2 + \sigma|g_{k+1}^T g_k| = -(1-\sigma)\|g_{k+1}\|^2 + \sigma|g_{k+1}^T g_k|$$

Since we are assumed that the Powell restart condition is not met, we have from (24) that $|g_{k+1}^T g_k| \le 0.2\|g_{k+1}\|^2$. Substituting into (30):

$$g_{k+1}^T d_{k+1}^{LS} \le -(1-\sigma)\|g_{k+1}\|^2 + \sigma\left(0.2\|g_{k+1}\|^2\right) = -(1-\sigma-0.2\sigma)\|g_{k+1}\|^2 = -(1-1.2\sigma)\|g_{k+1}\|^2$$

Let $c_1 = (1-1.2\sigma)$. Since σ<0.5, we have c_1>0. Therefore:

$$g_{k+1}^T d_{k+1}^{LS} \le -c_1\|g_{k+1}\|^2, \text{ where } c_1 > 0$$

Therefore, the search direction d_{k+1}^{LS} fulfills the sufficient descent condition.

Case 2: If ζ_k=0, ξ_k=1 and ω_k=0, then the relation (27) becomes

$$g_{k+1}^T d_{k+1} = g_{k+1}^T d_{k+1}^{FR}.$$

A sufficient descent condition for the FR method under strong Wolfe conditions was established in [1]. Thus, ∃ a constant $c_2 > 0$ such that

$$g_{k+1}^T d_{k+1}^{FR} \le -c_2\|g_{k+1}\|^2$$

Case 3: If ζ_k=0, ξ_k=0 and ω_k=1, then the relation (27) becomes

$$
\begin{aligned}
g_{k+1}^T d_{k+1} &= g_{k+1}^T d_{k+1}^{DY} \\
&= g_{k+1}^T \left(-g_{k+1} + \beta_{k+1}^{DY} d_k\right) \\
&= -\|g_{k+1}\|^2 + \beta_{k+1}^{DY} g_{k+1}^T d_k \\
&= -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{y_k^T d_k} g_{k+1}^T d_k \\
&\le -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{d_k^T y_k} \left| d_k^T g_{k+1} \right| \\
&\le -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{-(1-\sigma) d_k^T g_k} \left(-\sigma g_k^T d_k\right) (\textit{by using (5) and (26)}) \\
&\le -\frac{1-2\sigma}{1-\sigma} \|g_{k+1}\|^2
\end{aligned} \tag{31}
$$

Let $c_3 = \dfrac{1-2\sigma}{1-\sigma}$. Since $\sigma < 0.5$, we have $c_3 > 0$. Therefore:

$$g_{k+1}^T d_{k+1}^{DY} \le -c_3 \|g_{k+1}\|^2$$

Thus, it has been shown that d_{k+1}^{DY} fulfills the sufficient descent condition.

Case 4: If $\zeta_k=0$, $\xi_k=0$ and $\omega_k=0$ then the relation (27) becomes

$$g_{k+1}^T d_{k+1} = g_{k+1}^T d_{k+1}^{PRP}.$$

We demonstrate the sufficient descent condition of the PRP method by the following theorem.

Theorem 3.3. [19] Under Assumptions (1) and (2), suppose that the step sizes a_k satisfy conditions (4) and (5) with $\sigma < \dfrac{1}{2}$. If $\|s_k\| \to 0$ and there exist non-negative constants τ_1 and τ_2 such that

$$\|g_k\|^2 \ge \tau_1 \|s_k\|^2 \tag{32}$$

$$\|g_{k+1}\|^2 \le \tau_2 \|s_k\| \tag{33}$$

where $s_k = x_{k+1} - x_k = a d_k$. Subsequently, $\left(d_{k+1}^{PRP}\right)$ satisfies the sufficient descent condition $\forall k$.

Proof.

$$d_{k+1} = -g_{k+1}^T + \beta_k^{PRP} s_k \tag{34}$$

Now, multiplying (34) by g_{k+1}^T from the left, we get

$$g_{k+1}^T d_{k+1}^{PRP} = -\|g\|_{k+1}^2 + \beta_k^{PRP} g_{k+1}^T s_k \tag{35}$$

$$g_{k+1}^T d_{k+1}^{PRP} = -\|g_{k+1}\|^2 + \frac{g_{k+1}^T y_k}{\|g_k\|^2} g_{k+1}^T s_k \tag{36}$$

$$g_{k+1}^T d_{k+1}^{PRP} \le -\|g_{k+1}\|^2 + \frac{\left| g_{k+1}^T y_k \quad g_{k+1}^T s_k \right|}{\|g_k\|^2} \tag{37}$$

By the Cauchy-Schwarz inequality, we have

$$g_{k+1}^T d_{k+1}^{PRP} \le -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\| \|y_k\| \|g_{k+1}\| \|s_k\|}{\|g_k\|^2} \tag{38}$$

$$g_{k+1}^T d_{k+1}^{PRP} \le -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2 \|y_k\| \|s_k\|}{\|g_k\|^2} \tag{39}$$

From (20), $\|y_k\| \le L\|s_k\|$, so

$$g_{k+1}^T d_{k+1}^{PRP} \le -\|g_{k+1}\|^2 + \frac{L\|g_{k+1}\|^2 \|s_k\|^2}{\|g_k\|^2} \tag{40}$$

By using (33), we get

$$g_{k+1}^T d_{k+1}^{PRP} \le -\|g_{k+1}\|^2 + \frac{L\tau_2 \|s_k\|^3}{\|g_k\|^2} \tag{41}$$

Using (32), it further becomes

$$g_{k+1}^T d_{k+1}^{PRP} \le -\|g_{k+1}\|^2 + \frac{1}{\tau_1} L\tau_2 \|s_k\| \tag{42}$$

However, because assumption $\|s_k\| \to 0$, the second summand in (42) tends to zero, so there exists a number $0<\gamma<1$, such that

$$\frac{1}{\tau_1} L\tau_2 \|s_k\| \le \gamma \|g_{k+1}\|^2 \tag{43}$$

Now, (42) becomes

$$g_{k+1}^T d_{k+1}^{PRP} \le -\|g_{k+1}\|^2 + \gamma \|g_{k+1}\|^2 \tag{44}$$

$$g_{k+1}^T d_{k+1}^{PRP} \le (1-\gamma)\|g_{k+1}\|^2 \tag{45}$$

$$g_{k+1}^T d_{k+1}^{PRP} \le -c_4 \|g_{k+1}\|^2 \tag{46}$$

where $c_4 = 1-\gamma > 0$ (47)

Therefore, a sufficient descent condition holds for all k.

Case 5: If $\xi_k=0$, $\omega_k=0$ and $0<\zeta_k<1$, then the relation (27) becomes

$$g_{k+1}^T d_{k+1} = \zeta_k g_{k+1}^T d_{k+1}^{\mathrm{LS}} + (1-\zeta_k) g_{k+1}^T d_{k+1}^{\mathrm{PRP}} = g_{k+1}^T d_{k+1}^{\mathrm{LSPRP}}$$

The sufficient descent condition is satisfied and discussed in [36], such that

$$g_{k+1}^T d_{k+1}^{LSPRP} \le -c_5 \|g_{k+1}\|^2, \text{ where } c_5>0$$

Case 6: If $\zeta_k=0$, $0<\xi_k<1$ and $\omega_k=0$, then the relation (27) becomes

$$g_{k+1}^T d_{k+1} = \xi_k g_{k+1}^T d_{k+1}^{\mathrm{FR}} + (1-\xi_k) g_{k+1}^T d_{k+1}^{\mathrm{PRP}} = g_{k+1}^T d_{k+1}^{\mathrm{FRPRP}}$$

Djordjević established in [18] that the sufficient descent condition is satisfied for d_{k+1}^{FRPRP}

$$g_{k+1}^T d_{k+1}^{\mathrm{FRPRP}} \le -c_6 \|g_{k+1}\|^2 \text{ where } c_6>0$$

Case 7: If $\zeta_k=0$, $\xi_k=0$ and $0<\omega_k<1$, then the relation (27) becomes

$$g_{k+1}^T d_{k+1} = \omega_k g_{k+1}^T d_{k+1}^{\mathrm{DY}} + (1-\omega_k) g_{k+1}^T d_{k+1}^{\mathrm{PRP}} = g_{k+1}^T d_{k+1}^{\mathrm{DYPRP}}.$$

In Babando *et al.* [3], proved that the sufficient descent condition holds for d_{k+1}^{DYPRP}

$$g_{k+1}^T d_{k+1}^{DYPRP} \le -c_7 \|g_{k+1}\|^2, \text{ where } c_7>0$$

Case 8: If $\zeta_k+\xi_k=1$, $0<\zeta_k<1$, $0<\xi_k<1$ and $\omega_k=0$, then the relation (27) becomes

$$g_{k+1}^T d_{k+1} = \zeta_k g_{k+1}^T d_{k+1}^{\mathrm{LS}} + (1-\zeta_k) g_{k+1}^T d_{k+1}^{\mathrm{FR}} = g_{k+1}^T d_{k+1}^{\mathrm{LSFR}}$$

In Djordjević [19], Djordjević shown that the sufficient descent condition is satisfied for d_{k+1}^{LSFR}.

$$g_{k+1}^T d_{k+1}^{\mathrm{LSFR}} \le -c_8 \|g_{k+1}\|^2$$

Case 9: If $0<\zeta_k<1$, $0<\xi_k<1$, $\omega_k=0$ and $0<\zeta_k+\xi_k<1$, then the relation (27) becomes

$$g_{k+1}^T d_{k+1} = \zeta_k g_{k+1}^T d_{k+1}^{\mathrm{LS}} + \xi_k g_{k+1}^T d_{k+1}^{\mathrm{FR}} + (1-\zeta_k-\xi_k) g_{k+1}^T d_{k+1}^{\mathrm{PRP}} = g_{k+1}^T d_{k+1}^{\mathrm{LSFRPRP}}$$

Hanachi *et al.* proved in [22] that $d_{k+1}^{LSFRPRP}$ satisfies the sufficient descent condition $\forall k$, i.e.,

$$g_{k+1}^T d_{k+1}^{LSFRPPP} \le -c_9 \|g_{k+1}\|^2, \text{ where } c_9>0$$

Case 10: If $0<\zeta_k<1$, $0<\omega_k<1$, $\xi_k=0$ and $\zeta_k+\omega_k=1$, then the relation (27) becomes

$$g_{k+1}^T d_{k+1} = \zeta_k g_{k+1}^T d_{k+1}^{\mathrm{LS}} + (1-\zeta_k) g_{k+1}^T d_{k+1}^{\mathrm{DY}} = g_{k+1}^T d_{k+1}^{\mathrm{LSDY}}$$

Liu and Li in [20] shown that d_{k+1}^{LSDY} satisfies the sufficient descent condition $\forall k$, i.e.,

$$g_{k+1}^T d_{k+1}^{\mathrm{LSDY}} \le -c_{10} \|g_{k+1}\|^2, \text{ where } c_{10}>0$$

Case 11: If $0<\zeta_k<1$, $0<\omega_k<1$, $\xi_k=0$ and $0<\zeta_k+\omega_k<1$, then there exist constants $\varphi_1,\varphi_2,\varphi_3,\varphi_4$ such that $0<\varphi_1\le\zeta_k\le\varphi_3<1$ and $0<\varphi_2\le\omega_k\le\varphi_4<1$. Now, the relation (27) becomes

$$\begin{aligned} g_{k+1}^T d_{k+1}^{\mathrm{LSDYPRP}} &\le \varphi_1 g_{k+1}^T d_{k+1}^{\mathrm{LS}} + \varphi_2 g_{k+1}^T d_{k+1}^{\mathrm{DY}} \\ &\quad + (1-\varphi_3-\varphi_4) g_{k+1}^T d_{k+1}^{\mathrm{PRP}} \\ &\le -\varphi_1 c_1 \|g_{k+1}\|^2 - \varphi_2 c_3 \|g_{k+1}\|^2 \\ &\quad - (1-\varphi_3-\varphi_4) c_4 \|g_{k+1}\|^2 \end{aligned}$$

Hence

$$g_{k+1}^T d_{k+1}^{LSDYPRP} \le -\Omega \|g_{k+1}\|^2, \; with\, \Omega = \begin{bmatrix} \varphi_1 c_1 + \varphi_2 c_2 \\ +(1-\varphi_3-\varphi_4) c_4 \end{bmatrix} > 0$$

Thus, it has been shown that $d_{k+1}^{LSDYPRP}$ fulfills the sufficient descent condition.

Case 12: If $0<\xi_k<1$, $0<\omega_k<1$, $\zeta_k=0$ and $\zeta_k+\omega_k=1$, then the relation (27) becomes

$$g_{k+1}^T d_{k+1} = \xi_k g_{k+1}^T d_{k+1}^{\mathrm{FR}} + (1-\xi_k) g_{k+1}^T d_{k+1}^{\mathrm{DY}} = g_{k+1}^T d_{k+1}^{\mathrm{FRDY}}$$

Abubakar *et al.* in [37] shown that d_{k+1}^{FRDY} satisfies the sufficient descent condition$\forall k$, i.e.,

$$g_{k+1}^T d_{k+1}^{\mathrm{FRDY}} \le -c_{12} \|g_{k+1}\|^2, \text{ where } c_{12}>0$$

Case 13: If $0<\xi_k<1$, $0<\omega_k<1$, $\zeta_k=0$ and $0<\xi_k+\omega_k<1$, then the relation (27) becomes

$$g_{k+1}^T d_{k+1} = \xi_k g_{k+1}^T d_{k+1}^{\mathrm{FR}} + \omega_k g_{k+1}^T d_{k+1}^{\mathrm{DY}} + (1-\xi_k-\omega_k) g_{k+1}^T d_{k+1}^{\mathrm{PRP}} = g_{k+1}^T d_{k+1}^{\mathrm{FRDYPRP}}$$

Hanachi *et al.* proved in [21] that $d_{k+1}^{FRDYPRP}$ satisfies the sufficient descent condition $\forall k$, i.e.,

$g_{k+1}^T d_{k+1}^{FRDYPRP} \le -c_{13}\|g_{k+1}\|^2$, where $c_{13}>0$

Case 14: If $0<\zeta_k<1$, $0<\omega_k<1$ and $0<\xi_k<1$ with $\zeta_k+\xi_k+\omega_k=1$, then there exist constants $\gamma_1,\gamma_2,\gamma_3$ such that $0<\gamma_1\le\zeta_k\le1$, $0<\gamma_2\le\xi_k\le1$, and $0<\gamma_3\le\omega_k\le1$. Now, the relation (27) becomes

$$\begin{aligned} g_{k+1}^T d_{k+1}^{LSFRDY} &\le \gamma_1 g_{k+1}^T d_{k+1}^{LS} + \gamma_2 g_{k+1}^T d_{k+1}^{FR} + \gamma_3 g_{k+1}^T d_{k+1}^{DY} \\ &\le -\gamma_1 c_1 \|g_{k+1}\|^2 - \gamma_2 c_2 \|g_{k+1}\|^2 - \gamma_3 \|g_{k+1}\|^2 \end{aligned}$$

Hence

$$g_{k+1}^T d_{k+1}^{LSFRDY} \le -\Phi {g_{k+1}}^2, \; with\, \Phi = \left[\gamma_1 c_1 + \gamma_2 c_2 + \gamma_3 c_3\right] > 0$$

Thus, it has been shown that d_{k+1}^{LSFRDY} fulfills the sufficient descent condition.

Case 15: If $0<\zeta_k<1$, $0<\omega_k<1$ and $0<\xi_k<1$ with $0<\zeta_k+\xi_k+\omega_k<1$, then there exist constants $\varpi_1,\varpi_2,\varpi_3,\varpi_4,\varpi_5,\varpi_6$ such that $0<\varpi_1\le\zeta_k\le\varpi_4<1$, $0<\varpi_2\le\xi_k\le\varpi_5<1$, and $0<\varpi_3\le\omega_k\le\varpi_6<1$. Now, the relation (27) becomes

$$\begin{aligned} g_{k+1}^T d_{k+1} \le\ & \varpi_1 g_{k+1}^T d_{k+1}^{LS} + \varpi_2 g_{k+1}^T d_{k+1}^{FR} + \varpi_3 g_{k+1}^T d_{k+1}^{DY} \\ & + \left(1-\varpi_4-\varpi_5-\varpi_6\right) g_{k+1}^T d_{k+1}^{PRP} \\ \le\ & -\varpi_1 c_1 \|g_{k+1}\|^2 - \varpi_2 c_2 \|g_{k+1}\|^2 \\ & -\varpi_3 c_3 \|g_{k+1}\|^2 - \left(1-\varpi_4-\varpi_5-\varpi_6\right) c_4 \|g_{k+1}\|^2. \\ g_{k+1}^T d_{k+1} \le\ & -\Psi \|g_{k+1}\|^2 \\ & with\, \Psi = \left[\varpi_1 c_1 + \varpi_2 c_2 + \varpi_3 c_3 + \left(1-\varpi_4-\varpi_5-\varpi_6\right) c_4\right] > 0. \end{aligned}$$

Thus, it has been shown that $d_{k+1}^{LSFRDYPRP}$ fulfills the sufficient descent condition.

4. CONVERGENCE ANALYSIS

In this section, we study the global convergence properties of the proposed CG method. For further consideration, we need Assumptions (1) and (2), and the following important result, established by Zoutendijk [38] and Wolfe [4].

Theorem 4.1. Suppose that Assumptions (1) and (2) hold. Consider the iterative method (2), in which, $\forall k\ge0$, the search direction d_k is the descent direction, and the step size a_k satisfies the strong Wolfe conditions (4) and (5). Then

$$\sum_{k=0}^{\infty} \cos^2\phi_k \|g_k\|^2 < \infty \tag{48}$$

Where

$$\cos\phi_k = -\frac{g_k^T d_k}{\|g_k\|\|d_k\|^2} \tag{49}$$

Inequality (48) is called the Zoutendijk condition.

Lemma 1. Assume that Assumptions (1) and (2) hold. Let d_k be a descent direction (23) and suppose the step size a_k satisfies (5). Then, the step size a_k satisfies the following lower bound,

$$\alpha_k \ge \frac{(1-\sigma)}{L}\frac{\left|d_k^T g_k\right|}{\|d_k\|^2}, \forall k \ge 0 \tag{49}$$

Proof. From the (26), we have

$$-(1-\sigma) d_k^T g_k \le \sigma d_k^T g_k - d_k^T g_k = d_k^T \left(g_{k+1} - g_k\right)$$

By applying the Cauchy-Schwarz inequality and Lipschitz condition (20), we obtain

$$\begin{aligned} & d_k^T \left(g_{k+1} - g_k\right) \le \|d_k\| \cdot \|g_{k+1} - g_k\| \le \|d_k\| \cdot L\alpha_k \|d_k\| \\ & = L\alpha_k \|d_k\|^2 \end{aligned}$$

Combining the above inequalities yields

$$-(1-\sigma) d_k^T g_k \le L\alpha_k \|d_k\|^2$$

Rearranging gives the desired result (49).

According to Lemma (1) and Assumption (1), (2), the strong Wolfe condition (4), (5), and decent direction (23), we conclude that a_k, obtained from our innovative method, is non-zero, i.e., $\exists$ a constant $\eta > 0$ such that

$$a_k \ge \eta, \forall_k \ge 0 \tag{50}$$

The global convergence of the hRH method is demonstrated by the following theorem:

Theorem 4.2. Assume that Assumptions (1) and (2) holds. The sequence $\{x_k\}$ and $\{d_k\}$ generated by the hRH algorithm, and the search direction d_k is a descent direction, and the stepsize a_k satisfies the strong Wolfe conditions (4) and (5), then

$$\lim_{k\to\infty} inf g_k = 0 \tag{51}$$

Proof. A contradiction is used to prove this theorem. Assume that formula (51) does not hold. Therefore, ∃ a constant $r > 0$ such that

$$\|g_k\| \geq r, \ \forall k \geq 1 \tag{52}$$

According to Theorem (3.2), it can be deduced that

$$g_k^T d_k \leq -c\|g_k\|^2, \ \forall c > 0$$

We can write it as

$$-g_k^T d_k \geq c\|g_k\|^2, \forall c > 0$$

Then, by (26) and (53), we get

$$d_k^T y_k = d_k^T g_{k+1} - d_k^T g_k \geqslant -(1-\sigma) d_k^T g_k \geqslant c(1-\sigma)\|g_k\|^2 \tag{54}$$

From (20), we obtain

$$\|y_k\| = \|g_{k+1} - g_k\| \leq L\|s_k\| \leq LD \tag{55}$$

where $D = \max\{\|x-y\|, x,y \in N\}$ is the diameter of N and $s_k = x_{k+1} - x_k$.

We have

$$\begin{aligned} d_{k+1} &= -g_{k+1} + \beta_K^{\text{hRH}} d_k \\ \|d_{k+1}\| &\leq \|g_{k+1}\| + \left|\beta_K^{\text{hRH}}\right| \|d_k\| \end{aligned}$$

From (6), we obtain

$$\begin{aligned} \left|\beta_k^{\text{hRH}}\right| &= \left|\zeta_k \beta_k^{\text{LS}} + \xi_k \beta_k^{\text{FR}} + \omega_k \beta_k^{\text{DY}} + (1-\zeta_k - \xi_k - \omega_k)\beta_k^{\text{PRP}}\right| \\ &= \frac{\left|g_{k+1}^T y_k\right|}{\left|d_k^T g_k\right|} + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} + \frac{\|g_{k+1}\|^2}{\left|d_k^T y_k\right|} + \frac{\left|g_{k+1}^T y_k\right|}{\|g_k\|^2} \\ &\leq \frac{\|g_{k+1}\|\|y_k\|}{c\|g_k\|^2} + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} + \frac{\|g_{k+1}\|^2}{c(1-\sigma)\|g_k\|^2} + \frac{\|g_{k+1}\|\|y_k\|}{\|g_k\|^2} \end{aligned}$$

By using (52) and (54), we get

$$\leq \frac{\Gamma LD}{c\, r^2} + \frac{\Gamma^2}{r^2} + \frac{\Gamma^2}{c\,(1-\sigma)\, r^2} + \frac{LD}{r^2} = M.$$

The initial inequality is derived from the conditions $0<\zeta_k, \xi_k, \omega_k<1$ and $0<1-\zeta_k-\xi_k-\omega_k<1$. The second inequality uses the Cauchy-Schwarz inequality with (23). The final inequality utilizes (21), (52), and (55). Therefore, (3) and (50) imply that

$$\|d_{k+1}\| \leq \|g_{k+1}\| + \left|\beta_k^{\text{hRH}}\right| \|d_k\| \leq \|g_{k+1}\| + \frac{\left|\beta_k^{\text{hRH}}\right| \|s_k\|}{\alpha_k} \leq \Gamma + \frac{MD}{\eta} = W,$$

imply that

$$\sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} = \infty.$$

However, from (48), (52), and (23), we obtain

$$c^2 r^4 \sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} \leq \sum_{k=0}^{\infty} \frac{c^2 \|g_k\|^4}{\|d_k\|^2} \leq \sum_{k=0}^{\infty} \frac{\left(d_k^T g_k\right)^2}{\|d_k\|^2} < \infty.$$

This contradicts Theorem (4.1). Therefore, equation (52) is invalid, and the claim in (51) is proven.

5. THE NUMERICAL RESULTS

This section presents a numerical evaluation of the proposed hRH method compared to the established methods: DHF [23], HLSDY [20], HLSFR [19], and HDYCDHS [24]. The evaluations used 170 benchmark test problems obtained from [39-41] with different dimensions (n) ranging from 2 to 200,000. The measures adopted included the time required to solve the problem, the number of iterations, and the number of function and gradient evaluations. The implementation was performed using MATLAB (R2024b) on a personal computer with an Intel(R) Core (TM) i5-8250U CPU @ 1.60 GHz, 8.00 GB RAM, and a 64-bit Windows 11 Education operating system. The convergence criterion was set to $\|g_k\| \leq 0^{-6}$ or when the maximum number of iterations reaches 2000. A comprehensive list of these test problems is provided in Table 2. All of these algorithms implement the strong Wolfe parameters and are fixed at $\delta = 0.0001$ and $\sigma = 0.1$. The numerical comparisons were based on the performance profile framework proposed by Dolan and Moré [42]. In the analysis of the performance profiles, the upper curve indicates that the method is better. The effectiveness of the method is determined according to the performance metric $P_s(\tau)$, which measures the performance of the solver on a set of benchmark problems. Here, $P_s(\tau)$ denotes the proportion of problems solved by the solver with a performance ratio less than or equal to τ. Using this metric, we obtained the ratio of the solver's capacity

to properly assess how this optimization problems can be solved. Thus, efficient procedures must be considered and simplified. Taking $P_s(\tau) = P(\tau)$, it is also taken that $S = \tau$, The numerical results were compared graphically. Figs. 1-3 show the comparison results.

TABLE 2: List of test problems, dimensions, and initial points

Function name	Dimensions tested (n)	Initial point (s)
ARGLINB	2, 500, 1000	
BV	500, 1000, 10000, 20000, 30000	(0,...,0)
COSINE	10, 100, 1000, 10000, 100000	(1,...,1)
DIXMAANA	300, 30000, 60000, 90000, 120000	(2,...,2)
DIXMAANL	3000, 9000, 30000, 60000	(2,...,2)
DIXMAANK	3000, 9000, 30000, 60000, 120000	(2,...,2)
DIAGONAL 4	100, 1000, 10000, 100000, 200000	(1,...,1)
DIAGONAL 5	100, 1000, 10000, 50000, 200000	(1.1,...,1.1)
DIAGONAL 6	100, 1000, 10000, 50000, 100000	(−1.3,...,−1.3)
DIAGONAL 8	1000, 5000, 10000	(1,...,1)
DQRTIC	5000, 10000, 50000, 100000, 150000	(2,...,2)
EDENSCH	2, 100, 1000, 10000, 100000	(0,...,0)
DENSCHNF	10, 100, 1000, 10000, 100000	(0.02,...,0.02)
EDENSCHNB	2, 100, 10000, 100000	(0.3,...,0.3)
ETRIG1	2, 1000, 10000	(0.5,...,0.5)
ENGVAL1	10, 1000, 10000	(−1,...,−1)
EVF	2	(11,...,11)
EXPENALTY	100, 1000, 25000, 50000	(0.5,...,0.5)
EXTROSNB	2, 500, 1000, 10000	(−2,...,−2)
EXROSEN	10, 100, 1000, 5000, 10000, 50000, 100000	(2,...,2)
EXHIMMELBLAU	10, 1000, 10000	(−0.5,...,0.5)
GENHUMPS	2, 100, 500, 1500	(1,...,−1)
GENQUARTIC	1000, 25000, 75000	(1,...,1)
GQUARTIC	50, 100, 500, 1000	(1,...,1)
HARKERP	100, 1000, 10000, 50000	(−1,...,−1)
HIMMELBH	100, 10000, 100000	(0.2,...,0.2)
HIMMELBG	100, 1000, 10000, 100000	(1.5,...,1.5)
IE	100, 500, 1000	(0,...,0)
LIN	100, 250, 500	(1,...,1)
LIARWHD	1000, 50000, 100000	(4,...,4)
NONSCOMP	100, 5000, 25000, 50000, 100000, 200000	(−1,...,−1)
PEN1	100, 250, 500	(1. or 0.5.)
QP1	1000, 10000, 25000, 50000	(−1,...,−1)
QP2	100, 1000, 10000, 100000	(10,...,10)
QP3	100, 1000, 10000, 100000	(−1,...,−1)
QUARTC	100, 1000, 10000, 25000, 50000	(0.23,...,0.23)
QUARTICM	1000, 10000, 100000	(0.03,...,0.03)
RAYDAN1	50, 500, 50000	(1,...,1)
SINE	100, 1000, 5000, 50000, 100000	(1,...,1)
SHALLOW	25000, 50000, 100000	(10,...,10)
TRID	10, 100, 1000, 2000, 5000	(−1,...,−1)
WOODS	1000, 100000, 150000, 200000	(0.5,...,−2, 0,...,0)

In Fig. 1, we present the performance curves of the computational time. The vertical axis represents the fraction solved within a given ratio τ, while the horizontal axis represents the time ratio τ. When $0<\tau\leq0.5$, hRH shows better performance and solves 77% of the problems in less time than the other methods, while HLSDY and HLSFR 70% and HDYCDHS and DHF 67% and 51%, respectively. With τ rising to $\tau\geq0.5$, hRH remains the most efficient at 95%,

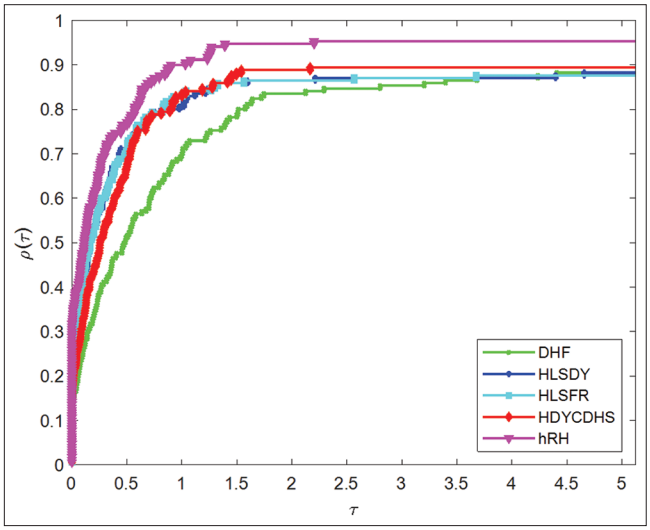


Fig. 1. Time performance for DHF, HLSDY, HLSFR, HDYCDHS, and hRH methods.

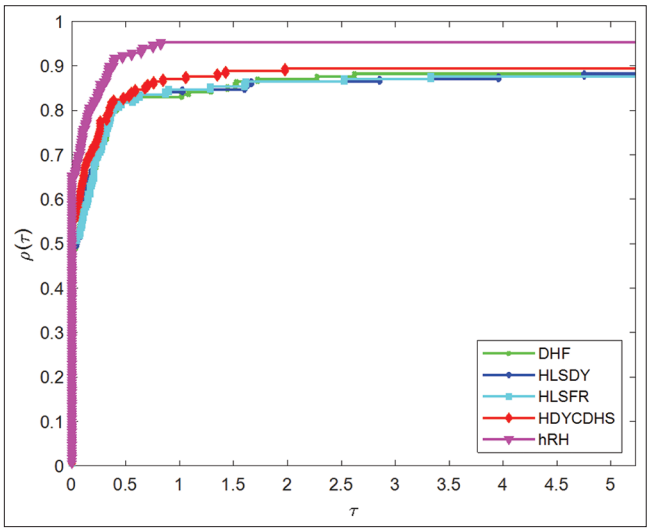


Fig. 2. Iteration count performance for DHF, HLSDY, HLSFR, HDYCDHS, and hRH methods.

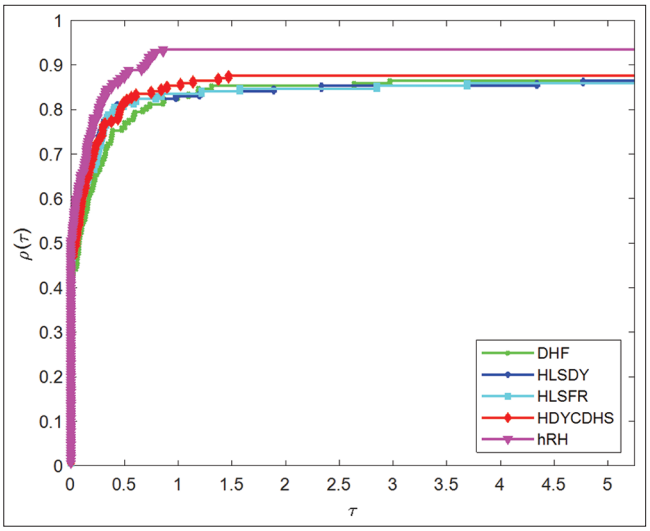


Fig. 3. Objective function/gradient calculation performance for DHF, HLSDY, HLSFR, HDYCDHS, and hRH methods.

While HLSDY takes the second place producing 89%, third place is taken by DHF and HLSDY 88%, and HLSFR in last place 87%. As expected, hRH is faster in terms of computational time than the other algorithms. Concurrently, if τ is between $0.5<\tau\leq1$, our method solved 90% problems, and that of the second-best method, HDYCDHS solved 82% problems.

Fig. 2 shows the relative measures, which are the iteration counts of each algorithm for convergence. In $0<\tau\leq0.5$, the hRH method demonstrated the highest resolution rate compared to the other methods, solving 92% of the problems with a smaller number of iterations than all the other methods, approximately 82%. Again, as $\tau\geq0.5$, the efficiency of the hRH method is 95% while for HDYCDHS, it is 89%, and for the other methods, it is 88%. Meanwhile, if τ is between $0.5<\tau\leq1$, the proposed method solved 95%, the HDYCDHS method 87% and the other methods 84%. These results demonstrate that the proposed hRH algorithm achieves a better solution with fewer iterations, particularly in high-dimensional or challenging problem domains.

The performance profile of the number of function/gradient evaluations is depicted in Fig. 3. The relative performance of hRH within the range $0<\tau\leq0.5$ is as follows: hRH solves 88% of the problems with fewer evaluations compared to HLSDY, HLSFR, and HDYCDHS solve 81% and DHF solves 76%. Up to $\tau\geq1$, the efficiency of hRH is 95%, that of the second-best method HDYCDHS is 89%, and that of the others is 87%. The results show that the proposed hRH algorithm requires fewer function/gradient evaluations than other algorithms, while significantly reducing the overall computational cost.

For the EXROSEN function, the proposed (hRH) method uniquely succeeded in solving the problem across the seven tested dimensions, whereas the comparative methods (DHF, HLSDY, HLSFR, and HDYCDHS) failed in all dimensions because of jamming, that is, generating many short steps without making significant progress toward the solution. The numerical simulation also confirmed the effectiveness of the hRH method for the investigated performance parameters. It can be observed that it always performs better than the other methods in terms of computational time, number of iterations, and number of function/gradient evaluations. These results demonstrate that the proposed hRH algorithm is stable and fast for optimizing unconstrained optimization problems.

6. IMAGE RESTORATION PROBLEMS

In this section, we evaluate the performance of the proposed hRH algorithm in restoring color images corrupted by impulse noise.

Following the spirit of [43], we treat high-density noise in digital images using a dual-phase framework. In the first phase, we used an adaptive median-based method to identify the corrupted pixels. Let us assume that Z is a color image of dimensions $M\times N$ and $\Omega = \{1,\ldots M\}\times\{1,\ldots N\}$ is the pixel index domain. Therefore, we assume that the subset of noisy pixels, denoted by D, is a subset of and choose $|D|$ to be the total number of noisy pixels. Every noisy pixel at location (i,j) has four immediate neighbors in the neighborhood W_{ij} around it: $\{(i-1,j),(i+1,j),(i,j-1),(i,j+1)\}$.

We observe the noisy pixel value, denoted as z_{ij}, and the recovered value, denoted as v_{ij}. The denoising process is formulated as a non-smooth optimization.

$$\min_{v}\sum_{(i,j)\in\mathcal{D}}\left[\left|v_{i,j}-z_{i,j}\right|+\frac{\lambda}{2}\left(2\cdot R_{i,j}^{(1)}+R_{i,j}^{(2)}\right)\right] \tag{56}$$

where the regularization terms are given by

$$R_{i,j}^{(1)}=\sum_{(m,n)\in\mathcal{W}_{i,j}\setminus\mathcal{D}}\psi_{\delta}\left(v_{i,j}-z_{m,n}\right),\ R_{i,j}^{(2)}=\sum_{(m,n)\in\mathcal{W}_{i,j}\cap\mathcal{D}}\psi_{\delta}\left(v_{i,j}-v_{m,n}\right).$$

The function $\psi_\delta(t) = \sqrt{t^2 + \delta}$ acts as a smooth penalty term that preserves the edges and is characterized by a positive parameter δ. The vector $v = [v_{i,j}]_{(i,j)\in D}$ is a lexicographically ordered list of unknowns. Directly minimizing (56) can be computationally expensive, particularly for high-resolution images with dense noise.

To resolve this issue, a smooth surrogate model was presented in [44], eliminating the non-smooth term and resulting in the following equivalent formulation:

$$\min_{\mathbf{v}} G_\delta(v) := \sum_{(i,j)\in\mathcal{D}} \left(\begin{array}{l} 2 \sum\limits_{(m,n)\in\mathcal{W}_{i,j}\setminus\mathcal{D}} \psi_\delta\left(v_{i,j} - z_{m,n}\right) + \\ \sum\limits_{(m,n)\in\mathcal{W}_{i,j}\cap\mathcal{D}} \psi_\delta\left(v_{i,j} - v_{m,n}\right) \end{array} \right) \tag{57}$$

With an increasing noise ratio, the number of corrupted pixels also increases, and the size and complexity of the problem (57) increase drastically. Nevertheless, as demonstrated in [44], CGMs remain efficient and scalable for this problem, even under severe corruption.

Nevertheless, color images require more attention than grayscale images because they are multichannel. Each of the three-color images contains three channels: Red, Green, and Blue (RGB), and there may be noise present in each of these channels differently. Therefore, the restoration process is much more complex. Therefore, we adopt a channel-wise restoration strategy where each RGB component is selectively restored independently using the same optimization procedure. To preserve reality in the final output, care is taken to ensure that the overall structure consists of what is present across channels. We show that this strategy can be used effectively to handle independently corrupted color channels without introducing artifacts in the reconstructed image.

We then applied this model to tackle salt-and-pepper noise, which is the most prevalent form of impulse corruption in color images. It was first operated using an adaptive median filter [45] to detect the noisy pixels across all color channels. The solution of (57) using our proposed hRH algorithm is compared in the reconstruction step to several classical CGMs: PRP, LS, and FR methods. We evaluated the performance on several widely used test images of size 768×512: Girl, Lighthouse, Hats, and Macaws. The termination condition is defined as:

$$Iterations > 300 \; or \; \frac{\left|G_\delta\left(\mathbf{v}_k\right) - G_\delta\left(\mathbf{v}_{k-1}\right)\right|}{\left|G_\delta\left(\mathbf{v}_k\right)\right|} \le 10^{-4}$$

The experimental platform matches the one described in section 5. Restoration quality is quantified using the Peak Signal-to-Noise Ratio (PSNR), computed as

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{255^2}{\frac{1}{MN}\sum_{i,j}\left(x^r_{i,j} - x^o_{i,j}\right)^2} \right)$$

where $x^r_{i,j}$ and $x^o_{i,j}$ denote the restored and original pixel values, respectively, and

TABLE 3: Image denoising results for different methods (hRH, PRP, LS, and FR)

Image	Noise	Methods											
		hRH			PRP			LS			FR		
		Itr	Time	PSNR	Itr	Time	PSNR	Itr	Time	PSNR	Itr	Time	PSNR
Girl	0.30	17	28.59	83.14	77	81.96	18.67	27	83.28	77.27	82	21.67	51.10
	0.50	21	54.45	80.79	74	79.97	17.67	28	81.01	73.68	80	20.00	43.03
	0.70	22	66.74	78.48	71	77.98	18.33	64	78.42	70.89	77	22.00	71.70
	0.90	31	139.79	75.78	69	75.33	23.00	72	69.19	68.67	75	22.33	75.34
House	0.30	21	32.75	81.46	81	83.06	20.33	28	81.57	81.69	83	15.33	16.08
	0.50	20	67.59	79.57	80	81.15	20.67	72	79.04	79.27	82	17.33	61.51
	0.70	26	51.05	76.89	77	79.10	13.33	41	75.46	75.60	79	21.00	41.35
	0.90	16	71.86	72.66	72	74.53	24.67	86	73.10	74.55	76	20.67	89.77
Macaws	0.30	22	24.31	84.30	85	84.66	20.33	18	83.98	84.97	83	24.33	18.99
	0.50	19	23.10	82.08	83	81.88	18.67	25	81.85	83.28	82	23.67	24.46
	0.70	23	48.26	79.33	80	80.45	20.00	33	80.52	79.53	78	14.00	36.27
	0.90	28	51.33	76.07	77	76.84	26.33	58	74.07	75.24	77	19.67	29.64
Hats	0.30	21	15.01	84.77	85	86.34	21.33	23	84.34	84.93	86	18.67	20.42
	0.50	20	18.97	82.81	84	84.27	19.00	28	81.67	83.88	85	16.33	21.44
	0.70	22	45.77	80.73	81	82.79	16.33	32	79.19	79.90	82	20.33	33.79
	0.90	25	25.49	77.37	78	73.91	22.00	42	77.20	75.88	76	24.67	62.55

The key metrics, including the iteration count, calculation time, and PSNR for all tested methods, are summarized in Table 3. We only provide visual results for noise levels of 70% and 90% in Figs. 4 and 5 to illustrate the original, noisy, and restored images, to conserve space.

The proposed hRH method demonstrates a marked advantage over the classical CG methods (PRP, LS, and FR) for image denoising. While the LS algorithm provides competitive PSNR values, our hRH method excels in high-noise scenarios, delivering superior or comparable image quality where it matters most. The most significant strength of hRH lies in its convergence efficiency, as it consistently requires a fraction of the iterations needed by the other methods. Although the FR method is frequently faster in terms of computation time, it suffers from poor accuracy. Conversely, PRP and LS are generally slower without consistently outperforming hRH in quality. Overall, the hRH method provides the best trade-off between accuracy and efficiency, proving to be the most robust and reliable algorithm tested, particularly in challenging denoising tasks.

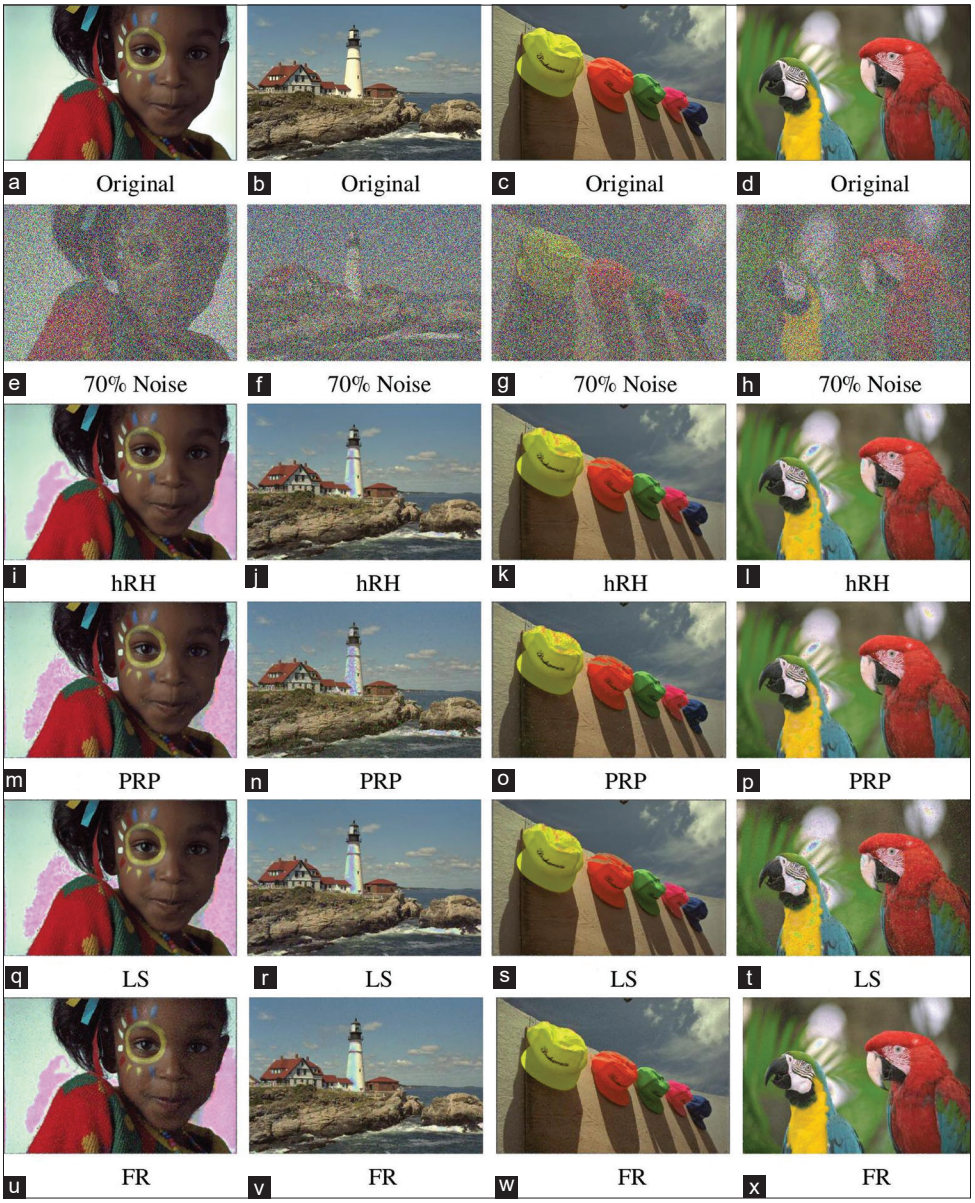


Fig. 4. (a-x) Original images (first row), noisy images with 70% salt-and-pepper noise (second row), and restored results with hRH (row 3), PRP, (row 4), LS (row 5) and FR (row 6).

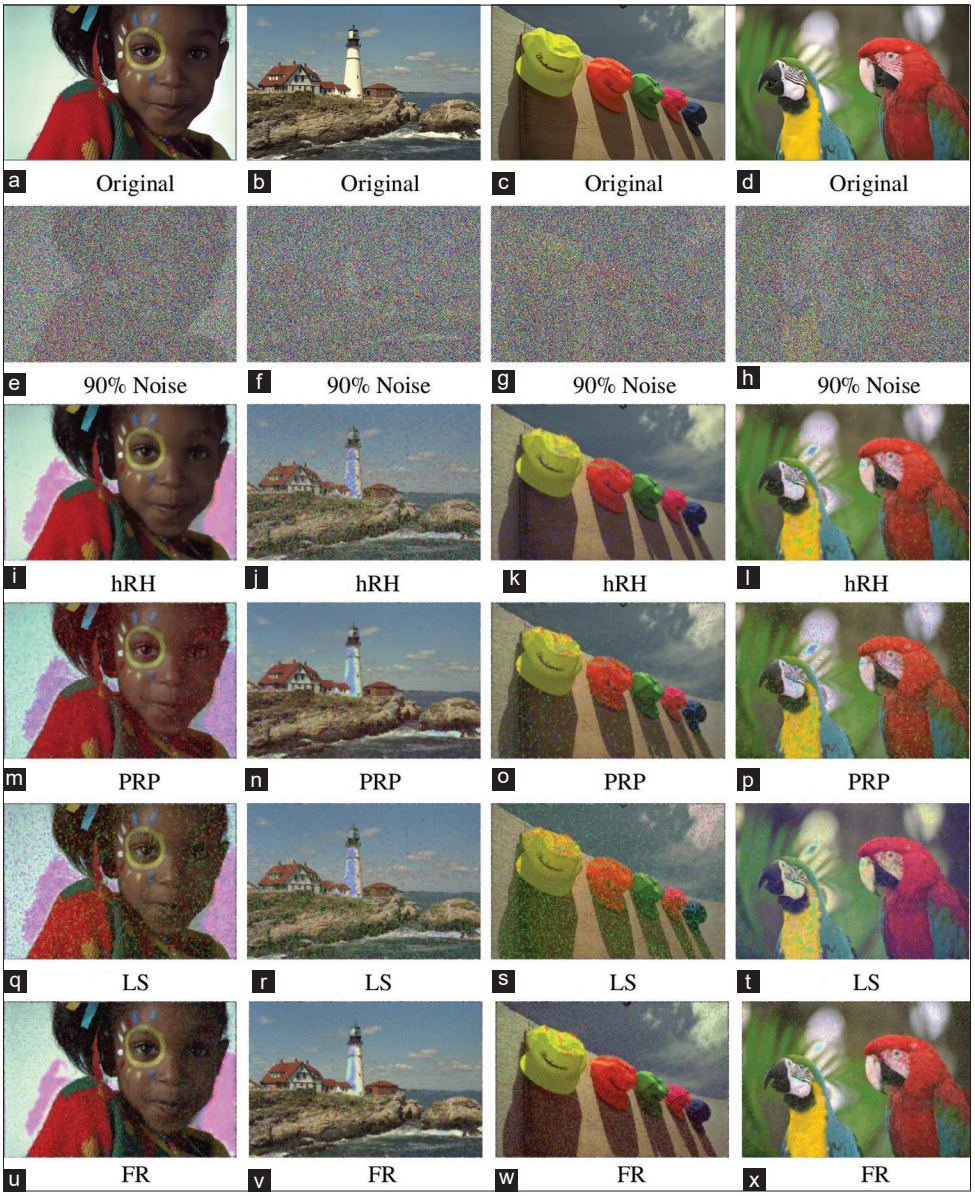


Fig. 5. (a-x) Original images (first row), noisy images with 90% salt-and-pepper noise (second row), and restored results with hRH (row 3), PRP, (row 4), LS (row 5) and FR (row 6).

7. CONCLUSION

In this study, a novel hybrid CG algorithm for unconstrained optimization is proposed and called β_k^{hRH}. The update parameter is a convex combination of the established CG formulas:

$$\beta_k^{\mathrm{hRH}} = \zeta_k \beta_k^{\mathrm{LS}} + \xi_k \beta_k^{\mathrm{FR}} + \omega_k \beta_k^{\mathrm{DY}} + \left(1 - \zeta_k - \xi_k - \omega_k\right) \beta_k^{\mathrm{PRP}},$$

where the coefficients are selected such that they satisfy the Dai-Liao conjugacy condition. The proposed method guarantees descent directions by an inexact line search and demonstrates global convergence under certain general conditions. Computational testing of the hRH method on a range of benchmark problems showed good performance, which was superior to that of conventional CG methods. In addition, the practical feasibility of the proposed method was demonstrated by its successful application to image impulse denoising, where the images were restored clearly despite the preservation of the intrinsic structural information. The above results thus show the versatility and applicability of the proposed method to standard optimization problems and real-world image processing problems.

ACKNOWLEDGMENT

We sincerely thank the anonymous reviewers for their thoughtful comments and insightful suggestions, which helped us improve the clarity and quality of this study. We are also grateful to the editor for their kind support and careful handling of the review process.

REFERENCES

[1] W. W. Hager and H. Zhang. "A survey of nonlinear conjugate gradient methods". *Pacific Journal of Optimization*, vol. 2, no. 1, pp. 35-58, 2006.

[2] M. R. Hestenes and E. Stiefel. "*Methods of Conjugate Gradients for Solving Linear Systems*". vol. 49. NBS, Washington, DC, 1952.

[3] H. A. Babando, M. Barma, S. Nasiru and S. Suraj. "A dai-liao hybrid prp and dy schemes for unconstrained optimization". *International Journal of Development Mathematics*, vol. 1, no. 1, pp. 41-53, 2024.

[4] P. Wolfe. "Convergence conditions for ascent methods". *SIAM Review*, vol. 11, no. 2, pp. 226-235, 1969.

[5] P. Wolfe. "Convergence conditions for ascent methods. II: Some corrections". *SIAM Review*, vol. 13, no. 2, pp. 185-188, 1971.

[6] R. Fletcher and C. M. Reeves. "Function minimization by conjugate gradients". *The Computer Journal*, vol. 7, no. 2, pp. 149-154, 1964.

[7] E. Polak and G. Ribiere. "Note sur la convergence de méthodes de directions conjuguées. Revue française d'informatique et de recherche opérationnelle". *Série Rouge*, vol. 3, no. 16, pp. 35-43, 1969.

[8] B. T. Polyak. "The conjugate gradient method in extremal problems". *USSR Computational Mathematics and Mathematical Physics*, vol. 9, no. 4, pp. 94-112, 1969.

[9] R. Fletcher. "*Practical Methods of Optimization*". John Wiley and Sons, Hoboken, 2000.

[10] Y. Liu and C. Storey. "Efficient Generalized conjugate gradient algorithms, part 1: Theory". *Journal of Optimization Theory and Applications*, vol. 69, pp. 129-137, 1991.

[11] Y. H. Dai and Y. Yuan. "A nonlinear conjugate gradient method with a strong global convergence property". *SIAM Journal on Optimization*, vol. 10, no. 1, pp. 177-182, 1999.

[12] J. Nocedal and S. J. Wright. "*Numerical Optimization*". Springer, Berlin, 2006.

[13] N. Andrei. "Numerical comparison of conjugate gradient algorithms for unconstrained optimization". *Studies in Informatics and Control*, vol. 16, no. 4, pp. 333-352, 2007.

[14] M. J. D. Powell. Nonconvex Minimization Calculations and the Conjugate Gradient Method. In: "*Numerical Analysis: Proceedings of the 10th Biennial Conference Held at Dundee, Scotland, June 28-July 1, 1983*", Springer, pp. 122-141, 1984.

[15] Y. Dai, J. Han, G. Liu, D. Sun, H. Yin and Y. X. Yuan. "Convergence properties of nonlinear conjugate gradient methods". *SIAM Journal on Optimization*, vol. 10, no. 2, pp. 345-358, 2000.

[16] S. Babaie-Kafaki. "A hybrid conjugate gradient method based on a quadratic relaxation of the dai-yuan hybrid conjugate gradient parameter". *Optimization*, vol. 62, no. 7, pp. 929-941, 2013.

[17] N. Andrei. "Another nonlinear conjugate gradient algorithm for unconstrained optimization". *Optimization Methods and Software*, vol. 24, no. 1, pp. 89-104, 2009.

[18] S. S. Djordjević. "New hybrid conjugate gradient method as a convex combination of FR and PRP methods". *Filomat*, vol. 30, no. 11, pp. 3083-3100, 2016.

[19] S. S. Djordjević. "New hybrid conjugate gradient method as a convex combination of LS and FR methods". *Acta Mathematica Scientia*, vol. 39. pp. 214-228, 2019.

[20] J. K. Liu and S. J. Li. "New hybrid conjugate gradient method for unconstrained optimization". *Applied Mathematics and Computation*, vol. 245, pp. 36-43, 2014.

[21] S. B. Hanachi, B. Sellami and M. Belloufi. "New iterative conjugate gradient method for nonlinear unconstrained optimization". *RAIRO-Operations Research*, vol. 56, no. 4, pp. 2315-2327, 2022.

[22] S. B. Hanachi, B. Sellami and M. Belloufi. "A new family of hybrid conjugate gradient method for unconstrained optimization and its application to regression analysis". *RAIRO-Operations Research*, vol. 58, no. 1. pp. 613-627, 2024.

[23] N. Salihu, M. R. Odekunle, A. M. Saleh, S. Salihu. "A dai-liao hybrid hestenes-stiefel and fletcher-revees methods for unconstrained optimization". *International Journal of Industrial Optimization*, vol. 2, no. 1, pp. 33-50, 2021.

[24] A. Hallal, M. Belloufi and B. Sellami. "An efficient new hybrid cg-method as convex combination of DY and CD and HS algorithms". *RAIRO-Operations Research*, vol. 56, no. 6, pp. 4047-4056, 2022.

[25] A. Perry. "A modified conjugate gradient algorithm". *Operations Research*, vol. 26, no. 6, pp. 1073-1078, 1978.

[26] Y. H. Dai and L. Z. Liao. "New conjugacy conditions and related nonlinear conjugate gradient methods". *Applied Mathematics and Optimization*, vol. 43, pp. 87-101, 2001.

[27] A. B. Abubakar and P. Kumam. "A descent dai-liao conjugate gradient method for nonlinear equations". *Numerical Algorithms*, vol. 81, pp. 197-210, 2019.

[28] N. Andrei. "A dai-liao conjugate gradient algorithm with clustering of eigenvalues". *Numerical Algorithms*, vol. 77, no. 4, pp. 1273-1282, 2018.

[29] S. Babaie-Kafaki. "On optimality of two adaptive choices for the parameter of dai-liao method". *Optimization Letters*, vol. 10, pp. 1789-1797, 2016.

[30] S. Babaie-Kafaki and R. Ghanbari. "The dai-liao nonlinear conjugate gradient method with optimal parameter choices". *European Journal of Operational Research*, vol. 234, no. 3, pp. 625-630, 2014.

[31] S. Babaie-Kafaki and R. Ghanbari. "A descent family of dai-liao conjugate gradient methods". *Optimization Methods and Software*, vol. 29, no. 3, pp. 583-591, 2014.

[32] S. Babaie-Kafaki and R. Ghanbari. "Two optimal dai-liao conjugate gradient methods". *Optimization*, vol. 64, no. 11, pp. 2277-2287, 2015.

[33] S. Babaie-Kafaki and R. Ghanbari. "Two adaptive dai-liao nonlinear conjugate gradient methods". *Iranian Journal of Science and Technology, Transactions A: Science*, 42, pp. 1505-1509, 2018.

[34] B. Ivanov, G. V. Milovanović and P. S. Stanimirović. "Accelerated dai-liao projection method for solving systems of monotone nonlinear equations with application to image deblurring". *Journal of Global Optimization*, vol. 85, no. 2, pp. 377-420, 2023.

[35] M. Y. Waziri, K. Ahmed and J. Sabi'u. "A dai-liao conjugate gradient method via modified secant equation for system of nonlinear equations". *Arabian Journal of Mathematics*, vol. 9, pp. 443-457, 2020.

[36] A. B. Abubakar, P. Kumam, M. Malik and A. H. Ibrahim. “A hybrid conjugate gradient based approach for solving unconstrained optimization and motion control problems”. *Mathematics and Computers in Simulation*, vol. 201, pp. 640-657, 2022.

[37] A. B. Abubakar, P. Kumam, M. Malik, P. Chaipunya and A. H. Ibrahim. “A hybrid FR-DY conjugate gradient algorithm for unconstrained optimization with application in portfolio selection”. *AIMS Mathematics*, vol. 6, no. 6, pp. 6506-6527, 2021.

[38] G. Zoutendijk. “Nonlinear programming, computational methods”. *Integer and Nonlinear Programming*”, pp. 37-86, 1970.

[39] N. Andrei. “An unconstrained optimization test functions collection”. *Advanced Modeling and Optimization*, vol. 10, no. 1, pp. 147-161, 2008.

[40] N. I. M, Gould, D. Orban and P. L. Toint. “Cuter and sifdec: A constrained and unconstrained testing environment, revisited”. *ACM Transactions on Mathematical Software*, vol. 29, no. 4, pp. 373-394, 2003.

[41] J. J. Moré, B. S. Garbow and K. E. Hillstrom. “Testing unconstrained optimization software”. *ACM Transactions on Mathematical Software*, vol. 7, no. 1, pp. 17-41, 1981.

[42] E. D. Dolan and J. J. Moré. “Benchmarking optimization software with performance profiles”. *Mathematical Programming*, vol. 91, pp. 201-213, 2002.

[43] R. H. Chan, C. W. Ho and M. Nikolova. “Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization”. *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1479-1485, 2005.

[44] J. F. Cai, R. Chan and B. Morini. Minimization of an Edge-Preserving Regularization Functional by Conjugate Gradient Type Methods. In: “*Image Processing Based on Partial Differential Equations: Proceedings of the International Conference on PDEBased Image Processing and Related Inverse Problems*”, CMA, Oslo, Springer, pp. 109-122, 2007.

[45] H. Hwang and R. A. Haddad. “Adaptive median filters: New algorithms and results”. *IEEE Transactions on Image Processing*, vol. 4, no. 4, pp. 499-502, 1995.

ORIGINAL ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

An Ensemble-based Machine Learning Framework for Advanced Distributed Denial of Service Attack Detection in Software Defined Networks

Aram Saleem[1], Hakem Beitollahi[2]

[1]MSc Student, Department of Computer Science, Soran University, Soran, Iraq, [2]Assistant Professor, Department of Computer Science, Soran University, Soran, Iraq

ABSTRACT

Distributed Denial of Service (DDoS) attacks pose a significant threat to modern network architectures, especially Software Defined Networking (SDN) due to its centralized controller. This study proposes an advanced framework for DDoS attack identification and prediction using state-of-the-art machine learning (ML) techniques in an SDN architecture. A comprehensive dataset was generated through a two-stage traffic generation procedure, simulating attack and normal scenarios over a 6-day period, from which fifteen were extracted to characterize network behavior. Multiple classifiers including Gradient Boosting Ensemble methods such as LightGBM, XGBoost, CatBoost, and Gradient Boosting Decision Trees, as well as additional ensemble methods such as AdaBoost and Bagging were evaluated alongside with One-Class SVM and Bayesian Networks. They were trained and evaluated using rigorous cross-validation. The results demonstrate near-perfect performance of ensemble models, achieving up to 99.98% accuracy with outstanding precision, recall, and area under curve metrics. To achieve efficient mitigation, the detection mechanism is deployed on local web servers, and a certificate authority-based secure communication channel transmits malicious IPs to the SDN controller, enabling low-latency, scalable, and real-time DDoS attack mitigation. This paper discusses the promise of applying cutting-edge ML models to enhance the robustness of SDN infrastructures against sophisticated cyber-attacks and offers a template for further research in dynamic network defense strategies.

Index Terms: Software Defined Networking, Distributed Denial of Service, Machine Learning, Ensemble Models, Traffic Classification

1. INTRODUCTION

Distributed denial of service (DDoS) attacks have evolved into a critical threat to the stability and security of modern network infrastructures [1], [2]. These attacks aim to render services unavailable to legitimate users by overwhelming network resources with malicious traffic. The impact of DDoS attacks extends beyond service disruption, causing significant financial losses, reputational damage, and disruptions across various sectors, including finance, healthcare, and cloud computing. Consequently, network administrators face increasing challenges in mitigating the sophistication and multi-layered nature of these attacks. Software Defined Networking (SDN) has revolutionized network management by decoupling the control plane from the data plane, enabling centralized control, programmability, and dynamic resource allocation [3], [4]. This architectural shift offers significant benefits, such as enhanced scalability,

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp184-197

E-ISSN: 2521-4217
P-ISSN: 2521-4209

Copyright © 2025 Saleem and Beitollahi. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Hakem Beitollahi, Assistant Professor, Department of Computer Science, Soran University, Soran, Iraq
hakem.beitollahi@soran.edu.iq

Received: 19-06-2025 Accepted: 01-09-2025 Published: 04-10-2025

adaptability, and operational efficiency, making SDN well-suited for complex network environments such as data centers and enterprise systems [3], [4]. However, the inherent centralization of SDN introduces critical security vulnerabilities. Specifically, the SDN controller acts as a single point of failure, the network's central control point. DDoS attacks targeting the controller can easily exhaust its resources, leading to network-wide outages. Fig. 1 illustrates the DoS attack on various planes of a SDN network. This vulnerability underscores the urgent need for effective DDoS attack detection and mitigation mechanisms within SDN environments [5], [6].

Traditional security controls, such as firewalls, signature-based intrusion detection systems, and rate-limiting techniques, often demonstrate limited effectiveness against contemporary DDoS attacks. These methods frequently fail to detect attacks that exploit protocol vulnerabilities or mimic legitimate traffic patterns. Furthermore, reactive security measures are typically implemented after significant damage has already occurred. In high-availability environments, where minimal downtime is crucial, such reactive approaches are inadequate [7], [8].

ML has emerged as a transformative approach to DDoS attack detection and prediction. Unlike traditional security solutions, ML algorithms analyze extensive network traffic data to identify patterns and anomalies that indicate malicious activity, including 0-day exploits and adaptive attack strategies [9]. Advanced techniques, such as deep learning and ensemble models, can detect subtle deviations from normal network behavior by analyzing complex features such as packet inter-arrival times and flow duration [10]. Furthermore, the predictive capabilities of ML enable proactive mitigation, allowing network administrators to neutralize threats before they escalate into full-scale attacks [6], [11].

Integrating ML-based detection mechanisms within SDN environments facilitates dynamic network responses to evolving attack patterns [12]. The centralized control of SDN allows ML models to provide real-time insights, enabling automated decisions that enhance network resilience and security [13], [14].

This study aims to address the aforementioned challenges by proposing a robust framework for DDoS attack detection and prediction specifically designed for SDN environments. The study's objectives are threefold:

1. Creation of a Comprehensive Statistical Dataset: To develop and publicly share a realistic dataset comprising both normal and attack traffic patterns relevant to SDN environments, thereby facilitating further research in this area.
2. Evaluation of ML Models: To implement and evaluate advanced ML techniques, including deep learning architectures and ensemble methods, to assess their effectiveness in real-time DDoS attack detection and prediction

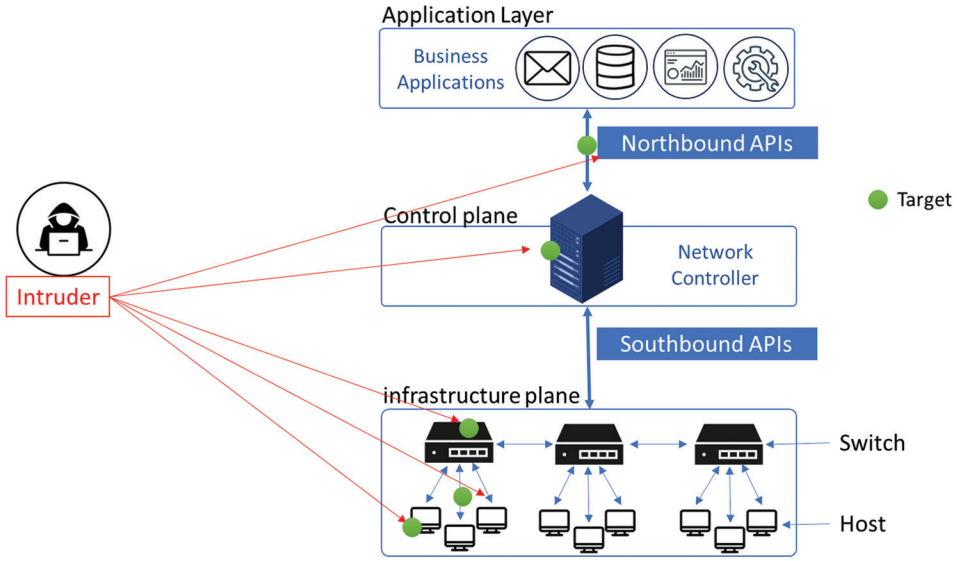


Fig. 1. Illustration of a denial of service attack targeting components of the Software Defined Networking architecture, including the centralized controller and data plane.

3. Comparative Analysis of Performance: To analyze the performance of selected ML models using key metrics such as accuracy, precision, computational efficiency, and scalability.

The main contributions of this study are:

- Realistic Dataset Creation: A publicly available dataset specifically designed for SDN environments, addressing the existing gap in realistic and diverse traffic patterns.
- Insights into ML Techniques: A comprehensive analysis of advanced ML methods, providing actionable insights into their respective strengths and weaknesses for DDoS attack detection.
- Enhanced Defense Framework: A practical, scalable, and adaptive solution for real-time DDoS attack mitigation, enhancing the resilience and reliability of SDN networks.

The rest of the paper is organized as follows: Section 2 reviews the related work with respect to the current techniques of DDoS detection and mitigation concerning the SDN environments and the respective limitations of those techniques. Section 3 describes in detail the proposed framework that also covers the creation of a dataset, ML methodologies, and how these techniques will be integrated into the SDN architecture. Section 5 discusses the experimental results, including the performance evaluation of different models and their comparison. Moreover, the discussion of the results, implications, and limitations is also discussed in this section. Finally, Section 6 concludes the paper.

2. RELATED WORKS

Numerous studies have explored the detection and mitigation of DDoS attacks in SDN environments using ML, deep learning, or hybrid techniques [15]–[18]. Although these methods demonstrate potential, they still exhibit certain limitations that necessitate further investigation.

Garba *et al.* [13] developed a framework that integrated ML-based detection into signature-based intrusion detection systems, achieving 99% classification accuracy with Decision Tree algorithms. However, their approach heavily relies on predefined attack signatures and supervised learning models, which significantly limits its adaptability to novel attack patterns not present in the training data.

Anley *et al.* [2] proposed an adaptive transfer learning-based CNN approach that improved detection accuracy across datasets. Nevertheless, this method incurs high computational overhead and requires extensive hyperparameter optimization, rendering it less suitable for real-time applications, particularly in resource-constrained environments.

Swami *et al.* [14] introduced an interquartile range threshold-based statistical approach for detecting spoofed flooding DDoS attacks. Although this method demonstrated reduced detection times, it struggles to accurately identify sophisticated attacks that closely resemble legitimate traffic patterns in dynamic scenarios, leading to a higher rate of false positives (FP). The reliance on static thresholds also limits the method's adaptability across diverse network environments. Al-Fayoumi and Abu Al-Haija [1] presented a lightweight detection model for MQTT-based low-rate DDoS attacks, achieving high accuracy. However, this model's narrow focus on specific attack types restricts its scalability and applicability to other DDoS categories.

Deep learning-based techniques are increasingly being utilized for DDoS detection in SDN environments. Clinton *et al.* [10] transformed network traffic into image data and achieved over 99% classification accuracy using image classification methods. However, this approach introduces significant computational complexity. The conversion of raw traffic to image formats increases processing time and is resource-intensive, making it unsuitable for high-throughput networks.

Songa and Karri [19] proposed a unified SDN framework employing feature elimination, clustering, and timeseries analysis, achieving 99.92% detection accuracy. Nevertheless, their work heavily relies on the CICDDoS2019 dataset, raising concerns about its generalization capability for practical implementation in diverse traffic scenarios.

Hybrid models that integrate multiple ML/DL techniques show considerable promise. For example, Mhamdi and Isa [6] explored a hybrid approach using a Deep Autoencoder in conjunction with a Random Forest classifier, achieving an anomaly detection rate exceeding 98%. However, the reliance on a centralized detection mechanism renders the system susceptible to controller-specific DDoS attacks, highlighting a critical vulnerability in SDN environments where the controller is a primary target. Furthermore, the substantial computational demands of hybrid approaches raise concerns about their scalability and feasibility for implementation in large, resource-constrained networks.

While ML-based approaches offer significant promise for DDoS attack detection, several challenges hinder their effective implementation within SDN environments. These challenges include high FP rates in anomaly detection, limited scalability of centralized detection, susceptibility of the SDN controller to targeted floods, and the inability of static models to adapt to evolving attack patterns in real-time. A key limitation is the scarcity of comprehensive and up-to-date datasets that accurately represent the unique characteristics of SDN traffic and modern DDoS attack strategies [10], [13]. Existing datasets often lack completeness or fail to encompass the wide range of attack scenarios, which limits the generalization capability of ML models. Furthermore, many studies do not evaluate advanced ML techniques in real time, dynamic SDN environments, resulting in substantial gaps in practical applicability [7], [20].

To address these limitations, our research focuses on the real-time deployment of trained ML models within an SDN architecture. In contrast to previous work, this study emphasizes scalable and adaptive detection mechanisms that operate efficiently in dynamic and heterogeneous network environments. The contributions presented in the following sections aim to mitigate the limitations of dataset dependency, centralized detection, and resource-intensive approaches without compromising detection accuracy and efficiency.

3. METHODOLOGY

This section details the design and implementation of our proposed framework for the detection of DDoS attacks within an SDN environment. The framework comprises an SDN controller, a comprehensive dataset encompassing both normal and malicious network traffic, a carefully selected set of statistical features to characterize traffic behavior, and various ensemble ML modules that utilize these features for real-time traffic classification. The subsequent subsections will explain the specifics of the SDN environment setup, the two-stage dataset creation methodology, the feature extraction process, and the integration of the ML-based classification engine with the SDN controller.

The key stages of the methodology are visually represented in the diagram depicted in Fig. 2, which highlights the sequential flow from the SDN environment setup to the immediate network-wide blocking of malicious traffic.

a. SDN setup: Depicts the "Ryu" controller, OpenFlow switch, and host configuration, including a dedicated web server.

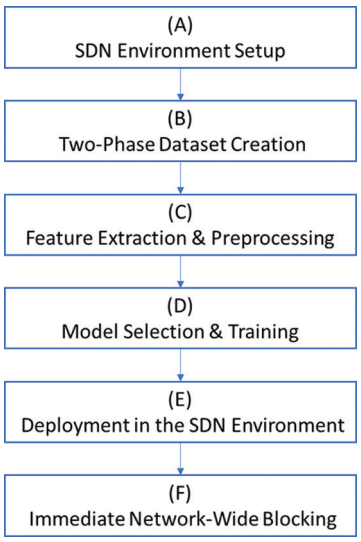


Fig. 2. Overview of the proposed methodology for distributed denial of service detection in Software Defined Networking, including dataset generation and model integration.

b. Two-phase dataset creation: Demonstrates the sequential generation of normal traffic and subsequent injection of malicious traffic, with packet captures labeled to create a comprehensive dataset.
c. Feature extraction and preprocessing: Details the computation of sixteen key features, followed by data cleaning and scaling procedures.
d. Model selection and training: Describes hyperparameter tuning, cross-validation techniques, and the selection of the optimal ML model from multiple candidates.
e. Deployment in SDN: Explains the distributed deployment of local ML classifiers on each server, the certificate authority (CA)-based encryption of IP lists, and their transmission to the controller for rule enforcement.
f. Immediate blocking: Illustrates the controller's rapid deployment of drop rules to the OpenFlow switch to mitigate attacks network-wide.

3.1. The Proposed Framework

The proposed framework leverages the centralized control plane paradigm inherent in SDN. The "Ryu" controller is employed in this study because of its Python-based modularity, adaptability for custom application development, and interoperability with various OpenFlow-enabled switches. The experimental network topology comprises multiple virtual hosts, a single OpenFlow-compliant switch, and the "Ryu" controller, which manages all flow rules. Specifically, one host functions as a web server hosting a basic HTTP application, while the remaining hosts generate both normal

and malicious traffic. OpenFlow facilitates communication between the switch and the hosts. The switch transmits flow-level statistics to the "Ryu" controller, which, in turn, dictates forwarding policies, installs flow rules, and implements traffic shaping or blocking. This centralized control plane empowers network administrators to enforce granular security policies and perform real-time anomaly detection. Furthermore, the host with interface "h2-eth0" is designated as a monitoring point to observe traffic, analyze packets, and relay aggregated statistics to the classification module. This design aims to provide the "Ryu" controller with comprehensive network awareness, enabling it to apply immediate countermeasures upon detecting malicious traffic sources.

To accurately characterize the behavior of both normal and attack traffic, the system extracts 15 statistical features from observed flows. Following each capture period, the packet capture script gathers and normalizes a range of metrics for each active source IP (excluding the web server). These metrics are then transformed into a set of features that serve as the basis for traffic classification. Each feature is designed to act as a potential indicator for distinguishing between normal and malicious DDoS traffic.

The 15 attributes constitute a comprehensive set of measures intended to quantify both high-level and fine-grained characteristics of network traffic. Specifically, rate-based attributes (e.g., Request Rate, Packet Arrival Rate) are used to identify volumetric traffic surges, while connection-oriented metrics (e.g., TCP SYN Count, Failed Connection Attempts) are used to detect manipulations of the TCP handshake process commonly exploited in DDoS attacks. Furthermore, distribution-based attributes (e.g., Unique IP Count, Port Usage Distribution) provide insights into the scope of malicious host scanning activities. Each attribute's mathematical derivation, practical significance, and specific utility in DDoS detection and mitigation are detailed in the subsequent subsections.

1. Request rate: The Request Rate measures the number of application-level requests (e.g., HTTP, FTP, or other protocol-specific requests) made by a source IP per unit time. Formally,

$$R_{req} = \frac{N_{req}}{T_{elapsed}} \tag{1}$$

Where N_{req} is the total number of requests observed during the capture interval and Telapsed is the duration of the interval (in seconds). High request rates are generally a sign of application-layer DDoS (e.g., HTTP floods). By monitoring (R_{req}), it is straightforward to detect hosts generating more requests than normal.

2. Packet arrival rate: The packet arrival rate (R_{pkt}) quantifies how frequently a host transmits packets in seconds:

$$R_{pkt} = \frac{N_{pkt}}{T_{elapsed}} \tag{2}$$

In volumetric DDoS attacks, attackers flood the network with an unusually large volume of packets. A sudden surge in (R_{pkt}) thus indicates beforehand that there can be a flood, marking normal bursts apart from more sustained malicious traffic.

3. Download rate: The download rate (R_{bytes}) is the average rate of data being transferred from a host over time:

$$R_{bytes} = \frac{B_{total}}{T_{elapsed}} \tag{3}$$

Here, B_{total} is the quantity of bytes received by the host during the capture window. Bandwidth-depletion DDoS attacks attempt to exhaust a victim's network link capacity by transferring large volumes of data. Tracking (R_{bytes}) helps identify this behavior.

4. Uptime: The uptime (U_{time}) measure indicates that the duration a host has been actively sending traffic during the period observed:

$$U_{time} = T_{last_packet} - T_{first_packet} \tag{4}$$

DDoS attacks in other instances are composed of high activity sustained over an extended period of time. Alternatively, legitimate clients will have more sporadic or short sessions. Thus, continually high Uptime can be an indicator of ongoing attack traffic as opposed to regular user-initiated behavior.

5. Flow duration variability: (V_{flow}) is the standard deviation of flow durations across hosts. If each flow i has duration Di, and $\bar{D}$ is their mean, then:

$$V_{flow} = \sqrt{\frac{1}{N_{flow}} \sum_{i=1}^{N_{flow}} (D_i - \bar{D})^2} \tag{5}$$

Although there are legitimate long-lived and persistent flows (e.g., media streaming), the flows that result from

DDoS attacks are normally high rates of extremely short or unusual flows. Therefore, analyzing (Vf low) reveals patterns indicating bursty or inconsistent traffic typical of malicious activity

6. Retransmission rate: The retransmission rate ($R_{retrans}$) is the amount of packets retransmitted by a host per second:

$$R_{bytes} = \frac{N_{retrans}}{T_{elapsed}} \tag{6}$$

Misbehaving activities such as SYN floods or poorly misconfigured bots could generate a lot of duplicate packets or half-opened connections. An elevated ($R_{retrans}$) can signal malicious floods or repeated attempts at scanning for a victim server without completing up connections.

7. Unique IP count: The unique IP count ($C_{Unique\ IP}$) reveals how many distinct destination IP addresses a source host target:

$$C_{Unique_IP} = \left|\left\{IP_{dest}\right\}\right| \tag{7}$$

In certain DDoS cases (or scanning attacks), an aggressive host rapidly probes many different IPs. However, a normal user tends to connect to fewer places in a short span of time, thus making an unusually high (CUnique IP) suspicious.

8. Sequential request patterns: Sequential Request Patterns (S seq) capture repeated requests to the same port from one host. Specifically,

$$S_{seq} = Max\left(\left\{N_{req}(P)\right\}\right) \tag{8}$$

Where Nreq(P) is the frequency of a host's repeated requests for the same port PPP consecutively. Automated scripts usually use repeated instances of the same actions that repeatedly hammer the same service endpoint a classic indication of possible DDoS or brute-force attacks.

9. Flow count per host: The flow count per host (Nflows) is the sum of flows originated by a source:

$$N_{flows} = \sum_{i-1}^{N_{sessions}} 1 \tag{9}$$

Malicious hosts may establish dozens or hundreds of ephemeral flows in an attempt to overwhelm a server's bandwidth. High rates of flows over a brief period are therefore an extremely strong indication of possible DDoS or network scanning.

10. Host communication frequency: Host Communication Frequency (fcomm) measures how many unique hosts a source communicates with per second:

$$f_{comm} = \frac{N_{hosts}}{T_{elapsed}} \tag{10}$$

This feature indicates whether a source is quickly propagating traffic to multiple targets (e.g., in a spread-based attack or scanning approach). In benign use cases, one typically sees smaller values that rise more slowly.

11. Port usage distribution: The Port Usage Distribution (Dports) gives the number of distinct destination ports utilized:

$$D_{ports} = |\{P_{dest}\}| \tag{11}$$

Attackers might attempt multi-port scanning or flooding to identify open services to attack. If one source is connecting to an unusually high number of different ports, it is either malicious probing or an advanced DDoS attack targeting multiple services.

12. TCP SYN count: The TCP SYN Count (Nsyn) measures how many TCP SYN packets a host sends.

$$N_{syn} = \sum_{i-1}^{N_{packets}} \delta_{SYN}(i) \tag{12}$$

Where δSYN (i) is 1 if the i-th packet is a SYN flag set, and 0 otherwise. SYN flood attacks are one of the most common types of DDoS, therefore large SYN counts within a short time interval a sure sign of potential flooding.

13. TCP ACK count: The TCP ACK Count (NACK) is the analogous measure for ACK packets:

$$N_{ACK} = \sum_{i-1}^{N_{packets}} \delta_{ACK}(i) \tag{13}$$

Although the ACK flag alone is not always malicious, analyzing its relationship to SYN packets (e.g., the ratio of SYN to ACK) can highlight incomplete handshakes or unusual session behaviors that typify DDoS attempts.

14. Connections per second: Connections per second (Rconn) is the sum of new connections established by a host in one second:

$$R_{conn} = \frac{N_{new_conn}}{T_{elapsed}} \quad (14)$$

Flooding attackers or active scanners have the tendency to establish many connections in quick succession to overload a target. Thus, a high (Rconn) distinguishes probable attackers from valid clients, which establish connections at a slower pace.

15. Failed connection attempts: Failed Connection Attempts (N_{failed}) refer to the number of connections attempted which fail to finish the TCP handshake. Monitoring can be done as follows:

$$N_{failed} = \frac{Failed\ Attempts}{Telpased} \quad (15)$$

Where NSYN is the number of SYN packets that were sent, and NSYN_ACKN is the number that was responded to with a SYN-ACK. In partial-handshake attacks (e.g., SYN floods), the attacker never finishes the server's response, so there are many incomplete or failed sessions.

The selection of features for this study was guided by a literature review as well as domain knowledge pertaining to DDoS attack patterns. Request Rate, Packet Arrival Rate, TCP SYN Count, and Port Usage Distribution are a few of the features used extensively in previous studies on DDoS detection in SDN and traditional networks [2], [13], [21], [22]. But on top of that, we also have some sophisticated or innovative features such as Sequential Request Patterns and Failed Connection Attempts that exceed the typical ones in benchmark research. The reasoning behind these features is to attack application-layer session behavior and incomplete handshake patterns, especially relevant to SDN environments but commonly omitted in earlier research works. That combined approach provides more comprehensive and SDN-focused detection capacity.

3.2. Dataset Creation

The development of a labeled dataset is crucial for the effective training and evaluation of ML models in DDoS attack detection. To generate realistic traffic data, we implemented a two-phase approach over 6 days: An initial phase for normal traffic generation, followed by a phase dedicated to creating malicious traffic.

As shown in Fig. 3, the normal traffic phase spanned 2 days, during which typical user interactions with the web server were simulated. 11,186 aggregated samples which users sent normal traffic to the web server during these 2 days. These activities included browsing, page requests, occasional form submissions, and SSH interactions. Concurrently, background operations such as normal file downloads and MySQL queries were initiated to emulate a more diverse network environment.

Upon capturing sufficient normal traffic data, the network configuration was transitioned to a malicious environment, which was sustained for 4 consecutive days. During this phase, various DDoS attack techniques were employed. A custom, multi-threaded Python script was utilized to generate continuous multi-threaded HTTP floods targeting the web server. In addition, high-volume SYN floods were launched using tools such as hping3 and iperf, alongside application layer resource exhaustion attacks. These attacks were designed to mimic common DDoS characteristics, 5 such as rapid request generation, incomplete TCP handshakes, and excessive consumption of server resources. In total, 11,469 aggregated samples in which bot machines participate in creating malicious traffic during the 4 days.

Throughout both phases, packet capture was consistently performed on the monitor host using Scapy. This ensured that all network activity originating from a source IP within a defined time window was recorded. Each captured network session was then labeled as either normal (0) or attack (1), resulting in a comprehensive dataset of benign and malicious network traffic.

The complete dataset used in this study is publicly available at https://www.kaggle.com/datasets/aramsaleem21/sddos-sdn-2025.

To clarify our choice of using a custom dataset instead of public alternatives like CICDDoS2019. A few of the engineered features used in this study are not computable or aggregable directly from the CICDDoS2019 dataset due to the granularity and content limitations of the available data. For example, Request Rate (req/sec) and Sequential Request Patterns depend on the application layer to recognize and count the distinct user interactions and their sequence. CICDDoS2019 does not have this application-level visibility.

Retransmission Rate depends on monitoring TCP sequence numbers to detect duplicate transmissions, which is not feasible without raw packet captures and sequence monitoring

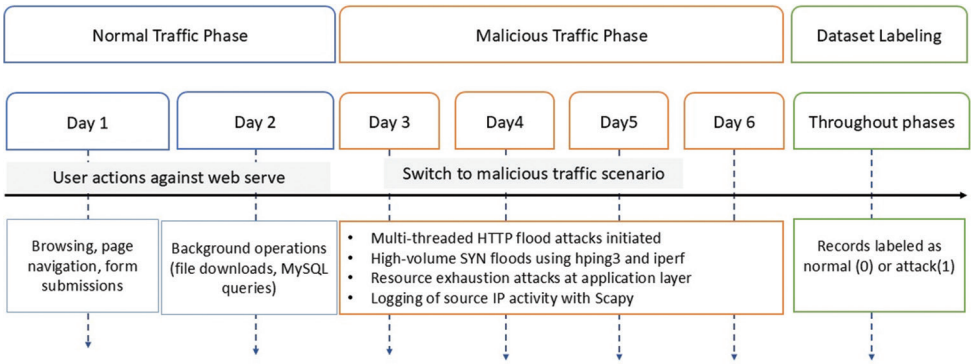


Fig. 3. (Dataset creation) highlights the two-phase normal vs. malicious data gathering.

not available in CICDDoS2019. Failed Connection Attempts Rate must maintain connection state on a per-host basis, i.e., unmatched SYNs or aborted handshakes through RST/FIN flags. CICDDoS2019 does not maintain this level of TCP control information.

In contrast, our dataset was built from full packet captures of a controlled SDN setup and captures both application and transport-layer behavior for a diverse set of traffic scenarios. This provides the ability to extract more expressive, real-time, and SDN-specific traffic features and hence provides a more fertile ground for evaluating advanced machine learning (ML) models under conditions more closely aligned with real-world network operation.

Following the collection of a comprehensive dataset comprising DDoS and normal traffic, the subsequent step involved data preprocessing. This process transformed raw packet-level data into a structured dataset suitable for ML-based classification. The following subsections detail the preprocessing and validation procedures.

1. Data cleaning: It includes two stages. (a) Removal of Duplicates: Overlapping capture intervals and sniffer buffering occasionally resulted in duplicate entries. These were identified and removed using (Timestamp, Source IP, Destination IP) tuples. (b) Handling Missing Values: Records with undetermined features due to missing session information were either dropped or imputed using the mean/median, depending on whether the ratio of missing data was low (1%).
2. Label encoding (binary labels): Each traffic sample was labeled as "0" for benign or "1" for attack. To mitigate potential label noise (e.g., suspicious-looking but benign traffic), cross-validation was performed using known malicious tools (iperf, hping3) during the attack phase.
3. Normalization/scaling: It includes two steps. (a) Feature scaling: To prevent feature scale sensitivity in certain ML algorithms, MinMax or Standard scaling was applied where necessary. This ensured that no single feature, such as raw byte counts, unduly influenced model training. (b) Outlier Treatment: Extreme values (e.g., an extremely high Packet Arrival Rate from a short, intense attack) can skew model training. While outliers can indicate attacks, they were capped at a maximum value (e.g., the 99.9th percentile) to stabilize training without eliminating the underlying malicious patterns.

This preprocessing pipeline ensured the final dataset's cleanliness, consistency, and appropriate scaling before classification.

3.3. Model Selection and Training Strategy

To comprehensively evaluate the effectiveness of ML techniques for DDoS detection, we selected eight classification models:

- Gradient boosting ensembles: LightGBM, XGBoost, CatBoost, and Gradient Boosting Decision Trees. These tree-based ensembles are recognized for their strong performance in handling diverse feature sets, robustness against overfitting due to boosting mechanisms, and interpretability of feature importance.
- Ensemble methods: AdaBoost and Bagging (Bootstrap Aggregating). While conceptually related to gradient boosting, these methods typically combine multiple weak estimators (e.g., Decision Trees) to achieve a favorable bias-variance trade-off.
- One-class SVM: A classical anomaly detection algorithm that learns the boundary of normal data to identify outliers. Although theoretically appealing for 0-day DDoS attack detection, it may exhibit a high FP rate if legitimate traffic deviates significantly from the training reference.
- Bayesian networks: A graphical probabilistic model that represents dependencies among features. While

often surpassed by ensemble methods in high-volume applications, Bayesian networks offer explainable insights into conditional probabilities of attack versus benign behavior.

The preprocessed dataset was partitioned into training (70%), validation (15%), and test (15%) sets. This partitioning strategy ensures an unbiased evaluation of model performance on unseen data.

Within the training set, k-fold cross-validation was employed. In each iteration, a model was trained on k-1 folds, and performance metrics were computed on the remaining fold. This process was repeated k times, and the resulting average metrics mitigate variance associated with a single train-test split.

We optimized the hyperparameters of each model to maximize performance. For Gradient Boosting models, a grid search was conducted across the number of estimators, learning rate, maximum tree depth (max depth), and subsample ratio to identify optimal configurations. For AdaBoost and Bagging, we varied the base estimator depth and the number of estimators to achieve a balance between model complexity and overfitting. For the One-Class SVM, we experimented with RBF and linear kernels and tuned the "nu" parameter, which controls the fraction of outliers. Finally, for Bayesian Networks, we explored score-based and constraint-based structure learning algorithms and employed partial hill-climbing to identify effective network structures. The accuracy, area under curve (AUC), and F1-score metrics guided the selection of optimal hyperparameter configurations. The final model configurations were then retrained using the combined training and validation sets and evaluated on the held-out test set.

4. IMPLEMENTATION IN THE SDN ENVIRONMENT

During the deployment phase, each web server within the SDN environment is equipped with its own ML model for DDoS detection. Upon detecting an attack signature, a web server promptly notifies the SDN controller. This design choice enhances scalability by distributing computational load to the servers and reduces detection latency through real-time classification of incoming traffic at the network edge, enabling instantaneous detection and immediate reporting of malicious IP addresses.

When a server identifies an IP address as malicious, the SDN controller immediately enforces drop rules across the network to mitigate potential damage. To ensure secure communication of identified malicious IP addresses, the web servers utilize a CA-based security mechanism, restricting participation in the detection and mitigation process to trusted entities. The SDN controller then installs drop rules on the OpenFlow switches throughout the network, effectively isolating attackers from all hosts within the SDN environment. This distributed deployment of classifiers serves two primary functions:

1. Controller load reduction: By delegating attack detection to individual web servers, the SDN controller is relieved of the burden of inspecting every request. This architecture minimizes the risk of overwhelming the controller with high-volume traffic and eliminates latency associated with packet forwarding to and from a centralized analysis unit.
2. Scalability and low-latency detection: As the number of web servers increases (e.g., due to data center expansion), the number of detection nodes scales proportionally. Each server performs traffic classification independently, imposing minimal overhead on the central controller and facilitating real-time or low-latency detection.

4.1. Local ML Detection

Each web server continuously monitors incoming traffic and analyzes features such as request rate, packet size, and unusual TCP flags. Upon detecting suspicious activity within a "five-second" window, the local ML model generates a list of the responsible IP addresses. This localized detection approach ensures rapid response times and avoids bottlenecks associated with centralized traffic analysis.

4.2. Secure IP List Transmission with CA-Based Sessions

Before transmitting the list of malicious IP addresses to the SDN controller, the web server establishes a secure communication channel using CA-issued certificates. The session establishment process involves the server first validating the controller's certificate (issued by a trusted CA). Subsequently, the server encrypts a randomly generated session key with the controller's public key. The controller then decrypts this encrypted session key using its corresponding private key, thereby acquiring the session key required for symmetric encryption. Finally, the server encrypts the list of malicious IP addresses using AES-256 encryption with the established session key, ensuring confidentiality and integrity during transmission.

4.3. Immediate Network-Wide Blocking

Once the IP list is decrypted, the SDN controller enforces network-wide blocking by commanding the OpenFlow

switches to discard all packets from the labeled IP addresses. The blocking occurs in a short time and thus prevents malicious hosts from further flooding or compromising network resources. Because each web server handles traffic classification independently, the network gains horizontal scalability: Adding more servers proportionally enhances the overall detection capability of the network, while the central controller remains focused on coordinating drop rules rather than performing high-volume traffic analysis.

Fig. 4 provides a more step-by-step illustration of the entire process, from when an attack is identified by a server to where the bad traffic is dropped in the network fabric.

In such an architecture, real-time detection at the server level prevents making the SDN controller a computational bottleneck. Whereas CA-based session establishment ensures mutual authentication and secure key exchange, AES256 encryption protects the IP lists against eavesdropping and tampering. With fast blocking of malicious traffic at the switch level, the entire network becomes resilient against DDoS disruption, illustrating how a distributed model of detection nicely complements SDN's centralization of policy enforcement.

5. RESULTS

This section outlines the results obtained from evaluating the eight ML models on the SDN-based DDoS dataset. Model performance is assessed using a variety of metrics, and further analysis is conducted using confusion matrices and ROC curves. The findings provide a comparative analysis of each model's ability to effectively detect and classify DDoS attacks. Next, a comprehensive discussion of the results is presented.

5.1. Overall Classification Outcomes

Table 1 provides a comparative overview of the performance of the eight ML models, evaluated on the SDN-based DDoS dataset. The performance of each model was assessed using the following metrics: accuracy, precision, recall, F1-score, AUC, Cohen's Kappa, Matthews correlation coefficient (MCC), FP rate (FPR), false-negative rate (FNR), and cross-validated accuracy.

5.2. Confusion Matrices: Minimal Misclassifications

To better visualize model performance, confusion matrices were generated for each of the eight classifiers as shown in Fig. 5. These matrices illustrate the distribution of true positives, FP, true negatives (TN), and false negatives (FN). The confusion matrices highlight that ensemble methods, particularly gradient boosting techniques, achieve a strong balance between high detection rates and low false alarm rates.

5.3. Cross-Validation and Consistency

Cross-validation accuracies remain consistently above 0.998 for LightGBM, CatBoost, Bagging, and XGBoost. This consistency across different folds indicates that their

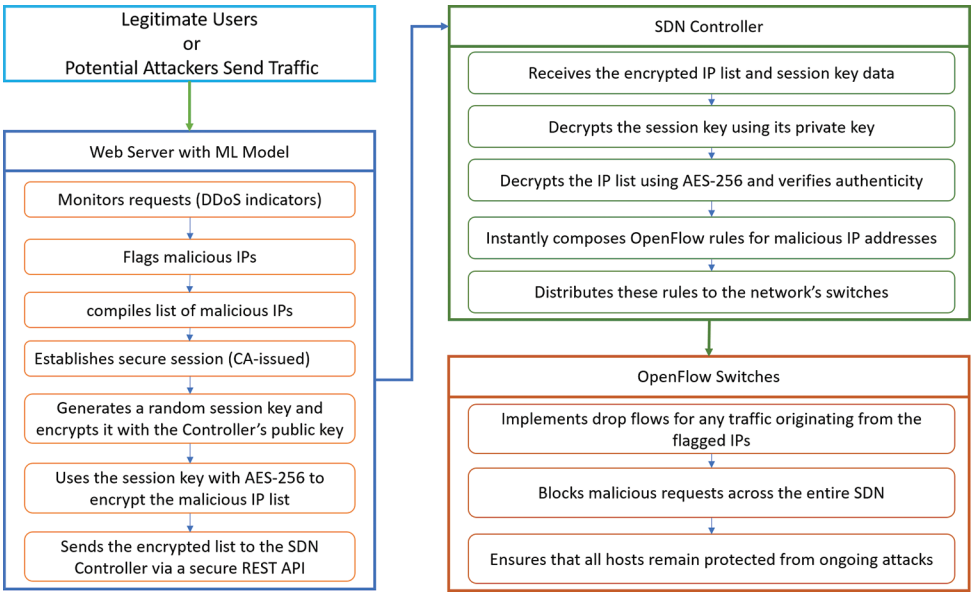


Fig. 4. Server-centric real-time distributed denial of service detection with CA-based security and immediate drop rules installed by the Software Defined Networking controller in this design, real-time detection at the server.

Table 1: Classification performance metrics across different models

Model	Accuracy	Precision	Recall	F1-Score	AUC	Cohen's Kappa	MCC	FPR	FNR	Cross-validated accuracy
LightGBM	0.9998	0.9999	0.9996	0.9998	0.9999	0.9996	0.9996	0.0001	0.0004	0.9992
XGBoost	0.9989	0.9996	0.9983	0.9989	0.9996	0.9978	0.9978	0.0004	0.0017	0.9985
CatBoost	0.9993	0.9999	0.9987	0.9993	0.9999	0.9987	0.9987	0.0001	0.0013	0.9986
Gradient boosting decision trees	0.9988	0.9991	0.9987	0.9989	0.9992	0.9978	0.9978	0.0009	0.0013	0.9993
AdaBoost	0.9987	0.9991	0.9983	0.9987	0.9992	0.9974	0.9974	0.0009	0.0017	0.9989
Bagging (bootstrap aggregating)	0.9991	0.9996	0.9987	0.9991	0.9999	0.9982	0.9982	0.0004	0.0013	0.9992
One-class svm	0.9823	0.9794	0.9998	0.9895	0.9998	0.9322	0.9343	0.1077	0.0002	0.9505
Bayesian Network	0.9982	1.0000	0.9965	0.9983	0.9983	0.9965	0.9965	0.0000	0.0035	0.9984

MCC: Matthews correlation coefficient, FPR: False-positive rate, FNR: False-negative rate, AUC: Area under curve

performance is robust and not attributable to overfitting on a specific train-test split. In contrast, cross-validation accuracy for One-Class SVM drops to 0.9505, revealing higher performance variability and emphasizing the importance of meticulous threshold tuning in anomaly-based detection systems.

6. DISCUSSION

The results illustrate clear differences in what each ML method is capable of doing for DDoS detection on SDN networks. With a dataset containing both traffic volume and behavior attributes, we can see which models are more accurate and more capable of preventing FP. The following section discusses these findings in more detail, starting with ensemble method advantages.

6.1. Ensemble Methods and Their Dominance

The results clearly demonstrate the superior performance of ensemble-based approaches, particularly LightGBM and CatBoost. Their near-perfect scores highlight two key strengths:

1. Robust feature handling: The 15 features, encompassing volumetric, connection, and distribution-based metrics, are well-suited to tree-based algorithms that inherently manage mixed data distributions.
2. Resilience to overfitting: Gradient boosting models iteratively construct weak learners (decision trees), enabling them to effectively identify subtle patterns while mitigating overfitting to noise.

The iterative refinement of gradient boosting, combined with the feature-rich dataset used in this study, likely contributes to the observed 0.9999 AUC scores in many cases. These high AUC values suggest that the malicious and benign traffic classes are distinctly separated within this dataset.

In computational efficiency, ensemble models such as LightGBM and CatBoost had fast convergence during training and low-latency predictions, which is suitable for real-time SDN deployments where response time is critical.

6.2. Trade-Offs with Anomaly Detection (One-Class SVM)

While One-Class SVM worked very well at identifying almost all attacks (recall = 0.9998), the sharp increase in FP (FPR = 0.1077) reflects a serious deficiency of anomaly-based detectors. It suggests that, when legitimate traffic significantly deviates from the baseline model, the model will end up marking benign requests as malicious. In operational scenarios, high FPR can lead to resource overhead in investigating benign events.

6.3. Bayesian Networks and Interpretability

Interestingly, the Bayesian Network achieved perfect precision (1.0000) at the expense of having a fairly elevated FNR (0.0035). That is, while it never misclassified benign traffic as an attack, it occasionally missed real attacks. Bayesian Networks do offer interpretability in the sense that they are graphical models whose conditional dependencies among features can make it easy to understand. For advanced research that aims to discover the underlying causal relations for attack patterns, this feature could provide new analytical contributions beyond bare classification.

Existing research has also explored the use of gradient boosting models such as LightGBM and CatBoost for DDoS detection. The majority of research, however, evaluates these models on typical data such as CICDDoS2019 or UNSW-15 and does not consider SDN-special traffic or real-time deployment. In contrast to our method, which is aimed at real SDN deployments and integrates these models with

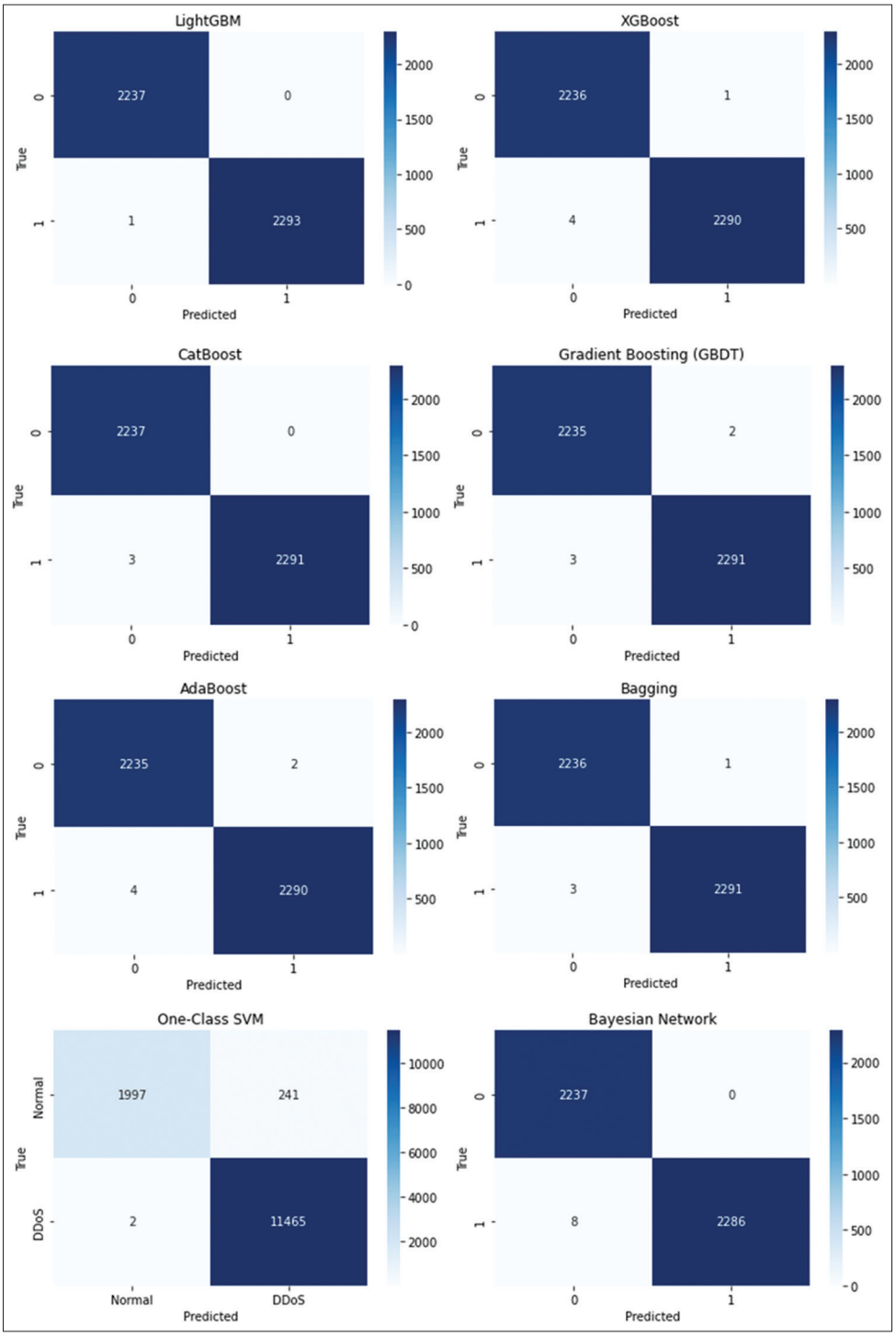


Fig. 5. Confusion matrices for each classifier showing the distribution of true positives, false positives, true negatives, and false.

an ongoing mitigation process, existing literature does not. Table 2 indicates a comparison between this proposal and existing works.

6.4. Implications for Advanced SDN-Based Research

With the network-level perspective of an SDN controller, high-performance ensemble models can be combined to offer

Table 2: Accuracy comparison of LightGBM and XGBoost across prior DDoS detection studies

Study	Model used	Dataset	Accuracy (%)
Han *et al*. [23]	LightGBM	CICDDoS2019	98.39
	XGBoost		97.95
Vaishali and Naik [24]	LightGBM	UNSW-15	99.18
Rozam and Riasetiawan [25]	XGBoost	A time series database	99.84
This study	LightGBM	Dataset used in this study	99.98
	XGBoost		99.89

DDoS: Distributed denial of service

a scalable and precise method for handling evolving DDoS threats. The distributed detection model with each server running the final classifier locally and securely notifying the controller of malicious IPs ensures low latency and effective utilization of resources. For mission-critical or large-scale deployments (e.g., data centers, cloud infrastructures), an extremely low FPR (essentially 0.0000) is required to avoid disrupting rightful traffic. Such minimal FPR (very nearly 0.0001) achieved by LightGBM and CatBoost shows that both models are well positioned for highly available scenarios fulfilling strict service-level agreements.

6.5. Limitations and Future Work

Although the ensemble methods have produced excellent results for this dataset, some of the potential avenues remain available for future work:

4. Temporal and incremental learning: Real networks have continuous traffic fluctuations. Online or incremental learning methods could be further explored to improve adaptability over time.
5. Multi-controller SDN: Even though in this research a single controller (Ryu) has been utilized, modern SDN deployments may have multiple controllers for fault tolerance and load balance. Classifier consistency in Mult controllers is an appealing area of future research.
6. Encrypted traffic analysis: Many application-layer protocols employ TLS/SSL encryption. Future work could be focused on feature extraction from encrypted flows, e.g., analysis of metadata or packet timing without analyzing payloads.

7. CONCLUSION

This study presents an ensemble-based DDoS detection framework tailored for SDN environments. It introduces a custom dataset collected in two phases, covering both normal and attack traffic and extracts 15 statistical features relevant to SDN traffic behavior. Eight ML models were evaluated, demonstrating high accuracy and suitability for real-time SDN deployment. Ensemble-based models, particularly LightGBM and CatBoost, significantly surpassed the performance of the other algorithms, achieving exceptional detection metrics (accuracy 0.9998, precision = 0.9999, and 0.0001 false alarms). These results underscore the suitability of advanced, tree-based methods for effectively processing the high-dimensional and heterogeneous data characteristic of network traffic logs. From an architectural perspective, distributing the ML detection process across web servers minimizes the processing load on the SDN controller and facilitates rapid attack detection. The use of CA-based secure delivery of IP lists ensures that only authenticated entities (controller and servers) are involved in the mitigation process. Upon receiving these IP lists, the controller implements network-wide blocking in real time, providing robust protection for the entire SDN environment against large-scale DDoS attacks.

REFERENCES

[1] M. Al-Fayoumi and Q. Abu Al-Haija. "Capturing low-rate DDoS attack based on MQTT protocol in software defined-IoT environment". *Array*, vol. 19, p. 100316, 2023.

[2] M. B. Anley, A. Genovese, D. Agostinello and V. Piuri. "Robust DDoS attack detection with adaptive transfer learning". *Computers and Security*, vol. 144, p. 103962, 2024.

[3] X. Etxezarreta, I. Garitano, M. Iturbe and U. Zurutuza. "Software-defined networking approaches for intrusion response in industrial control systems: A survey". *International Journal of Critical Infrastructure Protection*, vol. 42, p. 100615, 2023.

[4] N. N. Josbert, M. Wei, P. Wang and A. Rafiq. "A look into smart factory for industrial IoT driven by SDN technology: A comprehensive survey of taxonomy, architectures, issues and future research orientations". *Journal of King Saud University - Computer and Information Sciences,* vol. 36, no. 5, p. 102069, 2024.

[5] K. K. Karmakar, V. Varadharajan, M. Hitchens, U. Tupakula and P. Sariputra. "A trust-aware openflow switching framework for software defined networks (SDN)". *Computer Networks*, vol. 237, p. 110109, 2023.

[6] L. Mhamdi and M. M. Isa. "Securing SDN: Hybrid autoencoder-random forest for intrusion detection and attack mitigation". *Journal of Network and Computer Applications*, vol. 225, p. 103868, 2024.

[7] A. T. Phu, B. Li, F. Ullah, T. Ul Huque, R. Naha, M. A. Babar and H. Nguyen. "Defending SDN against packet injection attacks using deep learning". *Computer Networks*, vol. 234, p. 109935, 2023.

[8] X. Qin, F. Jiang, X. Qin, L. Ge, M. Lu and R. Doss. "Cgan-based cyber deception framework against reconnaissance attacks in ICS". *Computer Networks*, vol. 251, p. 110655, 2024.

[9] Z. Chen, M. Simsek, B. Kantarci, M. Bagheri and P. Djukic. "Machine learning-enabled hybrid intrusion detection system with host data transformation and an advanced two-stage classifier". *Computer Networks*, vol. 250, p. 110576, 2024.

[10] U. B. Clinton, N. Hoque and K. Robindro Singh. "Classification

of DDoS attack traffic on SDN network environment using deep learning". *Cybersecurity*, vol. 7, no. 1, p. 23, 2024.

[11] G. Srinivasa Rao, P. Santosh Kumar Patra, V. A. Narayana, A. Raji Reddy, G. N. V. Vibhav Reddy and D. Eshwar. "DDoSnet: Detection and prediction of DDoS attacks from realistic multidimensional dataset in IoT network environment". *Egyptian Informatics Journal*, vol. 27, p. 100526, 2024.

[12] J. Bhayo, S. A. Shah, S. Hameed, A. Ahmed, J. Nasir and D. Draheim. "Towards a machine learning-based framework for DDoS attack detection in software-defined IoT (SD-IoT) networks". *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106432, 2023.

[13] U. H. Garba, A. N. Toosi, M. F. Pasha and S. Khan. "SDN-based detection and mitigation of DDoS attacks on smart homes". *Computer Communications*, vol. 221, pp. 29-41, 2024.

[14] R. Swami, M. Dave and V. Ranga. "IQR-based approach for DDoS detection and mitigation in SDN". *Defence Technology*, vol. 25, pp. 76-87, 2023.

[15] A. Hirsi, L. Audah, A. Salh, M. A. Alhartomi and S. Ahmed. "Detecting DDoS threats using supervised machine learning for traffic classification in software defined networking". *IEEE Access*, vol. 12, pp. 166675-166702, 2024.

[16] A. A. Alashhab, M. S. Zahid, B. Isyaku, A. A. Elnour, W. Nagmeldin, A. Abdelmaboud, T. A. A. Abdullah and U. D. Maiwada. "Enhancing DDoS attack detection and mitigation in SDN using an ensemble online machine learning model". *IEEE Access*, vol. 12, pp. 51630-51649, 2024.

[17] H. M. Belachew, M. Y. Beyene, A. B. Desta, B. T. Alemu, S. S. Musa and A. J. Muhammed. "Design a robust DDoS attack detection and mitigation scheme in SDN-edge-IoT by leveraging machine learning". *IEEE Access*, vol. 13, pp. 10194-10214, 2025.

[18] Y. S. N. Fotse, V. K. Tchendji and M. Velempini. "Federated learning based DDoS attacks detection in large scale software-defined network". *IEEE Transactions on Computers*, vol. 74, no. 1, pp. 101-115, 2025.

[19] A. V. Songa and G. R. Karri. "An integrated SDN framework for early detection of DDoS attacks in cloud computing". *Journal of Cloud Computing*, vol. 13, p. 64, 2024.

[20] R. Abu Bakar, L. De Marinis, F. Cugini and F. Paolucci. "FTG-net-E: A hierarchical ensemble graph neural network for DDoS attack detection". *Computer Networks*, vol. 250, p. 110508, 2024.

[21] A. Ghorbanali and M. K. Sohrabi. "A comprehensive survey on deep learning-based approaches for multimodal sentiment analysis". *Artificial Intelligence Review*, vol. 56, no. Suppl 1, pp. 1479-1512, 2023.

[22] F. Nawshin, R. Gad, D. Unal, A. K. Al-Ali and P. N. Suganthan. "Malware detection for mobile computing using secure and privacy-preserving machine learning approaches: A comprehensive survey". *Computers and Electrical Engineering,* vol. 117, p. 109233, 2024.

[23] D. Han, H. Li and X. Fu. "Reflective distributed denial of service detection: A Novel Model utilizing binary particle swarm optimization—Simulated annealing for feature selection and gray wolf optimization-optimized LightGBM algorithm". *Sensors*, vol. 24, no. 19, p. 6179, 2024.

[24] R. Vaishali and S. M. Naik. "A Novel LightGBM-Bayesian Approach for DDoS Detection in SDN Environments". In: *2024 Moratuwa Engineering Research Conference (MERCon)*. pp. 7-12. 2024.

[25] N. Rozam and M. Riasetiawan. "XGBoost Classifier for DDoS attack detection in software defined network using sFlow protocol". *International Journal on Advanced Science, Engineering and Information Technology,* vol. 13, pp. 718-725, 2023.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Adapting F5-TTS Model to Kurdish Sorani: Diffusion-based Speech Synthesis with a Specialized Dataset

Hamreen Ahmad, Aree Mohammed

Department of Computer Science, University of Sulaimani, Sulaymaniyah, Iraq

ABSTRACT

The Kurdish language is one of the low-resource languages in the field of speech synthesis. Most of the currently available Kurdish text-to-speech (TTS) systems lack both accuracy and naturalness. To address this issue, this study fine-tunes the F5-TTS model for the Kurdish (Sorani) language, an advanced model that had not previously been fine-tuned for this language. The process began with the creation of a high-quality, well-constructed, single-speaker dataset containing more than 10 h of recorded speech that was collected from news, interviews, and short videos. The dataset was very carefully curated to ensure balanced sample durations, accurate transcriptions, emotional diversity, and a wide range of topics and speaking styles. After ensuring that the training data was sufficient, clean, and prepared, the F5-TTS model was fine-tuned on this data. Both objective and subjective evaluations were conducted to verify the model's performance. For the objective evaluation, the model achieved a character error rate of 4.3% and a word error rate of 20.37%, indicating a high transcription accuracy in the generated audio. In the subjective evaluation, the mean opinion score reached 4.72, showing that the synthesized speech is very close to the original speaker's voice. These results demonstrate that diffusion-based models like F5-TTS can be effectively adapted to low-resource languages when supported by a well-designed dataset.

Index Terms: Speech Synthesis, Diffusion-Based Model, Low-Resource Language, F5-Text-to-speech, Kurdish Sorani

1. INTRODUCTION

Text-to-speech (TTS), also known as speech synthesis, is a technology that converts written words into spoken language. It plays an important role in modern technology. Most artificial intelligence chatbots, smart devices, and even organizations use such systems to provide fast, more realistic, and enjoyable service. For example, by converting written text into natural-sounding speech, this technology is mainly utilized in human-computer interaction systems to make them more user-friendly [1]–[3]. An advanced TTS system should not only read what has been written, but also be able to maintain emotion, rhythm, punctuation, and many other properties to produce speech that sounds like a real human.

However, achieving these capabilities is not equally accessible across all languages. For widely spoken languages such as English, Spanish, and Arabic, TTS technology has been improved significantly [4], while low-resource languages are neglected and outdated [5]. In the case of Kurdish, while recent studies [6]–[8] have made noticeable progress, they often show limitations in fluency and emotional expression, as indicated by their reported mean opinion score (MOS) of 4.1 [8] and 3.9 [7]. The two major reasons for this gap are: unavailability of a high-quality, domain-diverse dataset, and

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp198-207

E-ISSN: 2521-4217
P-ISSN: 2521-4209

Copyright © 2025 Ahmad and Mohammed. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Hamreen Ahmad, Department of Computer Science, University of Sulaimani, Sulaymaniyah, Iraq. Email: hamreen.ahmad@univsul.edu.iq

Received: 29-07-2025 Accepted: 21-09-2025 Published: 06-10-2025

the use of outdated TTS models that are not suitable for low-resource languages [5].

The Kurdish Sorani script is similar to that of Arabic, Persian, and other related languages. These languages share some phonetic consistency [9], [10]. Most words in Kurdish Sorani script are pronounced as they are written. However, this phonetic transparency does not eliminate challenges faced when building a TTS dataset. For example, during natural speech or when expressing emotions, speakers often skip or merge letters, alter pronunciation, or blend sounds [11]. In addition, the Kurdish language also lacks a standardized punctuation convention [11], [12], which makes it more difficult to map Kurdish script to speech. These issues introduce more complexity to the process of creating a clean and usable dataset for TTS models.

Given these inherent challenges in Kurdish TTS development, this work presents the first fine-tuned F5-TTS model for the Kurdish Sorani language. F5-TTS, which stands for A Fairytaler that Fakes Fluent and Faithful Speech with Flow Matching, is a recently developed open-source model that leverages flow-matching and diffusion-based techniques to generate natural and emotional speech without the need for extensive training data [13]. The process of fine-tuning this model began with the collection of 10.11 h of high-quality speech from professional TV recordings, which were then segmented into over 4,800 carefully curated samples covering diverse topics and speech styles. Each sample was manually reviewed for transcription accuracy and balanced duration.

The following are key contributions of this study:

- Introduces the first fine-tuned F5-TTS model for Kurdish Sorani, a low-resource language.
- Constructs a high-quality single-speaker Kurdish Sorani dataset with balanced segment durations, emotional variability, and transcription precision.
- Provides the first Kurdish TTS evaluation that combines both objective metrics (Character error rate [CER] and word error rate [WER]) and subjective MOS ratings, offering a more comprehensive assessment than prior studies.

The rest of this paper begins with a review of related studies on speech synthesis for low-resource languages and prior efforts in Kurdish TTS. Next, the dataset design is discussed, including how the speech data was collected, segmented, and transcribed. This is followed by an explanation of the F5-TTS model architecture and the fine-tuning process. Finally, the evaluation results are presented, and the implications of the findings are discussed, with concluding remarks and directions for future research.

2. RELATED WORK

TTS technology has seen fast advancements in recent years. However, for low-resource languages such as Kurdish, progress has been slower due to limited attention and support in the field of speech synthesis. This section provides an overview of traditional and neural speech synthesis methods, with a focus on approaches adapted to low-resource settings, and mentions previous efforts related to Kurdish TTS systems.

2.1. Traditional and Neural Speech Synthesis Methods

In the past, concatenative speech synthesis methods were the most widely used. They could achieve a noticeable amount of naturalness [14]. Concatenative data-driven approaches even made their way into speech recognition and some musical synthesis applications. Although some statistical techniques like the hidden Markov model provided some advancements in the field [15], [16], they were still lacking naturalness and accuracy [17].

Recent advancements in neural network-based models have led to improvements in TTS models [18]. End-to-end architectures such as Tacotron and Tacotron 2 [19] were able to produce more realistic and natural speech, but they require large amounts of high-quality training data. This leads to more difficulty when using it with low-resource languages like Kurdish.

2.2. TTS for Low-Resource Languages

Due to limited data availability, fine-tuning TTS models for low-resource languages presents several difficulties [20]. A common approach to address these challenges is transfer learning [21]. Transfer learning involves utilizing a pre-trained model and extending it to a new language. For example, a recent study [22] compared multiple cross-lingual transfer techniques for some low-resource languages, including Bulgarian, Georgian, Kazakh, Swahili, Urdu, and Uzbek. The source languages used in their research were English, Finnish, Hindi, Japanese, and Russian. This variability helped identify the source language that produced the best quality results. Achieving CERs from 6.70% to 61.92% and predicted MOS scores between 2.24 and 3.02, the study demonstrated that phonological features offer better generalization and are more effective than conventional phone mapping in cross-lingual TTS. However, their approach relied on around 10 h of source-language data and 10 min of target data.

Another technique that supports low-resource languages is multilingual training [23]–[25]. Training a model on multiple languages simultaneously can improve performance on low-resource languages by sharing linguistic properties among them.

Un addition, zero-shot learning techniques, where a model performs a task without having been explicitly trained on that task or language, can also help models to generalize to unseen languages [26], [27] without requiring a large amount of training data.

2.3. Related Kurdish TTS Work

At present, only a limited number of neural network-based TTS models exist for the Kurdish language. Some organizations have privately developed datasets and models for the language to meet internal needs. For example, Rudaw TV has developed a TTS model for reading news on their website. On the other hand, some recent studies have made noticeable contributions in the field:

- Muhamad and Veisi [6]: Utilizing transfer learning within an end-to-end architecture, they used a pretrained HiFi-GAN vocoder, a model that converts intermediate acoustic features into audible waveforms, together with Tacotron 2 on LJ-Speech dataset [28] to develop a Kurdish TTS system. They used English character embedding from a pretrained English model while feeding Kurdish characters as input. The dataset used in this study consisted of approximately 10 h of Kurdish speech recordings and corresponding transcriptions. Results showed a MOS of 4.1, which indicates a high level of naturalness of the developed TTS system compared to TTS systems in high-resource languages.
- Ahmad and Rashid [7]: They introduced an end-to-end TTS system specifically for Kurdish Sorani dialect. Their approach uses a pretrained variational autoencoder, a generative model that learns compact latent representations of data and reconstructs them back into realistic outputs, for audio waveform reconstruction, combined with adversarial training, a technique where a generator and discriminator compete to improve realism, to enhance the quality of the synthesized speech. In addition, they incorporated a stochastic duration predictor to enhance the naturalness of the output audio. The method facilitates real-time generation of Kurdish speech audio with variations in pitch and rhythm. Evaluation on a custom dataset showed a MOS of 3.94, showing better performance compared to one-stage and two-stage models using subjective human evaluation.
- Abdullah *et al.* [8]: Instead of using pretrained English models, this study improves an existing Kurdish TTS architecture based on Tacotron by training a WaveGlow vocoder from scratch. The training data consist of a 21-h native Kurdish speech dataset, particularly for Central Kurdish (Sorani) dialect. In addition to optimizing WaveGlow architecture, the study also introduces advanced prosody modeling techniques to improve the rhythm, stress, and intonation of synthesized speech. The adapted model achieved a MOS of 4.91, which sets a new benchmark for Kurdish speech synthesis.

While these studies have shown significant progress in the field of speech synthesis, they either depend on English-based architectures or need a large amount of training data. In contrast, our study focuses on fine-tuning F5-TTS, a more modern, up-to-date, and efficient architecture for the Kurdish Sorani language using a carefully prepared dataset.

3. DATASET DESIGN

Building a speech dataset from scratch is a gradual and careful process in which every single detail matters [20], [29]. You must care about every single detail. In the case of TTS, every single word matters because the quality and the diversity of the training data will undoubtedly impact the performance of the resulting model.

3.1. Dataset Collection

As the goal of this work was to fine-tune the F5-TTS model with a single-speaker dataset, a consistent and high-quality audio source was needed. After multiple options were evaluated, the recordings of Shaho Amin, a well-known newsreader from Rudaw TV, were selected for the following reasons:

- Public Availability: Rudaw TV's archives are publicly available and can be easily collected from online sources.
- Clear Pronunciation: The speaker's voice is understandable, clear, and well expressed.
- Loudness: Audio levels are consistent, and additional normalization is not required.
- Recording quality: As a major media channel, Rudaw TV uses advanced recording equipment, resulting in clean and high-fidelity audio.
- Familiarity: The speaker's voice is widely recognized, which allows listeners to better judge the similarity between synthesized and real speech.

Although the final dataset contains 10.11 h of audio, more than 15 h of video were initially collected. During preprocessing,

low-quality and noisy sections were removed. The two main sources for the recordings were Rudaw TV's website and YouTube Channel. For further processing, all collected content was then converted into WAV format using a lossless conversion process to preserve audio quality.

The collected recordings covered a variety of topics, including news broadcasts, interviews, shorts, and other random videos. This diversity was intended to increase the model's generalization by providing emotions, intonation patterns, and different speech rates in the dataset. After it was observed that the speaker's vocal tone was evolving over time, recordings were selected from different time periods (2018–2024). This helped enhance the dataset by adding more variability to the data.

3.2. Data Segmentation

After data collection, the dataset passed several preparation stages, with segmentation and transcription being the most time-consuming steps. In data segmentation, audio files were divided into manageable clips while ensuring the following:

- Proper logical sentence boundaries
- Natural starting and ending tone
- A balanced distribution of clip durations.

To achieve a balanced dataset, segments were distributed approximately equally across different duration ranges (e.g., 2–4 s, 4–6 s, 6–8 s, etc.), as shown in Fig. 1. This prevented the model from overfitting to specific speech length ranges. During this process, Ocenaudio software was utilized, as it allows the manual selection and export of clean segments.

3.3. Transcription Process

After all segments were cleaned, finalized, and prepared, the transcription process was initiated. Initially, segments were transcribed quickly to match what was heard in the recordings. In later stages, each segment was carefully reviewed multiple times word by word to eliminate the remaining spelling mistakes.

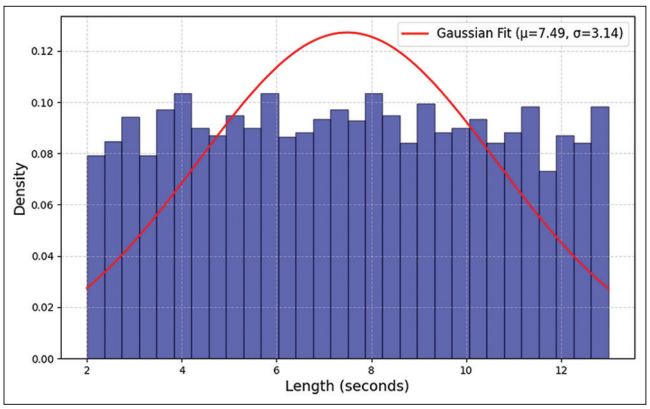


Fig. 1. Distribution of dataset samples.

Since punctuation plays an important role in expressing the tone and rhythm of synthesized speech, it was added where appropriate. Adding punctuation in the transcription helped the model to produce more realistic and expressive audio. Moreover, it allowed better control over speaking style when synthesizing new text. Most common punctuation use cases were:

- Adding a full stop (.) at the end of complete sentences to indicate closure
- Using a comma (,) to include short pauses or transitions between phrases
- Using a colon (:) when introducing reported speech
- Using question mark (?) for interrogative sentences
- Using an exclamation mark (!) for astonishing expressions.

While punctuation was added for expressiveness, each word itself was transcribed exactly as it was recited in the recordings. As Kurdish language is a non-phonemic writing system like Arabic and Persian, several challenges unique to the language were encountered despite efforts to avoid transcription mistakes:

- Some letters are written in a way but pronounced differently. For example, the letter (س) is pronounced as (ص) in words like شەست, موسولمان, سەت; similarly, (ت) may be pronounced as (ط) in تازە, سەتا, پێنجسەت.
- In fast speech, some letters are dropped or merged. For example, in چاوت تێژبێ, only one (ت) is clearly pronounced.
- Sometimes, when a letter comes twice repeatedly, it is hard to distinguish whether it is recited as one letter or two. For example, words like چاککردنەوە, رێککەوتنی contain consecutive (ک) sounds, but it is often unclear whether one or both are audibly pronounced.
- The English/ŋ/sound appears in some Kurdish words such as سنگ, مانگ, while there is not a dedicated letter for the sound in script.

3.4. Dataset Statistics

The final version of the dataset was 10.11 h long in 4,856 samples. Each sample was stored in WAV format. The dataset vocabulary had 2,567 unique tokens. Detailed statistics are provided in Table 1.

Fig. 1 demonstrates how samples are distributed evenly based on duration, which helps the model generalize across different speech lengths.

4. F5-TTS MODEL FINE-TUNING

Fine-tuning TTS models for low-resource languages like Kurdish Sorani can be challenging, primarily due to the lack of training data and limited computational resources. This section describes the steps taken to fine-tune the F5-TTS model for Kurdish Sorani using a specialized speech dataset. It includes an overview of the F5-TTS architecture, the data preparation and training environment, and a description of the fine-tuning process. An overview of the full training pipeline is presented in Fig. 2.

4.1. Overview of F5-TTS Architecture

Recent advancements in TTS technologies have improved the naturalness and flexibility of synthesized speech. Previous TTS systems were primarily based on auto-regressive (AR) models, which generate speech one token at a time, with each step depending on the last. Although models such as variational inference with adversarial learning for end-to-end text-to-speech and Tacotron2 [19] achieved high-quality results, they still suffer from inference latency and errors during sequential generation.

TABLE 1: Training dataset statistics

Metric	Value
Total duration	10.11 h
Number of samples	4,856
Shortest clip	2.0 s
Longest clip	13.0 s
Audio format	WAV
Vocabulary (Unique words)	2,567

F5-TTS is a fully non-autoregressive (NAR) TTS system, meaning it generates all audio tokens in parallel rather than sequentially, that produces high-quality speech using a flow matching framework, a generative modeling technique that learns to map noise into realistic data distributions [30], combined with a diffusion transformer (DiT) backbone, a transformer-based architecture tailored for diffusion models [31]. NAR approaches, particularly those based on diffusion-based models, which iteratively denoise random noise to produce realistic data [32], and flow matching techniques, allow for faster and more robust synthesis.

Unlike traditional TTS systems that rely on phoneme alignment, explicit duration models, or complex text encoders, F5-TTS presents a simplified architecture. It can directly learn how to speak from text characters and noisy speech. The model architecture can be described in five main phases, as shown below:

- Input Processing: The input text is turned into a sequence of characters, then padded with filler tokens to match the length of target speech frames. This enables the model to learn implicit alignment during training, which leads to a more natural rhythm.
- Feature Refinement: Input text features are refined using ConvNeXt blocks, a modern convolutional neural network design that improves feature extraction efficiency and alignment with speech frames.
- Denoising Process: A DiT gradually denoises a sampled noisy speech signal, conditioned on the refined text representation. It turns the noise into clear speech based on input text.

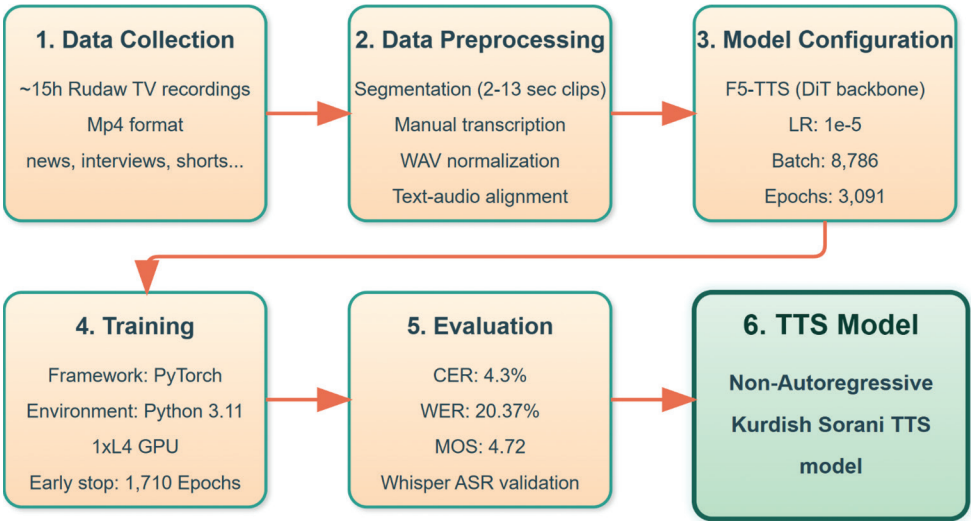


Fig. 2. Proposed model workflow.

- Training Objective: Training optimizes a flow-matching loss combined with a text-guided speech infilling task.
- Inference Strategy: Since diffusion models are typically slow due to step-by-step denoising, F5-TTS employs a sway sampling technique, an accelerated sampling method that reduces generation time while maintaining naturalness.

An overview of F5-TTS training (left) and inference (right) flow is illustrated in Fig. 3.

4.2. Data Pipeline and Training Environment

The dataset used for fine-tuning comprised 4,856 audio-text samples, totaling 10.11 h of Kurdish Sorani audio. Since the audio samples varied in duration, dynamic batching, a technique that groups sequences of similar lengths together to optimize training efficiency [33], was employed based on the number of spectrogram frames, with each batch being configured to contain approximately 8,786 frames to maximize GPU utilization [34]. As F5-TTS is a resource-intensive model, the fine-tuning process was conducted on Google Colab in a well-optimized environment. Hardware specifications and configuration parameters of the training process are provided in Tables 2 and 3.

4.3. Fine-tuning Process

After the environment was set up on Google Colab and 2 TB of Google Drive storage was allocated for saving checkpoints, the fine-tuning process began on April 18th, 2025, and continued for approximately 1 week without major difficulties.

The model started to produce reasonable speech within the first few thousand steps, and the quality gradually improved over time. While the fine-tuning was originally set to 3,091 epochs, it was terminated after 1,710 epochs (around 700,000 steps) because live output evaluations conducted every 5,000 steps revealed a close resemblance between the synthesized speech samples and the original recordings, indicating that further training was unnecessary.

As shown in Fig. 4, early in the training, the loss value hardly went under 0.4. However, by the final steps, the loss had decreased to 0.24272, which demonstrates a real

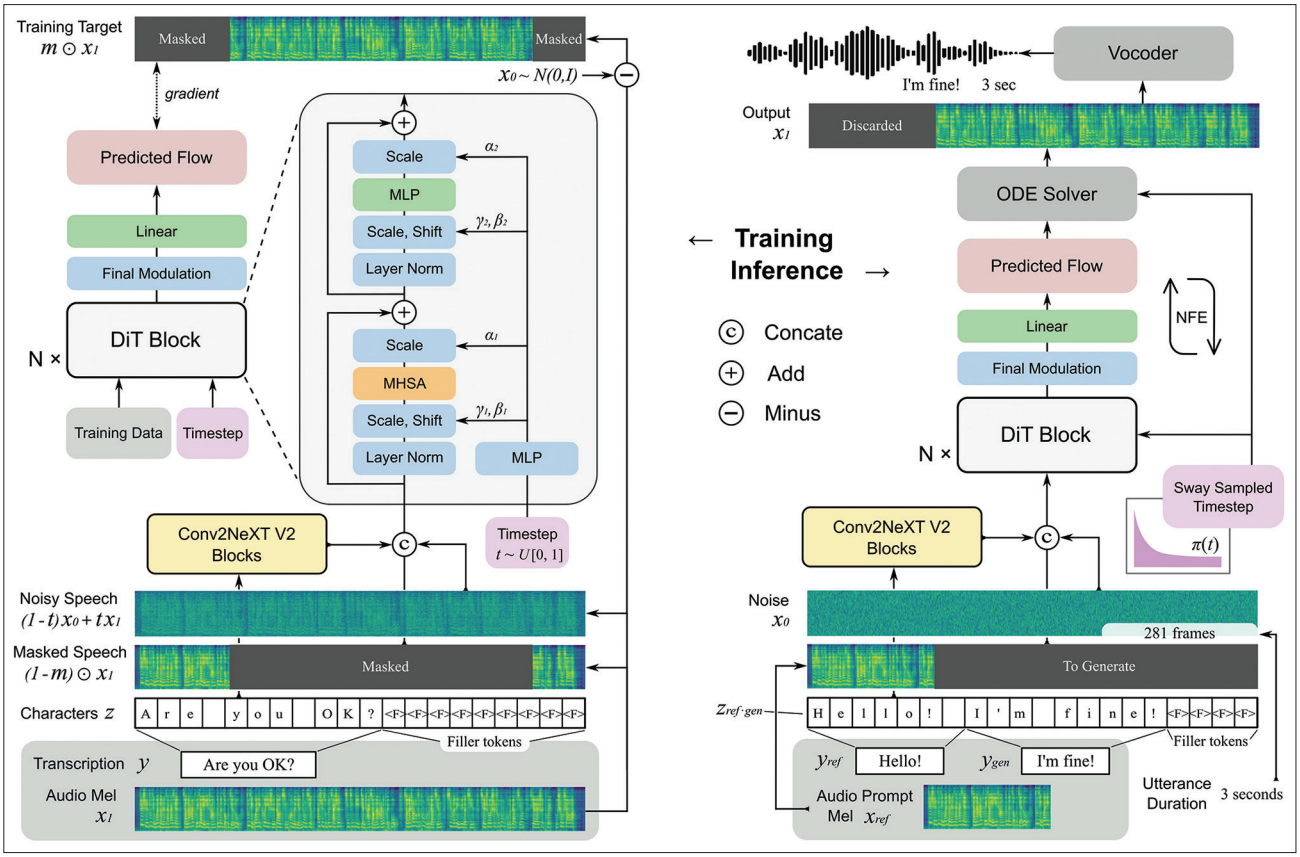


Fig. 3. F5-text-to-speech model architecture [13].

improvement in the model's convergence. This trend is also shown in Table 4, which presents the minimum loss values recorded across different fine-tuning intervals. In parallel, the learning rate schedule, demonstrated in Fig. 5, shows a gradual reduction in learning rate, contributing to a more stable training process. During the training, important metrics such as the loss curve, learning rate schedule, and other training statistics were monitored and recorded using the Weights and Biases (wandb) platform [35]. The decision to end the training early in the process was influenced by these tracking tools.

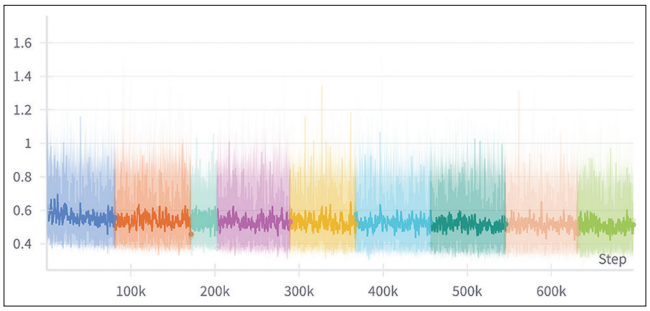


Fig. 4. Training loss value over steps.

TABLE 2: Training environment specifications

Component	Specification
GPU	NVIDIA L4 (22.5 GB VRAM)
CPU	6 physical cores, 12 logical threads
Operating system	Linux (Kernel 6.1.123+)
Python version	3.11.12
Training framework	PyTorch (with Weights and Biases (wandb) logging)

TABLE 3: Training configuration parameters

Parameter	Value
Batch size per GPU	8,786 frames
Batch size type	Frame-based
GPUs used	1
Gradient accumulation steps	1
Maximum samples	64
Learning rate	1e–5
Maximum gradient norm	1
Number of warmup updates	242
Number of epochs	3,091
Logging	Weights and Biases (wandb)
Checkpoint saving	Every 5,000 updates; save last at 2,000 updates

5. RESULTS DISCUSSION AND EVALUATION

The loss curve shown in Fig. 4 proves that the model has achieved excellent convergence and effective learning throughout the training process. To further assess the performance of the fine-tuned model, three evaluation metrics were used: CER, WER, and MOS. CER and WER are objective metrics that measure transcription accuracy at the character and word level, while MOS is a 1–5 scale for subjective rating of naturalness.

For objective metrics, the synthesized speech was transcribed using a pretrained Whisper Kurdish Sorani Automatic Speech Recognition (ASR) model ("PawanKrd/asr-large-ckb"), and the transcriptions were then compared to the original text. In contrast, the MOS metric was obtained based on human listener ratings, as the naturalness of speech may not be fully captured by objective metrics alone.

The results of CER and WER for three different samples are presented in Table 5. In both metrics, lower average scores indicate better synthesis quality. The average WER was 20.37%, and the average CER was 4.3%. These results demonstrate the model's strong transcription accuracy, particularly at the character level. The relatively high WER value of 20.37% may be due to limitations in the ASR model for Kurdish Sorani rather than actual quality issues, as demonstrated by the low CER and high MOS scores.

TABLE 4: Training loss value over time

From	To	Runtime	loss (Min)	lr
April 26, 2025 09:09	April 27, 2025, 02:59	17 h 50 m 16 s	0.24272	4.47E-06
April 25, 2025 02:54	April 26, 2025 02:46	23 h 52 m 27 s	0.25062	4.99E-06
April 23, 2025 21:46	April 24, 2025 21:34	23 h 47 m 59 s	0.25407	5.67E-06
April 22, 2025 20:36	April 23, 2025 20:30	23 h 54 m 15 s	0.27873	6.38E-06
April 21, 2025 19:49	April 22, 2025 19:14	23 h 25 m 6 s	0.28213	7.08E-06
April 20, 2025 17:58	April 21, 2025 17:49	23 h 51 m 46 s	0.28673	7.71E-06
April, 20 2025 07:47	April 20, 2025 16:52	9 h 4 m 31 s	0.30502	8.40E-06
April 19, 2025 06:42	April 20, 2025 06:08	23 h 26 m 19 s	0.31336	8.64E-06
April 18, 2025 06:50	April 19, 2025 04:49	21 h 58 m 36 s	0.3218	9.36E-06

TABLE 5: Word error rate (WER) and character error rate (CER) for synthesized Kurdish speech samples

Sample	Original text	Generated automatic speech recognition for the synthesized speech	WER (%)	CER (%)
1	هاوبیشتیهاتیەک بە پرۆداوی گوت زۆرم بڕی بوو، بەر لەبەر صەڤڵی شەبەت و وەجیبەی خورما دانیشتبووم هەل لەگەڵ ماموستا گوتی نەلاهو نەگبەر پەرم دایە ڕاست تا تێر نەبووم و لازم نەهێنا	ڕەسەرا هب ووب ێەبرب هزۆز ئەوگ ێرۆادوور هب کەێەناماۆێتشێەنزواە ائتسۆزام لەمگخل ر‌ەهووبەتشۆەنزاد امزوچ ێەهبچەو و ئەهبەشۆ تۆلەتسس ائێەهن هزاو هووبەهن رێێت ات ئسنارا هێاد هڕب زەو کەهو وانەەێ ێەتوگ	35.71	7.10
2	لە نیوێۆرک، ئەم پیاوە نەهو مۆزەی گڕی بە 6 ملیۆن و 200 هەزار دۆلار	رازەە 200 و ئۆۆعلەم 6 هب ێەڕک ێەزۆم وەێ موراێب هەێ کەرۆەوعەن هل رالۆد	14.29	3.08
3	و لەاس عەناکەراک ئەعرەتەز اوان هل کەێکەەی وکەهو ەەدروک ێەکەێزاک سەمک رازەە نادەص ننەۆبەەەێ ەەدروک هب ەاوۆزۆزۆک ێەکەۆرەێچ ەەهەنارەێگ	ئەعرەتەزاوان هل کەێکەەی وکەهو ەەدروک ێەکەێزاک سەمک رازەە نادەص ننەۆبەەەێ ەەدروک هب ەاوۆزۆزۆک ێەکەۆرەێچ ەەهەنارەێگ و لەاس عەناکەراک	11.11	2.73
Average			20.37	4.30

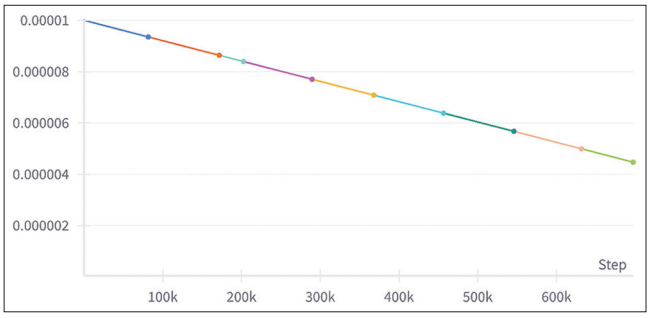


Fig. 5. Learning rate schedule during training.

To complement these objective measures with human perception assessments, a subjective evaluation was conducted using the MOS metric. Synthesized speech samples were shared online with more than 50 different native Kurdish Sorani speakers of different ages and genders. Each listener was asked to rate the naturalness of the same set of synthesized speech clips on a five-point scale ranging from 1 (Bad) to 5 (Excellent). The average score achieved in this metric was 4.72 out of 5, implying that the generated speech was perceived as highly natural. While this result is promising, it is important to remember that the number of listeners was limited, and no statistical significance testing was applied. Future evaluations with larger participant groups and formal statistical analysis would lead to more reliable results.

At the same time, these findings were obtained using clean single-speaker data in controlled settings. Several factors that may impact system performance in real-world scenarios, including background noise, overlapping speech, as well as speaker and dialect variability. Handling these challenges may require additional techniques and extended datasets, which can be explored in future research directions.

In comparison to older TTS approaches for low-resource languages such as transfer learning, the results of our diffusion-based model show significant improvements. Table 6 provides a comparison of various methods applied to low-resource languages, including different architectures and learning strategies. Notably, while prior Kurdish TTS studies have reported their MOS scores, none have evaluated their systems using objective evaluation metrics, such as CER or WER. This work is the first to apply these objective measures to Kurdish Sorani TTS for a more precise evaluation.

Beyond the technical achievements, this work has significant cultural and linguistic importance. Developing high-quality TTS systems for underrepresented Kurdish Sorani can

TABLE 6: Comparison of text-to-speech systems for low-resource languages

Study	Target language	Model architecture	Training data	Learning type	Character error rate (%)	Mean opinion score (1–5)
Do *et al*. [22]	Bulgarian (bg) Georgian (ka) Kazakh (kk)	FastSpeech 2+HiFi-GAN	Source: ~10 h/ speaker Target: ~10 min/speaker	Transfer learning	6.70 32.05 18.83	3.02 2.43 2.37
Abdullah *et al*. [8]	Kurdish Sorani (ku)	Tacotron 2+HiFi-GAN	10 h	Transfer learning	N/A	4.1
Ahmad and Rashid [7]	Kurdish Sorani (ku)	Variational autoencoder+Adversarial	Custom dataset	Stochastic duration predictor	N/A	3.9
Proposed model	Kurdish Sorani (ku)	F5-TTS (diffusion-based)	10.11 h	Diffusion-based	4.30	4.72

support digital inclusion, education, and even cultural preservation. It ensures the language is not left behind in the advancements of TTS technology. The fine-tuned model becomes a core module in many applications such as audiobooks, e-learning materials, and automated dubbing tools. In addition, building computational resources for Kurdish helps document and standardize the language, which faces dialectal diversity and limited digital resources. For this reason, the benefits of this research go beyond speech synthesis only; it can serve both practical applications and the preservation of Kurdish language identity.

6. CONCLUSION AND FUTURE WORKS

In this study, the first F5-TTS model was successfully fine-tuned for Kurdish Sorani, a low-resource language that has received less attention in the field of speech synthesis. The process began by creating a 10.11-h speech dataset. The work focused mostly on balancing segments, topic variety, emotions, transcription accuracy, punctuation, and other details that optimized the dataset quality. Using this high-quality dataset, the open-source F5-TTS model was fine-tuned in a well-optimized environment and configuration settings. Evaluations showed that the system learned to synthesize natural, emotional, and human-like Kurdish speech.

Compared to previous Kurdish TTS works, this study indicates that modern diffusion-based models like F5-TTS can improve the naturalness and reliability of speech synthesis for low-resource languages without requiring a massive amount of training data or using a pre-trained English-based architecture. The results show that a carefully constructed dataset can overcome the typical limitations of low-resource TTS systems.

While the fine-tuned model provides good results, there are still some areas for further improvement. One limitation is the use of a single-speaker dataset, which may reduce generalizability to different voices or dialects. Using a multi-speaker dataset that includes various emotions and accents would enable additional training to improve this fine-tuned model. This can improve the model's generalization for different speaking styles. In addition, real-world deployment introduces challenges such as background noise, overlapping speech, and speaker variability, which may reduce accuracy. Addressing these conditions in future work will improve the system's robustness and support its integration into practical Kurdish applications such as audiobooks, dubbing, and other voice-based systems.

REFERENCES

[1] L. Mohasi and D. Mashao. "Text-to-Speech Technology in Human-Computer Interaction". In: 5th *Conference on Human Computer Interaction in Southern Africa, South Africa (CHISA 2006, ACM SIGHI)*, pp. 79-84, 2006.

[2] V. R. Reddy and K. S. Rao. "Better Human Computer Interaction by Enhancing the Quality of Text-to-Speech Synthesis". In: *2012 4th International Conference on Intelligent Human Computer Interaction (IHCI)*, IEEE, United States, pp. 1-6, 2012.

[3] R. Zhen, W. Song, Q. He, J. Cao, L. Shi and J. Luo. "Human-computer interaction system: A survey of talking-head generation". *Electronics*, vol. 12, no. 1, p. 218, 2023.

[4] T. Xie, Y. Rong, P. Zhang, W. Wang and L. Liu. "*Towards Controllable Speech Synthesis in the Era of Large Language Models: A Systematic Survey*". [arXiv Preprint]; 2025.

[5] T. Reitmaier, E. Wallington, D. K. Raju, O. Klejch, J. Pearson, M. Jones, P. Bell, S and Robinson. "Opportunities and Challenges of Automatic Speech Recognition Systems for Low-Resource Language Speakers". In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1-17, 2022.

[6] S. Muhamad and H. Veisi. "End-to-end kurdish speech synthesis based on transfer learning". *Passer Journal of Basic and Applied Sciences*, vol. 4, no. 2, pp. 150-160, 2022.

[7] H. A. Ahmad and T. A. Rashid, "Central Kurdish text-to-speech synthesis with novel end-to-end transformer training". *Algorithms*, vol. 17, no. 7, p. 292, 2024.

[8] A. A. Abdullah, S. S. Muhamad and H. Veisi. "*Enhancing Kurdish*

Text-to-Speech with Native Corpus Training: A High-Quality Waveglow Vocoder Approach". [ArXiv Preprint]; 2024.

[9] M. K. Mahmood, A. Q. H. Rash, M. A. Q. H. Rash, S. K. Mahmood and H. Güler. "Kurdish and Persian: Dialects or separate languages?" *International Journal of Social Science and Human Research*, vol. 6, p. 2216, 2023.

[10] G. Tavadze. "Spreading of the Kurdish language dialects and writing systems used in the Middle East". *Bulletin of the Georgian National Academy of Sciences*, vol. 13, no. 1, pp. 170-174, 2019.

[11] K. S. Esmaili. "*Challenges in Kurdish Text Processing*". [arXiv Preprint]; 2012.

[12] E. M. Qadir and H. H. Padar. "Punctuation in English and Kurdish: A contrastive study". *Koya University Journal of Humanities and Social Sciences*, vol. 5, no. 1, pp. 41-61, 2022.

[13] Y. Chen, Z. Niu, Z. Ma, K. Deng, C. Wang, J. Zhao, K. Yu and X. Chen. "*F5-tts: A Fairytaler that Fakes Fluent and Faithful Speech with Flow Matching*". [Preprint]; 2024.

[14] A. J. Hunt and A. W. Black. "Unit selection in a concatenative speech synthesis system using a large speech database". In: *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*. IEEE, Atlanta, GA, USA, pp. 373-376, 1996.

[15] A. Falaschi, M. Giustiniani and M. Verola. "A hidden Markov model approach to speech synthesis". In: *First European Conference on Speech Communication and Technology, Presented at the EUROSPEECH*. pp. 2187-2190, 1989.

[16] K. Tokuda, Y. Nankaku, T. Toda, H. Zen, J. Yamagishi and K. Oura. "Speech synthesis based on hidden Markov models". *Proceedings of the IEEE*, vol. 101, no. 5, pp. 1234-1252, 2013.

[17] H. Zen, K. Tokuda and A. W. Black. "Statistical parametric speech synthesis". *Speech Communication*, vol. 51, no. 11, pp. 1039-1064, 2009.

[18] Y. Wang, R. J. Skerry-Ryan, D. Stanton, Y. Wu, R. J. Weiss, N. Jaitly, Z. Yang, Y. Xiao, Z. Chen, Q. Le, Y. Agiomyrgiannakis, R. Clark and R. A. Saurous. "*Tacotron: Towards End-to-End Speech Synthesis*". [ArXiv Preprint]; 2017.

[19] J. Shen, R. Pang, R. J. Weiss, M. Schuste, N. Jaitly, Z. Yang, Z. Chen, Y. Zhang, Y. Wang, R. J. Skerry-Ryan, R. A. Saurous, Y. Agiomyrgiannakis and Y. Wu. "Natural tts Synthesis by Conditioning Wavenet on Mel Spectrogram Predictions". In: *Presented at the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, United States, pp. 4779-4783, 2018.

[20] H. Ali, S. Subramani, R. Varahamurthy, N. Adupa, L. Bollinani and H. Malik. "*Collecting, Curating, and Annotating Good Quality Speech Deepfake Dataset for Famous Figures: Process and Challenges*". Cornell University, United States, 2025.

[21] K. Azizah and W. Jatmiko. "Transfer learning, style control, and speaker reconstruction loss for zero-shot multilingual multi-speaker text-to-speech on low-resource languages". *IEEE Access*, vol. 10, pp. 5895-5911, 2022.

[22] P. Do, M. Coler, J. Dijkstra and E. Klabbers. "*Strategies in Transfer Learning for Low-Resource Speech Synthesis: Phone Mapping, Features Input, and Source Language Selection*". [arXiv Preprint]; 2023.

[23] R. Huang, C. Zhang, Y. Wang, D. Yang, J. Tian, Z. Ye, L. Liu, Z. Wang, Z. Jiang, X. Chang, J. Shi, C. Weng, Z. Zhao and D. Yu. "Make-a-Voice: Revisiting Voice Large Language Models as Scalable Multilingual and Multitask Learners". In: *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*. Vol. 1. [Long Papers], pp. 10929-10942, 2024.

[24] T. Saeki, G. Wang, N. Morioka, I. Elias, K. Kastner, F. Biadsy, A. Rosenberg, B. Ramabhadran, H. Zen, F. Beaufays and H. Shemtov. "Extending multilingual speech synthesis to 100+ languages without transcribed data". In: *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, United States, pp. 11546-11550, 2024.

[25] Y. Zhang, R. J. Weiss, H. Zen, Y. Wu, Z. Chen, R. J. Skerry-Ryan, Y. Jia, A. Rosenberg and B. Ramabhadran. "*Learning to Speak Fluently in a Foreign Language: Multilingual Speech Synthesis and Cross-Language Voice Cloning*". [arXiv Preprint]; 2019.

[26] T. Saeki, S. Maiti, X. Li, S. Watanabe, S. Takamichi and H. Saruwatari. "*Learning to Speak from Text: Zero-Shot Multilingual Text-to-Speech with Unsupervised Text Pretraining*". [arXiv Preprint]; 2023.

[27] Y. Xian, B. Schiele and Z. Akata. "*Zero-Shot Learning -- the Good, the Bad and the Ugly*". [arXiv Preprint]; 2020.

[28] "*The LJ Speech Dataset*". Available from: https://www.kaggle.com/datasets/mathurinache/the-lj-speech-dataset [Last accessed on 2025 Apr 28].

[29] A. Katumba, S. Kagumire, J. Nakatumba-Nabende, J. Quinn and S. Murindanyi. "A curated crowdsourced dataset of Luganda and Swahili speech for text-to-speech synthesis". *Data Brief*, vol. 62, p. 111915, 2025.

[30] Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel and M. Le. "*Flow Matching for Generative Modeling*". [arXiv Preprint]; 2023.

[31] W. Peebles and S. Xie. "*Scalable Diffusion Models with Transformers*". [arXiv Preprint]; 2023.

[32] J. Ho, A. Jain and P. Abbeel. "*Denoising Diffusion Probabilistic Models*". [arXiv Preprint]; 2020.

[33] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and L. Polosukhin. "*Attention is All You Need*". [arXiv Preprint]; 2023.

[34] Z. Ge, L. Kaushik, M. Omote, and S. Kumar. "Speed up Training with Variable Length Inputs by Efficient Batching Strategies". In: *Interspeech*. pp. 156-160, 2021.

[35] "*Weights and Biases: The AI Developer Platform*". Weights and Biases. Available from: https://wandb.ai/site [Last accessed on 2025 May 04].

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Antimicrobial Susceptibility in Acute Myeloid Leukemia Patients in Erbil

Muna M. Najeeb[1,2]* and Sevan O. Majed[3]

[1]*Medical Microbiology Department, Collage of Health Sciences, Hawler Medical University, Erbil, Kurdistan, Iraq,*
[2]*Department of Pharmacy Technology, Kirkuk Technical Medical Institute, North Technical University, Kirkuk, 36001, Iraq,*
[3]*Department of Biology, College of Education, Salahaddin University-Erbil, Erbil, 44001, Kurdistan, Iraq.*

ABSTRACT

Patients with acute myeloid leukemia (AML) demonstrate significant sensitivity to bacterial infections. Appropriate and efficient antibiotic treatment is essential to diminish the morbidity and mortality rate of the disease. This study aimed to determine the antibacterial susceptibility profile of bacterial isolates in AML patients in Erbil City. From August 2024 to January 2025, samples were obtained from the blood, gut, and urine of 40 AML patients at Nanakali Hospital in Erbil city. Among 40 cases of AML, 49 bacteria were isolated. Gram-negative bacteria (63.27%) were more prevalent than Gram-positive bacteria (36.73%). Most Gram-negative isolates were *Escherichia coli* (34.70%) and *Klebsiella pneumoniae* (16.43%), while *Staphylococcus hominis* (10.20%) was the most common Gram-positive. Against Gram-negative isolates, colistin (100%) showed the best antibacterial action, while vancomycin and imipenem both with percentage of 100% were most successful against Gram-positive isolates. Colistin, vancomycin, and imipenem had significant efficacy, confirming their application in therapy. Continuous surveillance of resistance is essential.

Index Terms: Acute Myeloid Leukemia, Culture Media, Effectiveness, Antibiotic Susceptibility Test

1. INTRODUCTION

Acute myeloid leukemia (AML) is a severe hematologic cancer marked by an increase in the number of myeloid precursor cells in the bone marrow, leading to inefficient hematopoiesis and consequent cytopenias [1]. The disease advances quickly, and its treatment generally entails aggressive chemotherapy protocols that further weaken the immune system, making patients extremely vulnerable to opportunistic infections, especially those induced by bacterial pathogens [2]. Infections continue to be a primary cause of morbidity and mortality in AML patients, particularly during the neutropenic phase resulting from chemotherapy [3]. The impaired integrity of mucosal barriers, neutrophil dysfunction, and the recurrent use of central venous catheters create ideal conditions for microbial translocation and bloodstream infections [4]. Bacterial pathogens, including both Gram-negative and Gram-positive types, are the most common causes of infectious and often show resistance to multidrug resistance (MDR), making treatment more difficult [5]. MDR describes the ability of microorganisms, especially bacteria, to resist the effects of various structurally and functionally different antimicrobial drugs. MDR is usually defined as not responding to at least one drug from three or more different groups of antibiotics, which greatly reduces treatment options and increases morbidity, mortality, and costs associated with healthcare [6]. Bacteria acquire MDR through many molecular processes that allow them to avoid the effects of numerous antibiotics. The primary mechanisms consist of 1 - efflux pumps, bacteria utilize membrane proteins to actively expel antibiotics from the cell, hence diminishing

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp208-215

E-ISSN: 2521-4217
P-ISSN: 2521-4209

Copyright © 2025 Najeeb and Majed. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Muna M. Najeeb, Department of Pharmacy Technology Institute, Kirkuk Technical Medical Institute, North Technical University, Kirkuk,36001, Iraq. E-mail: muna@ntu.edu.iq

Received: 29-06-2025 Accepted: 20-09-2025 Published: 06-10-2025

intracellular drug concentration. 2 - Enzymatic degradation or modification, bacteria synthesize enzymes that breakdown or chemically alter antibiotics, like β-lactamases that degrade β-lactam drugs. 3 - Target modification bacteria modify the antibiotic's target site by mutation or enzymatic alteration, hence diminishing drug binding efficacy. 4 - Decreased permeability mutations in porin proteins diminish the outer membrane's permeability, hence limiting antibiotic penetration [7]. Gram-negative bacilli, especially *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa,* and *Acinetobacter baumannii*, have been progressively recognized in severe neutropenic phases among AML patients [8]. Conversely, Gram-positive bacteria, including *Staphylococcus aureus*, *Enterococcus* spp., and *coagulase-negative staphylococci*, are major contributors to nosocomial infections in this group of patients [9]. The regular and widespread use of broad-spectrum antibiotics in these patients has led to an increase in MDR organisms, such as extended-spectrum beta-lactamase producers and carbapenem-resistant enterobacteriaceae. [10]. Resistance to antimicrobial agents constitutes a global health issue and presents a particularly serious risk to immunocompromised individuals, including those with hematologic malignancies. Antibiotic resistance is identified as one of the principal worldwide public health problems [11]. This problem grows more severe in low- and middle-income nations, where diagnostic capabilities are limited, infection control measures are frequently insufficient, and surveillance data are poor [12]. In Iraq, particularly in the Kurdistan Region, there is an escalating concern about the high incidence of antibiotic-resistant bacterial strains; nevertheless, comprehensive regional statistics, particularly worrying for sensitive populations such as AML patients, remain scarce [13]. Recognizing local antimicrobial susceptibility patterns is essential for directing actual antibiotic treatment, particularly in high-risk patients. Numerous studies have highlighted the necessity of local surveillance to customize antimicrobial regimens efficiently, mitigate resistance development, and enhance patient outcomes [14]. Without this data, healthcare providers might continue using general treatments that could cause resistant bacteria to develop and make treatment less effective. No comprehensive studies have been undertaken in Erbil City to evaluate the variety of bacterial isolates and their antibiotic susceptibility profiles in AML patients. This study eliminates the gap by identifying bacterial infections in AML patients in Erbil City and comparing the *in vitro* efficacy of commonly utilized antibiotics, thereby providing data to inform treatment decisions, enhance care, optimize antibiotic utilization, and mitigate infection-related mortality in this high-risk population.

2. MATERIALS AND METHODS

2.1. Data and Specimen Collecting

During the period of August 2024–January 2025, three different biological samples including blood, gut (stool), and urine from the same patient were collected from 40 AML patients at Nanakali Hospital in Erbil City from both genders (19 males and 21 females) for different age groups under or equal to 90 years (≤90 years) after patients' approval. All samples were taken in sterile containers under sterile conditions. Blood specimens were collected in blood culture bottles, gut (stool) specimens were collected in sterile stool cups, while urine specimens were collected in urine specimen containers. Later, the samples were sent to the microbiology laboratory at Nanakali Hospital for the standard cultivation method process.

2.2. Detection and Identification of Isolated Bacteria

Each sample underwent a different microbiological process. They were promptly inoculated using appropriate selective and differential media (blood agar base, cetrimide agar, eosin methylene blue agar, MacConkey agar, mannitol salt agar, Mueller–Hinton agar, and nutrient agar) at a temperature of 35–37°C for 24–48 h. Gram staining was achieved. All isolated bacteria were recognized depending on their cultural characteristics and biochemical analysis, including lipase, indole, motility, oxidase, catalase, and coagulase, manually to confirm the results using the VITEK 2 Compact System.

2.3. Antimicrobial Susceptibility Test for Isolated Bacteria

Ten different antibiotics from several classes of different modes of action on all of the bacteria that were isolated – imipenem, ceftriaxone, vancomycin, clindamycin, rifampin, ciprofloxacin, levofloxacin, amikacin, gentamicin, and colistin – were used to confirm the resistance profiles of the isolates. Results were interpreted according to the Clinical and Laboratory Standards Institute (CLSI) (CLSI, 2023) recommendations after the testing was conducted on Mueller–Hinton agar using the Kirby–Bauer disk diffusion method. After that, we confirmed the results of the resistance profile using the VITEK 2 compact system. The VITEK 2 compact system facilitated automated identification and susceptibility testing, enabling a more efficient investigation of bacterial samples. This comprehensive approach not only verified resistance patterns but also enabled the finding of potential treatment options suited to handle the particular types of bacteria identified.

2.4. Approval of Ethical Guidelines and Authorization

for Participation

The local Human Research Ethics Committee at the Medical Microbiology department, Health Sciences College, Hawler Medical University-Erbil has approved and allowed the whole process carried out in this research project involving human participants, human materials, or human data (Reference number 4c/132). Every technique used during the study followed the Declaration of Helsinki, as revised in 2013, and all research subjects had written informed consent for publishing.

3. RESULTS

3.1. Distribution of AML Patients According to Age and Gender

Table 1 shows the distribution of 40 cases collected from patients with AML cancer. The distribution of cases according to different age groups showed that the highest proportions of cases were 10 (25%) belonged to the 21–30-year age groups, followed by a proportion of 5 (12.5%), which was the same proportion of four different age groups including 0–10 years, 11–20 years, 31–40 years, and 71–80 years. 41–50-year age groups had the lowest proportion 2 (5%) while no cases were recorded in the 81–90-year age group. In terms of gender distribution, male cases composed 19 (47.5%) of all samples, while females accounted for 21 (52.5%), indicating a slightly higher female proportion, as shown in Table (1). Among males, the highest proportion, 7 (17.5%) was observed in the 21–30 age group, followed by the 0–10 age group, 4 (10%), whereas female cases showed that 11–20, 31–40, and 71–80 age groups each had the highest proportion, 4 (10%). No cases of either gender were recorded in the 81–90-year age groups, as in (Table 1).

3.2. Frequency of Positive and Negative Results Across Various Sample Types

Table 2 shows the frequency of culture-positive and culture-negative results from a total of 120 biological samples obtained from patients, including 40 blood samples, 40 urine samples, and 40 gut samples. A total of 55 samples (45.8%) exhibited positive cultures, signifying bacterial growth, whereas 65 samples (54.2%) demonstrated negative cultures. Among the blood samples, 19 (47.5%) exhibited bacterial growth (culture-positive), while 21 (52.5%) showed no microbial growth (culture-negative). Urine samples exhibited the highest positivity rate, with 28 out of 40 samples (70%) demonstrating microbial growth, whereas 12 samples (30%) indicated no growth. In contrast, gut samples displayed the lowest incidence of positive outcomes, with merely 8 samples (20%) indicating bacterial growth and 32 samples (80%) producing negative culture results.

3.3. Distribution of Bacterial Species across Diverse Sample Types

A total of 49 bacterial strains were recovered from 40 patients diagnosed with AML, utilizing diverse biological materials, including blood, urine, and gut. The bacterial diversity is summarized in Fig. 1. *E. coli* was the most frequently found harmful bacterium, comprising 17 (34.70%) of the total isolates. *K. pneumoniae* was the second most prevalent pathogen,

TABLE 1: Distribution of AML patients according to age and gender

Age groups		Gender	
Time intervals	No (%)	Male-No (%)	Female-No (%)
0–10	5 (12.5)	4 (10)	1 (2.5)
11–20	5 (12.5)	1 (2.5)	4 (10)
21–30	10 (25)	7 (17.5)	3 (7.5)
31–40	5 (12.5)	1 (2.5)	4 (10)
41–50	2 (5)	2 (5)	0 (0)
51–60	4 (10)	2 (5)	2 (5)
61–70	4 (10)	1 (2.5)	3 (7.5)
71–80	5 (12.5)	1 (2.5)	4 (10)
81–90	0 (0)	0 (0)	0 (0)
Total	40 (100)	19 (47.5)	21 (52.5)

AML: Acute myeloid leukemia

TABLE 2: Frequency of positive outcomes across various sample types

Types of samples	Growth No (%)	No growth No (%)	Total No (%)
Blood	19 (47.5)	21 (52.5)	40 (100)
Urine	28 (70)	12 (30)	40 (100)
Gut	8 (20)	32 (80)	40 (100)

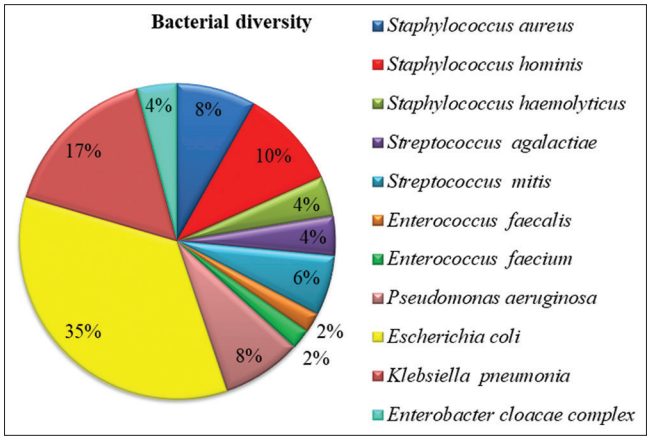


Fig. 1. Distribution of bacterial species across diverse sample types.

accounting for 8 cases (16.43%), followed by *Staphylococcus hominis* with 5 cases (10.20%), and both *P. aeruginosa* and *S. aureus*, each with 4 cases (8.16%). *Streptococcus agalactiae* and *Staphylococcus haemolyticus*, both Gram-positive bacteria, were each found in 2 (4.08%) of the total isolates. *Streptococcus mitis* was identified in 3 (6.12%). *Enterococcal* species, specifically *Enterococcus faecium* and *Enterococcus faecalis*, were infrequently observed, each representing 1 (2.04%) of the isolated pathogens. Based on the distribution of bacterial isolates among biological specimens. Blood samples tested positive for several microorganisms, including *S. aureus*, *S. hominis*, *S. mitis,* and *P. aeruginosa*. Urine samples exhibited the highest prevalence of *E. coli* and *K. pneumoniae* pathogens. Gut-related samples tested positive for *E. coli* and *the Enterobacter cloacae complex*.

The pie chart shows the distribution of bacterial species isolated across the study. *E. coli* was recognized as the predominant species, followed by *K. pneumoniae* and *S. hominis*. Additional commonly isolated bacteria comprised *S. aureus*, *P. aeruginosa*, and *S. mitis*. Species found less frequently were *S. haemolyticus*, *S. agalactiae*, *E. faecium*, and *E. faecalis*, among others. The isolates include a wide spectrum of both Gram-positive and Gram-negative bacterial pathogens, underscoring the microbial diversity among the samples.

3.4. Characterization of Various Bacterial Species on Different Types of Culture Media

For microbiological research and diagnostics, bacterial isolates must be identified and characterized. The morphological and biochemical characteristics of several bacterial species are investigated in this study utilizing a variety of culture media, such as Cetrimide agar, Eosin Methylene Blue (EMB) agar, Mannitol Salt agar, MacConkey agar, and Blood agar. With different cellular structures, including clustered cocci, short chains, diplococci, and rod-shaped forms, Gram staining was used to classify the isolates as either Gram-positive or Gram-negative bacteria. *P. aeruginosa* was selectively isolated on Cetrimide agar, *E. coli* produced distinctive metallic green colonies on Eosin Methylene Blue agar, and *S. aureus* displayed β-hemolysis on Blood agar and mannitol fermentation on Mannitol Salt agar. MacConkey agar successfully distinguished lactose fermenters such as *K. pneumoniae*, *E. coli*, and Enterobacter cloacae complex which showed red or pink colonies as shown in Table 3.

3.5. Antibiotic Susceptibility Profile for Isolated Pathogenic Bacteria

The antibiotic susceptibility profile was conducted on 18 isolated Gram-positive and 31 Gram-negative bacteria to determine their sensitivity, intermediate response, and resistance to different antibiotics. The selected antibiotics were categorized based on their mechanism of action. For Gram-positive bacteria18 bacteria were isolated, the first class of antibiotics that inhibited cell wall synthesis was classified into three groups, each containing different examples. Among those, imipenem and vancomycin were both effective to all cases of Gram-positive bacteria 18 (100%). However, ceftriaxone was not effected in 17 (94.44%) cases. The second class of antibiotics targeted protein synthesis inhibition and consisted of two groups with different antibiotics. Among those, clindamycin and rifampin showed high resistance proportions of 11 (61.11%) and 16 (88.89%), respectively. The last class of antibiotics focused on the inhibition of nucleic acid synthesis. This class included only one group with two different types of antibiotics. Among them, ciprofloxacin and levofloxacin both exhibited high resistance rates, with 14 (77.78%) and 10 (55.56%) cases, respectively. However, Levofloxacin also showed a notable proportion of sensitivity in 8 (44.44%) cases, suggesting its potential as a good treatment option. These findings are shown in Table 4. A total of 31 Gram-negative bacterial isolates were tested for antibiotic susceptibility, revealing varying levels of sensitivity and resistance across different antibiotic classes. The first class of antibiotics working on the inhibition of cell wall synthesis contains two antibiotics, imipenem and ceftriaxone. Imipenem was highly sensitive in 19 (61.29%) cases while ceftriaxone showed high resistance, with 28 (90.32%) cases being resistant. For the second class, which was inhibition of protein synthesis, aminoglycosides were used. This group contains two antibiotics, amikacin and gentamicin. Both antibiotics had the same resistance rate of 7 (22.58%) cases. Gentamicin was highly sensitive in 24 (77.42%) cases, whereas amikacin showed lower sensitivity in comparison. Another class was the disruption of the cell membrane which contained only one group. Polymyxin E like colistin was tested and showed no resistance, indicating 100% sensitivity. The last class was the inhibition of nucleic acid synthesis. Fluoroquinolones and ciprofloxacin exhibited high resistance in 29 (93.55%) cases; levofloxacin showed high sensitivity in 19 (61.29%) cases. These findings shown in Table 5 suggest that imipenem, gentamicin, polymyxin E, and levofloxacin may be more effective treatments.

3.6. Antibiotic Resistance in Isolated Pathogenic Bacterial Strain

The antibiotic resistance profile of 11 different bacterial species was estimated against a total of 10 antibiotics. The results indicate varying resistance levels among different isolates. *S. aureus* and *K. pneumoniae* exhibited the highest resistance, with 6 out of 10 antibiotics (60%) showing

TABLE 3: Characterization of various bacterial species on different types of culture media

Cetrimide agar	Mannitol salt agar	Eosin methylene blue	Nutrient agar	MacConkey agar	Blood agar	Gram stain	Cell morphology	Isolated bacteria
N. G	Round, creamy to golden, media yellow	N. G	Round, golden yellow, or creamy	N. G	Round, golden yellow β-hemolysis	G +ve	Cocci in cluster	*Staphylococcus aureus*
N. G	Round, white to off white media pink	N. G	Round, white to creamy	N. G	Round, off white, gamma hemolysis	G +ve	Cocci in cluster	*Staphylococcus hominis*
N. G	Round, white to creamy media pink	N. G	Round, creamy, or off white	N. G	Round, off white, β-hemolysis	G +ve	Cocci in cluster	*Staphylococcus haemolyticus*
N. G	N. G	N. G	Round, off white or grayish	N. G	Round, creamy, or grayish, β-hemolysis	G +ve	Diplococci short chain	*Streptococcus agalactiae*
N. G	N. G	N. G	Round, smooth, grayish	N. G	Round, off white, γ-hemolysis	G +ve	Cocci, chain	*Streptococcus mitis*
N. G	N. G	N. G	Round, gray, or cream color	N. G	Round, off white, γ-hemolysis	G +ve	Cocci, diplococcic (Pair)	*Enterococcus faecium*
N. G	N. G	N. G	Round, creamy, or grayish	N. G	Round, grayish γ-hemolysis	G +ve	Cocci, diplococci (Pair)	*Enterococcus faecalis*
Moist, smooth, greenish-blue, metallic sheen	N. G	Smooth, off white, slight sheen	Round, greenish	Smooth, round, green	Round, green, or greenish blue, γ-hemolysis	G -ve	Bacillus, single rods	*Pseudomonas aeruginosa*
N. G	N. G	Smooth, moist, greenish metallic sheen	Round, moist creamy, or grayish	Smooth, moist, pink or red	Round, off white, γ-hemolysis	G -ve	Bacillus, single rods or chain	*Escherichia coli*
N. G	N. G	Mucoid, dark pink or purple with metallic sheen	Mucoid, off white or grayish	Large moist, mucoid, pink or red	Mucoid, creamy, off-white hemolysis	G -ve	Bacillus, single rods or chain	*Klebsiella pneumoniae*
N. G	N. G	Smooth, moist, light pink	Moist, creamy, or off white	Large moist, off white to pink	Smooth, pale yellow, γ- hemolysis	G -ve	Bacillus, short chain or rods	*Enterobacter cloacae complex*

G+ve: Gram positive, G –ve: Gram negative, N.G: No growth, β-hemolysis: Beta hemolysis, γ-hemolysis: Gamma hemolysis

TABLE 4: Antibiotic susceptibility profile for isolated pathogenic Gram-positive bacteria

Classification according to the mechanism of action	Groups	Antibiotics	Interpretation		
			Sensitive (%)	Inter-mediate (%)	Resistance (%)
Inhibition of cell wall synthesis	Beta-lactam (carbapenem)	Imipenem	18 (100)	/	/
	Beta-lactam (cephalosporin)	Ceftriaxone	1 (5.56)	/	17 (94.44)
	Glycopeptide	Vancomycin	18 (100)	/	/
Inhibition of protein synthesis	Lincosamides	Clindamycin	7 (38.89)	/	11 (61.11)
	Rifamycin	Rifampin	2 (11.11)	/	16 (88.89)
Inhibition of nucleic acid synthesis	Fluoroquinolones	Ciprofloxacin	/	4 (22.22)	14 (77.78)
	Fluoroquinolones	Levofloxacin	8 (44.44)	/	10 (55.56)

Number of Gram-positive bacteria=18

TABLE 5: Antibiotic susceptibility profile for isolated pathogenic Gram-negative bacteria

Classification according to the mechanism of action	Groups	Antibiotics	Interpretation		
			Sensitive (%)	Inter-mediate (%)	Resistance (%)
Inhibition of cell wall synthesis	Beta-lactam (Carbapenem)	Imipenem	19 (61.29)	1 (3.23)	11 (35.48)
	Beta-lactam (cephalosporin)	Ceftriaxone	3 (9.68)	/	28 (90.32)
Inhibition of protein synthesis	Aminoglycoside	Amikacin	21 (67.74)	3 (9.68)	7 (22.58)
	Aminoglycoside	Gentamicin	24 (77.42)	/	7 (22.58)
Disruption of cell membrane	Polymyxin E	Colistin	31 (100)	/	
Inhibition of nucleic acid synthesis	Fluoroquinolones	Ciprofloxacin	2 (6.45)	/	29 (93.55)
	Fluoroquinolones	Levofloxacin	19 (61.29)	/	12 (38.71)

Number of Gram-negative bacteria=31

ineffectiveness. *S. hominis*, *S. mitis*, *E. faecium*, *E. coli*, and *Enterobacter cloacae* demonstrated resistance to 5 antibiotics (50%). *Staphylococcus haemolyticus* and *P. aeruginosa* showed intermediate resistance, with 4 antibiotics (40%) being ineffective. *S. agalactiae* and *E. faecalis* exhibited the lowest resistance rates, with only 3 antibiotics (30%) showing ineffectiveness. This was indicated for multidrug-resistant bacteria; most of those bacteria were resistant to more than three antibiotics. This is all summarized in Table 6.

TABLE 6: Antibiotic resistance in isolated pathogenic bacterial strain

Isolated bacteria	Number of antibiotic resistances (Total number of antibiotics=10) (%)
Staphylococcus aureus	6 (60)
Staphylococcus hominis	5 (50)
Staphylococcus haemolyticus	4 (40)
Streptococcus mitis	5 (50)
Streptococcus agalactiae	3 (30)
Enterococcus faecium	5 (50)
Enterococcus faecalis	3 (30)
Escherichia coli	5 (50)
Pseudomonas aeruginosa	4 (40)
Klebsiella pneumoniae	6 (60)
Enterobacter cloacae	5 (50)

4. DISCUSSION

This study's findings offer essential insights into the antimicrobial susceptibility patterns of AML patients in Erbil, Iraq, while also emphasizing significant similarities with global antimicrobial resistance trends. The significant frequency of Gram-negative bacteria (63.27%), especially *E. coli* (34.70%) and *K. pneumoniae* (16.43%), reflects trends noted in other areas where these pathogens predominantly cause infections in immunocompromised individuals. A multicenter European investigation showed analogous data, indicating that Enterobacteriaceae constituted 58% of bloodstream infections in patients with hematological malignancies [8]. The prevalence of these organisms likely indicates their capacity to inhabit the gastrointestinal tract and translocate after mucosal damage generated by chemotherapy, a feature well-documented in neutropenic patients [3]. The prevalence of Gram-positive isolates is 36.73%, with *S. hominis* at 10.20%, consistent with U.S. hospital findings indicating that *coagulase-negative staphylococci* comprised 31% of nosocomial bloodstream infections in cancer patients. The prevalence of *P. aeruginosa* (8.16%) is particularly alarming because of its inherent resistance mechanisms and correlation with adverse outcomes in neutropenic individuals. This discovery aligns with Italian data indicating that *P. aeruginosa* is the etiological agent in 12% of Gram-negative infections inside hematological units [15]. The antibiotic resistance patterns identified in this study indicate both regional particularities and global alarming tendencies. The notably elevated resistance to ceftriaxone (90.32% in Gram-negative isolates) surpasses the rates documented in adjacent Middle Eastern nations, where resistance to third-generation cephalosporins varies from 60% to 75% [13]. The ciprofloxacin resistance incidence of 93.55% significantly exceeds the 65% reported in Chinese hematological centers [16]. Potentially indicating variations in antibiotic prescribing policies and infection control protocols. The total sensitivity to colistin (100%) and the significant efficiency of imipenem (61.29%) against Gram-negative isolates align with findings from European studies, where these antibiotics are considered last-resort choices [17]. The 35.48% resistance to imipenem is concerning and exceeds the 15–20% resistance rates documented in Western Europe, indicating the emergence of carbapenem resistance in our region that requires urgent attention. The 100% susceptibility of Gram-positive isolates to vancomycin and imipenem is promising and aligns with global trends, since these antibiotics continue to be fundamental in managing resistant Gram-positive infections [6]. The elevated resistance rates to rifampin (88.89%) and ciprofloxacin (77.78%) indicate the repercussions of extensive prophylactic usage, akin to

trends noted in Indian cancer centers, where fluoroquinolone resistance escalated to 80% after the introduction of routine prophylaxis [12]. The patterns of MDR are notably concerning, since 60% of *S. aureus* and *K. pneumoniae* isolates exhibit resistance to six or more classes of antibiotics. This surpasses the 40–50% MDR rates documented in Turkish hematological patients [5] and nears the 70% resistance shown in certain Indian studies [12]. situating our region among those with the most acute resistance challenges worldwide. Various causes may contribute to these concerning resistance patterns. The absence of effective antimicrobial stewardship programs in Iraqi hospitals, along with the accessibility of over-the-counter antibiotics, undoubtedly contributes to the misuse of antibiotics [13]. Moreover, the scarcity of infection management resources in our environment may promote the propagation of resistance strains, as evidenced in other resource-limited settings [12]. The elevated incidence of extended-spectrum β-lactamase-producing Enterobacteriaceae, indicated by ceftriaxone resistance rates, aligns with global patterns but is more pronounced in our location. These findings possess significant clinical ramifications. The elevated resistance rates to frequently employed antibiotics such as ceftriaxone and ciprofloxacin indicate that they should be discontinued for empirical usage in our context. Treatment guidelines ought to account for local resistance trends, potentially prioritizing combination therapy or novel medicines when accessible. The sustained efficacy of colistin, however promising, prompts apprehension over the necessity of increased utilization of this last-resort antibiotic, thereby hastening the emergence of resistance [17]. The study's shortcomings comprise its single-center methodology and limited sample size, potentially impacting generalizability. The absence of molecular characterization of resistance mechanisms hinders a comprehensive knowledge of resistance transmission patterns. Subsequent research should integrate genomic techniques to pinpoint individual resistance genes and clonal affiliations, as demonstrated in European institutions [18].

This study indicates significantly elevated antibiotic resistance rates among AML patients in Erbil, surpassing other international reports. The results highlight the critical necessity for extensive antimicrobial stewardship initiatives, improved infection control protocols, and investment in swift diagnostic technologies. Given the substantial regional variation in resistance patterns, our findings underscore the necessity of local surveillance to inform treatment plans. The circumstances necessitate prompt measures to conserve the remaining effective antibiotics and avert a post-antibiotic era in hematological malignancies.

5. CONCLUSION

This study highlights the bacterial diversity and antibiotic resistance in AML patients in Erbil. *E. coli* and *K. pneumoniae* were the most common isolates. High resistance to ceftriaxone, ciprofloxacin, and rifampin was observed, while colistin, imipenem, and vancomycin remained effective. The findings stress the need for local antimicrobial stewardship, regular surveillance, and improved infection control. In the absence of national data, this research fills a critical gap and calls for molecular studies to guide future treatment and policy.

6. ACKNOWLEDGMENT

The authors are thankful to Hawler Medical University and the College of Health Sciences for giving us the opportunity to carry out this research paper.

7. FUNDING

No funding.

REFERENCES

[1] H. Döhner, A. H. Wei, F. R. Appelbaum, C. Craddock, C. D. DiNardo, H. Dombret, B. L. Ebert, P. Fenaux, L. A. Godley, R.P.Hasserjian,R.A.Larson,R.L.Levine,Y.Miyazaki,D.Niederwieser, G. Ossenkoppele, C. Röllig, J. Sierra, E. M. Stein, M. S. Tallman, H. F. Tien, J. Wang, A. Wierzbowska and B. Löwenberg. "Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN". *Blood*, vol. 140, pp. 1345-1377, 2022.

[2] E. H. Estey. "Acute myeloid leukemia: 2021 update on risk-stratification and management". *American Journal of Hematology*, vol. 95, pp. 1368-1398, 2020.

[3] C. Gudiol and J. Carratalà. "Antibiotic resistance in cancer patients". *Expert Review of Anti-Infective Therapy*, Vol. 12, pp. 1003-1016, 2014.

[4] A. G. Freifeld, E. J. Bow, K. A. Sepkowitz, M. J. Boeckh, J. I. Ito, C. A. Mullen, I. I. Raad, K. V. Rolston, J. A. Young, J. R. Wingard and Infectious Diseases Society of America. "Clinical practice guideline for the use of antimicrobial agents in neutropenic patients with cancer: 2010 update by the infectious diseases society of America". *Clinical Infectious Diseases*, vol. 52, pp. e56-e93, 2011.

[5] D. Mert, S. Ceken, G. Iskender, D. Iskender, A. Merdin, F. Duygu and M. Ertek, F. "Altuntas. Epidemiology and mortality in bacterial bloodstream infections in patients with hematologic malignancies". *The Journal of Infection in Developing Countries*, vol. 13, pp. 727-735, 2019.

[6] E. Tacconelli, E. Carrara, A. Savoldi, S. Harbarth, M. Mendelson, D. L. Monnet, C. Pulcini, G. Kahlmeter, J. Kluytmans, Y. Carmeli, M. Ouellette, K. Outterson, J. Patel, M. Cavaleri, E. M. Cox, C. R. Houchens, M. L. Grayson, P. Hansen, N. Singh, U. Theuretzbacher, N. Magrini and WHO Pathogens Priority List

Working Group. “Discovery, research, and development of new antibiotics: The WHO priority list of antibiotic-resistant bacteria and tuberculosis”. *The Lancet Infectious Diseases*, vol. 18, pp. 318-327, 2018.

[7] W. C. Reygaert. “An overview of the antimicrobial resistance mechanisms of bacteria”. *AIMS Microbiology*, vol. 4, p. 482, 2018.

[8] E. M. Trecarichi, G. Giuliano, C. Cattaneo, S. Ballanti, M. Criscuolo, A. Candoni, F. Marchesi, M. Laurino, M. Dargenio, R. Fanci, M. Cefalo, M. Delia, A. Spolzino, L. Maracci, V. Bonuomo, A. Busca, M. I. D. Principe, R. Daffini, E. Simonetti, G. Dragonetti, M. E. Zannier, L. Pagano, M. Tumbarello and Hematologic Malignancies Associated Bloodstream Infections Surveillance (HEMABIS) Registry - Sorveglianza Epidemiologica Infezioni Fungine in Emopatie Maligne (SEIFEM) group, Italy. “Bloodstream infections due to Gram-negative bacteria in patients with hematologic malignancies: Updated epidemiology and risk factors for multidrug-resistant strains in an Italian perspective survey”. *International Journal of Antimicrobial Agents*, 61, pp. 106806, 2023.

[9] H. Wisplinghoff, T. Bischoff and S. M. Tallent. “Nosocomial bloodstream infections in US hospitals: Analysis of 24,179 cases from a prospective nationwide surveillance stu. “dy”. *Clinical Infectious Diseases*, vol. 39, pp. 309-317, 2004.

[10] E. M. Trecarichi, L. Pagano, B. Martino, A. Candoni, R. Di Blasi, G. Nadali, L. Fianchi, M. Delia, S. Sica, V. Perriello, A. Busca, F. Aversa, R. Fanci, L. Melillo, F. Lessi, M. I. Del Principe, C. Cattaneo, M. Tumbarello and Haematologic Malignancies Associated Bloodstream Infections Surveillance (HEMABIS) Registry - Sorveglianza Epidemiologica Infezioni Funginein Emopatie Maligne(SEIFEM) Group, Italy. “Bloodstream infections caused by *Klebsiella pneumoniae* in onco-hematological patients: Clinical impact of carbapenem resistance in a multicentre prospective survey”. *American Journal of Hematology*, vol. 91, pp. 1076-1081, 2016.

[11] S. F. Nadeem, U. F. Gohar, S. F. Tahir, H. Mukhtar, S. Pornpukdeewattana, P. Nukthamna, A. M. Moula Ali, S. C. B. Bavisetty and S. Massa. “Antimicrobial resistance: More than 70 years of war between humans and bacteria”. *Critical Reviews in Microbiology*, 46, pp. 578-599, 2020.

[12] R. Laxminarayan, A. Duse, C. Wattal, A. K. Zaidi, H. F. Wertheim, N. Sumpradit, E. Vlieghe, G. L. Hara, I. M. Gould, H. Goossens, C. Greko, A. D. So, M. Bigdeli, G. Tomson, W. Woodhouse, E. Ombaka, A. Q. Peralta, F. N. Qamar, F. Mir, S. Kariuki, Z. A. Bhutta, A. Coates, R. Bergstrom, G. D. Wright, E. D. Brown and O. Cars. “Antibiotic resistance-the need for global solutions”. *The Lancet Infectious Diseases*, vol. 13, pp. 1057-1098, 2013.

[13] I. A. Naqid, A. A. Balatay, N. R. Hussein, H. A. Ahmed, K. A. Saeed and S. A. Abdi. “Bacterial strains and antimicrobial susceptibility patterns in male urinary tract infections in Duhok province, Iraq”. *Middle East Journal of Rehabilitation and Health Studies*, vol. 7, pp. e103529, 2020.

[14] T. Vieceli and J. Rello. “Optimization of antimicrobial prescription in the hospital”. *European Journal of Internal Medicine*, vol. 106, pp. 39-44, 2022.

[15] E. M. Trecarichi and M. Tumbarello. “Antimicrobial-resistant Gram-negative bacteria in febrile neutropenic patients with cancer: Current epidemiology and clinical impact”. *Current Opinion in Infectious Diseases*, 27, pp. 200-210, 2014.

[16] S. Zhang, Q. Wang, Y. Ling and X. Hu. “Fluoroquinolone resistance in bacteremic and low risk febrile neutropenic patients with cancer. *BMC Cancer*, vol. 15, p. 42, 2015.

[17] T. Tängdén. “Combination antibiotic therapy for multidrug-resistant Gram-negative bacteria”. *Upsala Journal of Medical Sciences*, vol. 119, pp. 149-153, 2014.

[18] C. Costantini, E. Nunzi, A. Spolzino, M. Palmieri, G. Renga, T. Zelante, L. Englmaier, K. Coufalikova, Z. Spáčil, M. Borghi, M. M. Bellet, E. Acerbi, M. Puccetti, S. Giovagnoli, R. Spaccapelo, V. N. Talesa, G. Lomurno, F. Merli, L. Facchini, A. Spadea, L. Melillo, K. Codeluppi, F. Marchesi, G. Marchesini, D. Valente, G. Dragonetti, G. Nadali, L. Pagano, F. Aversa and L. Romani. “Pharyngeal microbial signatures are predictive of the risk of fungal pneumonia in hematologic patients”. *Infection and Immunity*, vol. 89, pp. e0010521, 2021.

RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

LectBench-95: Preparing a University-Lecture Corpus for A/B Evaluation of Lecture-Processing Techniques

Ari A. Aziz, Aree A. Mohammed

Department of Computer Science, College of Science, University of Sulaimani, Sulaymaniyah, Kurdistan Region, Iraq.

ABSTRACT

Tools for processing educational lectures are rapidly advancing, but there is a need for a diverse, balanced, and high-quality university lecture transcript corpus. The existing datasets are either limited to K-12, tutorial styled, and lack interactivity, focus on narrow disciplines, paywalled/non-accessible, or are impractically large. We introduce LectBench-95, a publicly available corpus of 95 video lecture transcripts spanning 3 disciplines and 17 specific subjects within them. With strict filtering for high audio quality (SNR ≥25 dB), transcription confidence (Mean 0.84, Min 0.7), transcript quality controls, and a power analysis-guided sample size, the 94-h dataset aims to detect ≥20% performance differences between competing systems with 95% confidence for head-to-head A/B experiments. LectBench-95 contains 816 k words (unique ≈ 123 k), a mean Measure of Textual Lexical Diversity of 54.5, and a mean speech rate of 144 words/min, mirroring real-world university lectures. A toy A/B test on zero-shot summarization (Gemini-1.5-Flash vs. 1.5-Flash-8B) shows the corpus's utility, resulting in a statistically significant 43% win-rate gap with $P = 3 \times 10^{-5}$. Released under CC BY-NC-SA 4.0, LectBench-95 provides a modest yet statistically robust dataset for future educational natural language processing research and prototyping.

Index Terms: University Lecture Corpus, A/B Evaluation, Lecture Processing Techniques, Educational Natural Language Processing

1. INTRODUCTION

In the information age, advancing pedagogical tools are more important than ever, requiring every step in the development process to be efficient and effective. Evaluation, in particular, can strongly impact the success of such tools. For lecture processing techniques, this means a dataset that is reflective of real-world university lectures. Such techniques have been on the rise, especially with the advent of Large Language Models (LLMs) [1], [2]. Examples of educational LLM-based tools include: question answering teacher assistant [3]; screening children's language development levels [4]; providing lesson plans, activities, and materials [5]; and automatic grading of student responses [6]. However, the advancement of these tools requires a solid dataset that provides accurate feedback for pinpointing improvement areas.

Current university-lecture corpora often lack the properties needed for reliable A/B comparisons of lecture processing techniques on classroom discourse, including (1) interactive, professor-led lectures rather than tutorials and monologues; (2) multidisciplinary coverage that reveals subject-specific blind spots; (3) rigorous filtering and transcript quality controls; (4) public accessibility at a practical size. This

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp216-230 E-ISSN: 2521-4217 P-ISSN: 2521-4209

Copyright © 2025 Ari A. Aziz. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Ari A. Aziz, Department of Computer Science, College of Science, University of Sulaimani, Sulaymaniyah, Kurdistan Region, Iraq. E-mail: ari.aryan@univsul.edu.iq

Received: 26-06-2025 Accepted: 23-08-2025 Published: 11-10-2025

limits both model evaluators who are selecting systems and researchers studying classroom discourse.

To address this, we introduce LectBench-95, a dataset of 95 carefully curated, diverse, and high-quality university lecture transcripts, sourced from various institutions. It is designed for head-to-head A/B experiments in mind, with pre-declared acceptance criteria: (a) ≥3 subjects per area across STEM/Humanities/Social Sciences; (b) SNR ≥25 dB and mean Whisper confidence ≥0.70 per lecture; (c) 30–120 min lecture durations; and (d) a statistical power-aware sample size that targeting the detection of moderate win-rate differences in pairwise tests. Each transcript includes metadata and segment-level timestamps, allowing for use cases that go beyond those outlined in this paper. this dataset is publicly available at this link https://osf.io/astqv/.

This paper provides a comprehensive overview of the LectBench-95 dataset, including its collection, processing, and the rationale behind its design choices. Detailed statistics are presented to characterize the dataset and its applications. To evaluate the statistical significance of the results obtained from experimenting with this dataset, a toy A/B experiment is conducted that demonstrates the dataset's benefits when it comes to evaluating two simple lecture summarization techniques.

The videos are collected from publicly available university lectures and transcribed using a state-of-the-art speech-to-text system. The rest of the paper details the process.

Our contributions include:

1. A curated, interactive university-lecture corpus spanning 17 subjects with timestamps and quality summaries
2. A transparent method that allows selecting and auditing for interactivity
3. A power-aware evaluation recipe for paired win-rate comparisons
4. A demonstration of zero-shot summarization showing statistically significant model differentiation.

The remainder of the paper is organized as follows: Section 2 discusses the current educational lecture datasets. Section 3 details the dataset's construction, including collection and processing, statistical design, and lecture selection criteria. Section 4 provides both a quick statistical summary and a more detailed breakdown of the dataset's characteristics, including lexical, speech rate, and quality statistics. Section 5 presents an A/B experiment using the dataset to demonstrate its utility in evaluating lecture processing techniques. Section 6 outlines the intended use cases (including implementation and analysis opportunities) and limitations of the dataset, both of which suggest future work.

2. RELATED RESOURCES

Diverse open-source datasets in the domain of university lectures are somewhat scarce, especially those that focus on traditional teaching. This scarcity creates obstacles for researchers who require rich, varied data reflective of real classroom dynamics for developing various types of educational technologies. The current corpora often focus narrowly on specific subjects or delivery formats (e.g., online tutorials); they may even focus more thoroughly on pre-university level courses. While there are also some that can tick these boxes, they may fall short in terms of accessibility or usability. In this section, such datasets are explored, and arguments will be made for the limitations that may hold true in this context.

The National Center for Teacher Effectiveness (NCTE) Transcripts [7] is one dataset, which contains 1660 elementary math classroom transcripts, each 45–60 min long, collected by the NCTE between 2010 and 2013. The lessons are from 4th and 5th grade math classes in the United States. It includes rich turn-level metadata of dialogic discourse, such as questioning and prompting. Furthermore, it includes classroom observation scores, teacher and student demographics, survey responses, and student test scores. However, the dataset is limited only to elementary math classes and does not include higher education or other diverse subjects. Another somewhat similar dataset is the MET dataset [8], which contains 2500 4–9th grade classroom recordings collected between 2009 and 2011 in the US. It spanned multiple subjects, including English Language Arts and mathematics. Its downsides are that it requires a subscription; it only focuses on 4th–9th-grade lessons, and it does not provide transcripts. The TalkMoves dataset [9] is another math classroom transcript dataset containing 567 K-12 math lectures, also collected in the US. Despite being publicly available, it is limited in terms of subjects and to K-12 education. Another dialogue dataset is the CIMA dataset [10], which is designed specifically with training deep learning models in mind for Italian tutoring. It was collected through asynchronous role-playing by crowd-workers meant to depict Italian tutoring lessons. It is ideal for tutor agent training, but is limited to Italian. Rai *et al.* [11] investigate word error rates (WER) between YouTube Automatic Captions and OpenAI's Whisper model. It focuses on NPTEL (a large

MOOC platform in India) videos, and it collects 8740 h of Indian English lecture audio covering 9800 lectures sourced from the aforementioned platform. The Technical Indian English dataset mentioned includes audio, transcripts, and speaker metadata. This dataset focuses mostly on non-native varieties of English (specifically Indian English) that are technically dense for the purpose of improving ASR systems. The Autoblog 2020 [12] and its subsequent version Autoblog 2021 [13] are two corpora—the latter is publicly available on Kaggle—that focus on online learning video lectures. The main aim of Autoblog 2020 was to convert lectures into easy-to-consume blog posts in an automated manner, while its subsequent version aimed to increase the dataset by adding manual transcription for ASR benchmarking of spontaneous speech. Autoblog 2021 includes 63 lecture videos and audio files, manual and automatic transcriptions, slide images, and more. Its limitations include that it is not a traditional in-class lecture with teacher-student engagements in a classroom; rather, the teaching is performed in a monologous manner in an online setting. Furthermore, it is limited only to Medical Engineering and Pattern Recognition. The Automatically Recognizing Lecture Highlighting Corpus [14] is another corpus meant for training speech processing models for recognizing moments where the lecturer highlights or emphasizes a certain word using vocal emphasis. The dataset features 104 different English speakers from various disciplines derived from YouTube tutorial videos. Despite the authors saying that it would be made freely available to the community, no access link was found to the dataset as of the time of writing this paper. Furthermore, it focuses on tutorials rather than formal university lecture transcripts. VT-SSum [15] is a benchmark dataset that contains 125K transcript-summary pairs derived from 9,616 video transcripts from videolectures.net across 26 different categories. It focuses on spoken language summarization and segmentation and leverages the accompanying slides as weak supervision. However, it uses Microsoft speech-to-text, while this dataset utilizes the more modern, accurate, and open-source OpenAI Whisper [16]. The AVLectures dataset [17] is another dataset that focuses on audio-visual lecture segmentation and summarization. It has 86 courses (15 of them manually segmented), encompassing more than 2350 lectures (2,200 h total) spanning STEM fields such as EECS, Physics, and Mathematics. Despite being publicly accessible and despite the authors claiming that you can access each of the course's tarballs individually, the dataset webpage[1] provides 3 tarballs only where one is 340 GB and the other two are 79 GB and 1 GB, respectively. This makes it impractical to use for researchers seeking smaller or modest-sized datasets. This argument is true for the Slide Speech dataset [18] as well.

In addition, several other datasets comprise MOOC lecture transcripts, which do not align with the goal of collecting traditional university lectures. For example, the Khan Academy Corpus [19] contains a large collection of transcripts from Khan Academy videos, which are primarily tutorial style and also require affiliation with certain universities to access. Similarly, the TED-LIUM [20] and TED-LIUM 3 [21] datasets contain a large amount of audio and transcripts from TED talks, which are also not traditional university lectures. Lee *et al.* [22] also have the problem of including untraditional tutorial-style lectures with no teacher–student interactions.

Collectively, these resources highlight many shortcomings with the current status quo of university-level educational corpora and the lack of accessible, rich, and interactive dialogue within traditional university lecture settings. This work aims to fill these gaps by curating LectBench-95. Table 1 summarizes the limitations of the aforementioned datasets for better readability.

3. DATASET CONSTRUCTION

In this section, a detailed overview of data collection, processing pipeline and the statistical planning that went into the dataset construction is provided. The criteria used for lecture selection will also be discussed.

3.1. Collection and Processing Pipeline

The videos were downloaded, and their audio was extracted using the yt-dlp[2] library. Which is then passed on to OpenAI's Whisper[3] [23] library. The choice of Whisper was guided by its robustness and low WER [24]. This pipeline yields a segmented transcript containing the text along with other useful metadata such as the start–end times of a segment, temperature, tokens, and segment IDs. However, only the text itself, the start- and end-times, and the segment IDs are needed. The segments are concatenated in the following format:

[Segment ID] [Start Time - End Time] Text Content

where:

- Segment ID: A zero-padded, three-digit identifier for

1 https://india-data.org/dataset-version/52e67bb5-5acd-438d-a13b-f255cee17432/5f9005e0-8a6b-4180-bc52-fdda80a6917a

2 https://github.com/yt-dlp/yt-dlp

3 https://github.com/openai/whisper — Specifically the small variant.

TABLE 1: Limitations of existing educational lecture datasets

Dataset	Limit. Subjects	Pre-Univ.	Access Restricted	No Transcripts	Non-Interactive	Lang. Limit.	Large Size
NCTE Transcripts	✓	✓					
MET Dataset		✓	✓	✓			
TalkMoves	✓	✓					
CIMA						✓	
Technical Indian English						✓	
Autoblog 2021	✓				✓		
Automatically Recognizing Lecture Highlighting Corpus			✓		✓		
VT-SSum					✓		
AVLectures	✓				✓		✓
Slide Speech					✓		✓
Khan Academy Corpus			✓		✓		
TED-LIUM					✓		
Lee et al.							

Limitations key:
1. Limited Subjects: Focuses on specific subjects (e.g., only math/STEM)
2. Pre-University: Primarily K-12 education level
3. Access Restricted: Requires subscription, affiliation, or unavailable
4. No Transcripts: Only audio/video available
5. Non-Interactive: Monologue without teacher–student interactions
6. Language Limitations: Focuses on specific languages or dialects
7. Large Size: Impractical for modest computing resources

the segment (e.g., 001, 002, etc.). It is used to uniquely identify each segment and to allow the LLM to reference it during the evaluation

- Start Time and End Time: The timestamps indicating the beginning and end of the segment, formatted as MM: SS.ss, where MM represents minutes, SS represents seconds, and ss represents fractional seconds rounded to two decimal places
- Text Content: The transcribed text of the segment, with leading and trailing whitespace removed.

Because interactivity was an afterthought for this dataset, and its value became apparent after going through the literature review, it was necessary to recheck all the videos to determine whether or not they are interactive. The traditional way would be to open each video and check if the lecturer is in a traditional classroom setting with students present, classifying it as interactive. Conversely, a MOOC-style video would be considered non-interactive. In some cases, however, the lecture may be in a Zoom-like setting where the lecturer is talking to a camera without any students present, but the lecturer may still be interactive with the audience. In these cases, we would still classify it as an interactive lecture[4]. Regardless, this method is cumbersome and time-consuming; therefore, we adopted a different approach. A Google Sheet was created with the video links, titles, broad subject, and specific subject, and a column was created to extract the video's link using the following cell formula:

=REGEXEXTRACT(B2, "v=([a-zA-Z0-9_-]+)")

Where B2 is the cell containing the video link. Afterward, the video ID is used to create 4 other cells that fetch the video's storyboard from YouTube servers; each cell fetches a different storyboard image using the following formula:

=IMAGE("https://i.ytimg.com/vi/" & D2 & "/0.jpg", 2)

Here, D2 is the cell containing the video ID, and the 0.jpg in the URL represents the storyboard's image index. The other 3 cells use 1.jpg, 2.jpg, and 3.jpg, respectively. Then, another column is created to hold a boolean value indicating whether the video is interactive or not, as judged by the storyboard images. If the storyboard images show a classroom setting with students present, it is marked as TRUE (i.e., interactive); otherwise, it is marked as FALSE (i.e., non-interactive). After that, Google Sheets' data-validation feature was used to create a Checkbox for that column, allowing toggling the interactivity status with a click instead of editing. This way, determining whether the video is interactive or not is done in a matter of seconds, without the need to open each video, waiting for it to load, and skipping to the relevant parts.

4 These videos make up a very small portion of the dataset, and were only included in order to meet the subjects quota. Preference was given to videos within physical classrooms

This approach is not perfect, but it is a good approximation that saves us a lot of time and effort. A screenshot of the Google Sheet is shown in Appendix Fig. 1. (Appendix A).

Now, for those videos that appear to be in a zoom-like setting, the storyboard images may not be enough to determine interactivity; manual checks were performed to determine their interactivity. If a video is non-interactive, it is simply discarded, and an interactive alternative on the same subject is sought.

Appendix B describes the attempts made at anonymizing the dataset. However, anonymization was not feasible given the current resources and time constraints we had. Furthermore, the lectures are all publicly available on YouTube, thus anonymization is not strictly necessary in this case. As a cautionary note, users are strongly encouraged to anonymize the dataset if their use case requires it.

3.2. Statistical Design

Each transcript is considered a single sample and is used in a contest between the different techniques. With 95 transcripts, each pairwise comparison between two techniques is performed across all 95 transcripts, providing 95 independent contests per pair. The following paragraphs explain how this number is determined and why it is sufficient to distinguish the performance superiority of a technique in an A/B test.

The sample size was not randomly chosen; the aim was to have a sample size that would be large enough to be statistically significant and small enough to be feasible to collect and process. Thus, a power analysis was used to determine the sample size needed to achieve 80% power with a significance level of 0.05, which corresponds to a 95% confidence level. Where the win-rate of technique A is $p_1 = 0.4$ and technique B is $p_2 = 0.6$, i.e., 20% effect size. p_1 and p_2 were chosen to cover the most demanding case in terms of sample size (more on this next). The main goal here is to ensure (with 95% confidence) that a 20% difference in the win rates of a pair of techniques will be detected if it exists. Larger effects (e.g. 30% or 40%, etc.) are inherently easier to detect as they require smaller sample sizes.

A two-proportion z-test is used as the basis to determine the sample size needed, the best fit for our scenario, as it assumes a binomial distribution of binary outcomes (i.e., win/lose scenario of an A/B experiment) and independent samples. Originally, the two-proportion z-test is used to determine the statistical significance of the difference between two proportions. However, by rearranging the formula[5], it can be used to determine the sample size needed to achieve a certain level of statistical significance. Equation 1 shows the formula used to calculate the sample size needed for each group.

Sample size calculation formula:

$$n = \frac{(Z_{\alpha/2} + Z_{\beta})^2 . (p_1(1-p_1) + p_2(1-p_2))}{(p_1 - p_2)^2} \quad (1)$$

Where:

- n is the sample size needed for each group
- $z_{\alpha/2}$ is the z-score corresponding to the significance level (0.05). Which is 1.96 for a two-tailed test
- z_{β} is the z-score corresponding to the power (0.80). Which is 0.84 for a power of 80%
- p_1 is the baseline proportion (e.g., win rate of technique A). Set to 0.4 in this case
- p_2 is the expected proportion (e.g., win rate of technique B). Set to 0.6 in this case.

The choice of a 20% effect size was not arbitrary; rather, it was made to strike the aforementioned balance between statistical significance and data collection feasibility derived, as shown in Fig. 1. The Figure illustrates the different effect sizes and baselines along with their corresponding sample sizes. As can be seen from the plot, the required sample size increases when the baselines are near 0.5, where it is the most difficult to distinguish which technique is superior, thus more samples are needed. In contrast, when the baseline is near 0 or 1, the required sample size is smaller as it is easier to detect a difference. Typically, the smallest effect size with the largest sample size is sought to ensure the most demanding case is covered. In this case, that would be the 10% effect size with a baseline of 0.5, which requires a sample size of $n = 385$. However, collecting 385 samples would be time-consuming and resource-intensive. Due to the impracticality of gathering such a big sample size, the 20% effect size with a baseline of 0.4, which requires a sample size of $n = 95$ will be chosen. This should give us a good balance between statistical significance and data collection feasibility.

3.3. Data Selection Criteria

To ensure the transcripts are relevant, high-quality, and diverse, the following criteria were put in place:

- The transcripts should be only traditional, professor-led

5 Assuming equal sample sizes for both proportions, i.e., $n_1 = n_2$ which is standard for experiment designs.

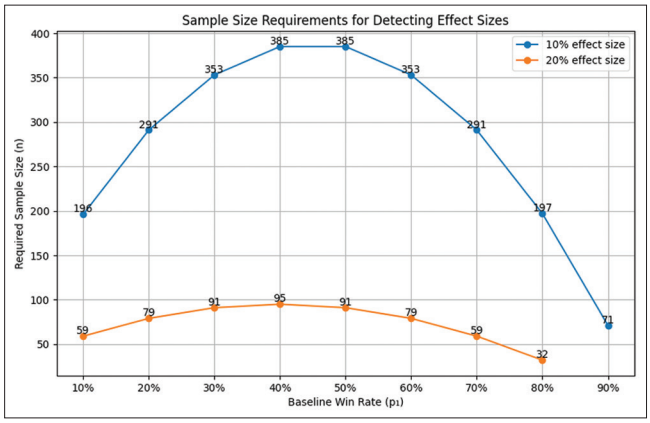


Fig. 1. Required sample size for different effect sizes and baseline win-rates ($p1$). Note that p2 = p1+effect size.

academic lectures with a clear presence of both students and teachers in the classroom–be it physical or virtual. No tutorials or non-academic content were included. This is to ensure that the content is relevant to the intended use cases

- The transcripts should be in English, as the majority of the academic content in the public domain is in English
- The transcripts range from 30 min all the way up to 2 h, which is the typical length of a university lecture
- The transcripts must have clear audio quality evidenced through a high signal-to-noise ratio (SNR) and Whisper's own segment-level average confidence scores (derived from avg_logprob). The thresholds are ≥25 dB for SNR and ≥70% for the average confidence level across all the segments. This is to ensure that mistranscription does not skew any results
- The transcripts should be from a diverse set of universities (public and private) and professors across different academic fields. To be more precise, a distribution of 40% STEM (38 transcripts), ≈ 30% Humanities (28 transcripts), ≈ 30% Social Sciences (29 transcripts) was targeted. Fig. 2 shows the full breakdown of the transcripts across different academic fields.

4. DATASET ANALYSIS

In this section, an overview of the LectBench-95 dataset will be provided, making clear its composition, characteristics, and key statistics.

4.1. General Stats

Table 2 provides a quick overview of the dataset, showing the key statistics such as the total number of lectures, total audio duration, and average lecture duration, among others.

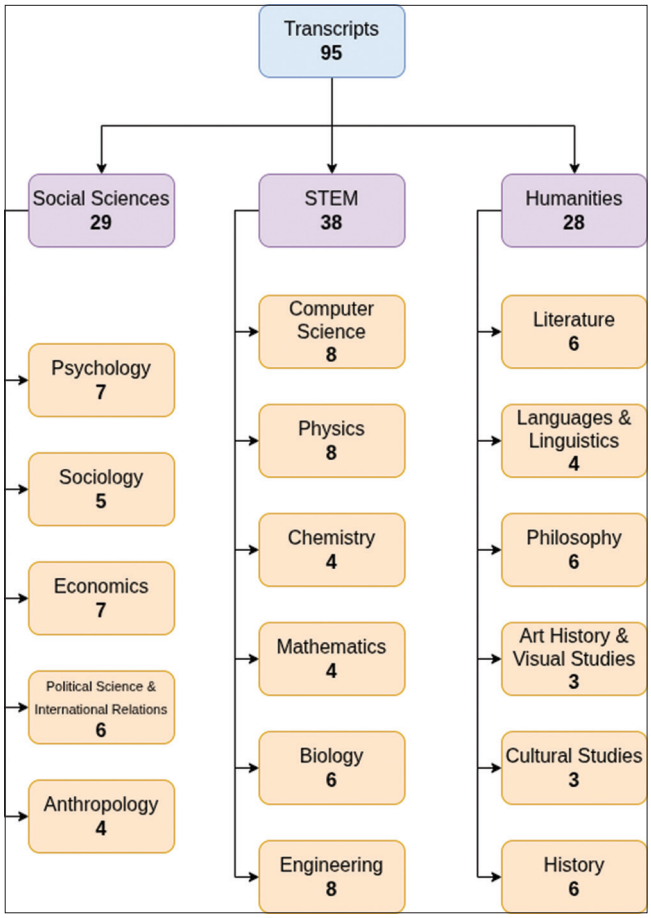


Fig. 2. Distribution of the transcripts across different academic fields.

Meanwhile, Table 3 provides a more detailed breakdown of the dataset's statistics, including temporal characteristics, speech characteristics, lexical diversity, and quality metrics.

The dataset contains 95 lecture transcripts, each with segments and metadata such as video publisher, title, URL, duration (in minutes), broad subject (i.e., discipline), and specific subject. With a total lecture audio duration of 94.06 h and an average lecture duration of 59.41 min (SD: 15.67), three lectures exceeded 90 min. The lectures span 3 disciplines and 17 subject areas, with each subject area having at least 3 lectures, as shown in Fig. 2. Furthermore, the dataset is lexically rich, with a total word count of 816,214 and a unique word count of 122,696. The Measure of Textual Lexical Diversity (MTLD) [25] is mainly used to assess lexical diversity and yields a mean score of 54.49 in this case. The flow of speech in the dataset is also noteworthy, with an average speech rate of 144.3 words per minute (WPM) and a standard deviation of 21.0 WPM, which aligns with typical speech rates in academic settings [26]. The transcripts were

generated using OpenAI's Whisper model, which provides an average log probability for each segment, from which the confidence scores are derived. The mean confidence score across all transcripts is 0.8402 (SD: 0.0508, Min: 0.7018), indicating a solid level of transcription quality. Segment-level timestamps are also available, allowing for any temporal analysis or alignment tasks.

These quick statistics should give insights into the dataset's scale and composition. The upcoming sections will dive even deeper into the dataset's statistics, including lexical, speech rate, and quality metrics.

TABLE 2: LectBench-95 dataset summary statistics

Metric	Value
General characteristics	
Total number of lectures	95
Total audio duration	94.06 h (5,644 min)
Average lecture duration	59.41 min (SD: 15.67)
Disciplines	3
Subject areas (all ≥3 lectures)	17
Lexical statistics	
Total word count	816,214
Unique word count	122,696
Vocabulary richness	MTLD: 54.49 (mean), TTR: 0.1578 (mean)
Speech rate	144.3 WPM (SD: 21.0)
Transcript quality	
Confidence score (Whisper)	Mean: 0.8402 (SD: 0.0508, Min: 0.7018)
Segment-level timestamps	Available

4.2. Statistical Breakdown and Quality Validation

In this section, the quality of the dataset will be validated by comparing the statistics against existing literature and benchmarks. This will tell us whether or not the dataset is reflective of real-world university lectures. The dataset validation is approached from three main angles: lexical statistics, speech rate statistics, and quality metrics. For each, a general overview and a subject-specific breakdown will be provided.

4.2.1. Lexical statistics

In terms of lexical statistics, the dataset shows a high degree of vocabulary richness, as evidenced in Fig. 4, which categorically illustrates the distribution of MTLD scores across lectures. As can be seen, around 97.9% of the lectures fall within the medium to high diversity categories; the majority being high diversity (64.2%), with only 2.1% classified as low-diversity.

Besides being an indicator of transcript quality, lexical diversity is also crucial for assessing the ability of lecture processing tools to handle the rich and varied vocabulary common in academia.

Turning our attention to subject-specific analysis, notable variations in lexical diversity are observed across different academic disciplines and specific subject areas within them.

TABLE 3: Full descriptive statistics of the LectBench-95 dataset

Metric	n	Mean	Standard	Min	Q1	Median	Q3	Max
Duration and temporal								
Duration (minutes)	95	59.41	15.75	33.05	47.81	54.82	72.20	104.57
Duration (hours)	95	0.99	0.26	0.55	0.80	0.91	1.20	1.74
Segment Count	95	803	366	299	553	700	955	2181
Avg segment duration (seconds)	95	5.05	2.04	2.13	3.53	4.42	6.07	13.02
Speech and linguistic rate								
Speech rate (WPM)	95	144.3	21.1	96.9	130.6	141.7	156.6	201.2
Syllable rate (SPM)	95	203.6	27.4	127.6	184.7	205.8	219.0	266.6
Word count	95	8591	2693	3965	6623	7960	10340	15585
Avg word length	95	4.19	0.25	3.66	4.03	4.18	4.34	4.95
Lexical diversity								
MTLD score	95	54.49	11.62	32.28	46.42	52.47	60.34	90.35
Type-Token ratio	95	0.158	0.043	0.074	0.127	0.153	0.187	0.282
Vocabulary size	95	1291	345	513	1086	1254	1484	2381
Hapax ratio	95	0.079	0.031	0.024	0.057	0.075	0.098	0.173
Sophistication ratio	95	0.165	0.033	0.098	0.146	0.163	0.183	0.282
Transcription quality								
Avg confidence	95	0.840	0.051	0.702	0.816	0.846	0.882	0.912
Low confidence ratio	95	0.060	0.100	0.00	0.008	0.025	0.058	0.467
Min confidence	95	0.552	0.163	0.004	0.448	0.580	0.647	0.830
Max confidence	95	0.924	0.043	0.748	0.913	0.937	0.951	0.970

Notes: *n*=Sample size, Q_1=First quartile, Q_3=Third quartile. WPM: Words per minute, SPM: Syllables per minute, MTLD: Measure of textual lexical diversity. Duration metrics measured across 95 educational video transcripts. Low confidence ratio indicates the proportion of transcript segments with confidence<0.7

Fig. 3 presents a ranking of subjects based on their mean MTLD scores, showing how diverse each subject's vocabulary is. For instance, art History and Visual Studies exhibit the highest mean MTLD score of 88.49 (±1.75), whereas Mathematics is the lowest with a mean MTLD score of 40.26 (±2.67), indicating a more specialized vocabulary. A trend that can easily be observed is that the more technical the subject, the lower the MTLD score. This trend is also backed by the discipline-based analysis conducted, in which Humanities and Social Sciences both have higher mean MTLD scores (58.5 and 58.3, respectively) compared to STEM subjects (48.6). This trend is consistent with existing literature that suggests PhD dissertations in the Humanities and Social Sciences tend to have higher lexical diversity compared to those in STEM fields, where more specialized and repetitive words are used [27].

Appendix Table 1 (Appendix C) provides a table with more statistics on the lexical diversity of the dataset.

4.2.2 Speech rate statistics

The dataset also exhibits substantial diversity when it comes to speech rate. Fig. 6 shows the distribution of lectures across different speech rate categories. As is evident, the majority of lectures (approximately 70%) fall within the Normal (120–160 WPM) range. In addition, a good portion of lectures fall in the slow (<120 WPM) and fast (>160 WPM) categories, with approximately 10% and 20%, respectively. This diversity in speech rate allows researchers both to evaluate the performance of lecture processing techniques on the most common speech rates (i.e., Normal) and to test their robustness on the tails of the distribution (i.e., Slow and Fast). This is crucial for a comprehensive evaluation of a certain technique.

As shown in Fig. 5, the variance of the average speech rate across different subjects can be noted. The subjects with the fastest speech rates are Computer Science (mean: 170.3 WPM, SD: 25.1), followed by Languages and Linguistics (mean: 161.2 WPM, SD: 12.3), whereas the slowest are Art History and Visual Studies (mean: 129.9 WPM, SD: 8.7) and Physics (mean: 134.0 WPM, SD: 22.9). It can be seen that the two fastest and two slowest subjects both fall in STEM and Humanities. This variation tells us that consistency cannot be detected in speech rate across disciplines, as the speech rate seems to be more dependent on the specific subject area. Other factors such as the lecturer's personal style, the complexity of the subject matter, and the intended audience may also play a significant role in determining speech rate. Looking at a subject such as Art History and Visual Studies, under the light of the previous lexical diversity analysis, wherein it was shown to have the richest vocabulary, coupled with the slowest speech rate, it can be concluded that this subject is more likely to be delivered in a more deliberate and measured manner, the lecturers are most likely very selective in their words and phrases, and might allow pauses for reflection, both of which which may contribute to the slower speech rate and higher lexical diversity. This might be in contrast to subjects such as Computer Science and Languages and Linguistics, which were ranked fifth and sixth, respectively, in terms of lexical diversity, but have the fastest speech rates. Mathematics and Physics, on the other hand, fall into the slow speech rate and low lexical diversity quarter. We point out such

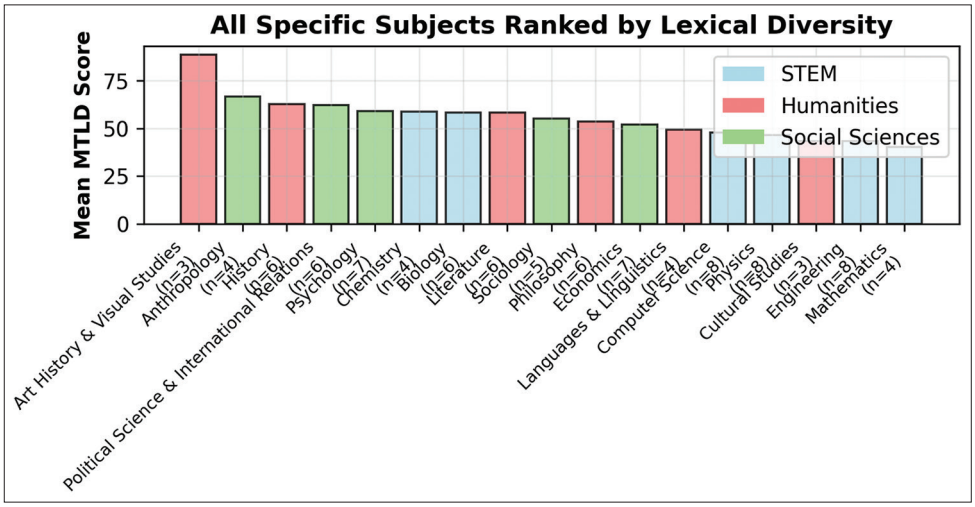


Fig. 3. Ranking of subjects by measure of textual lexical diversity scores.

analysis in the Intended Use Cases section (6.1) and leave it for future work.

Appendix Table 2 (Appendix D) contains further statistics.

4.2.3. ASR quality statistics

The distribution of average confidence scores across the dataset is shown in Fig. 8. This figure illustrates the spread of confidence scores, highlighting the overall quality of the transcriptions. The minimum confidence score is 0.7018, while the mean is 0.8402. It is evident that majority of the transcripts are clustered around the higher end of the confidence scale, indicating a generally high level of transcription accuracy.

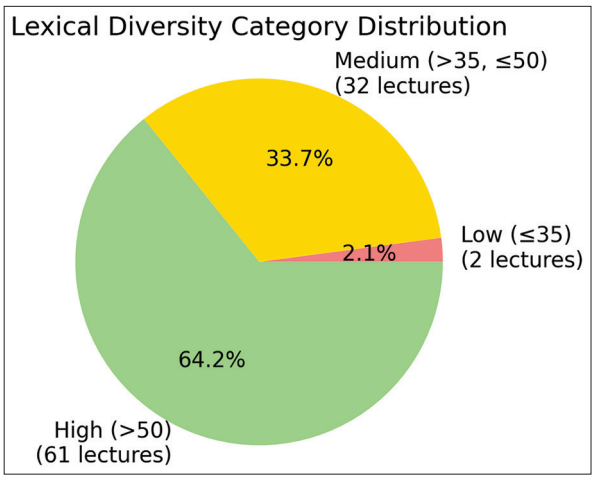


Fig. 4. Categorical breakdown of lectures by measure of textual lexical diversity scores.

As for the subject-specific average confidence scores, Fig. 7 shows the average confidence scores for each subject area. It is evident that most subjects have high average confidence scores, with most falling between 0.8 and 0.9. This indicates that the transcript reliably reflects the spoken content. Furthermore, it supports confidence in evaluations or analyses that will be conducted on the dataset.

Further statistics are available in Appendix Table 3 (Appendix E).

5. TOY A/B DEMONSTRATION

To validate the effectiveness of the dataset in providing statistically significant results on head-to-head comparisons between the techniques, a toy A/B evaluation is conducted using two different LLM models on the task of zero-shot summarization of the transcripts. The two models used were Google's Gemini-1.5-Flash and Gemini-1.5-Flash-8B. This choice was guided by Google's claim that, despite the Gemini-1.5-Flash-8B model being smaller than the Gemini-1.5-Flash model, it is more efficient, faster, and nearly as capable as the larger model[6]. Because of the inherent performance similarity of the two models, distinguishing the superior model is harder, making this an appropriate test of the dataset's ability to detect small performance differences. Fig. 9 illustrates the prompt used for the zero-shot summarization task.

BERTScore [28] was used to evaluate the faithfulness of the

6 https://developers.googleblog.com/en/gemini-15-flash-8b-is-now-generally-available-for-use/

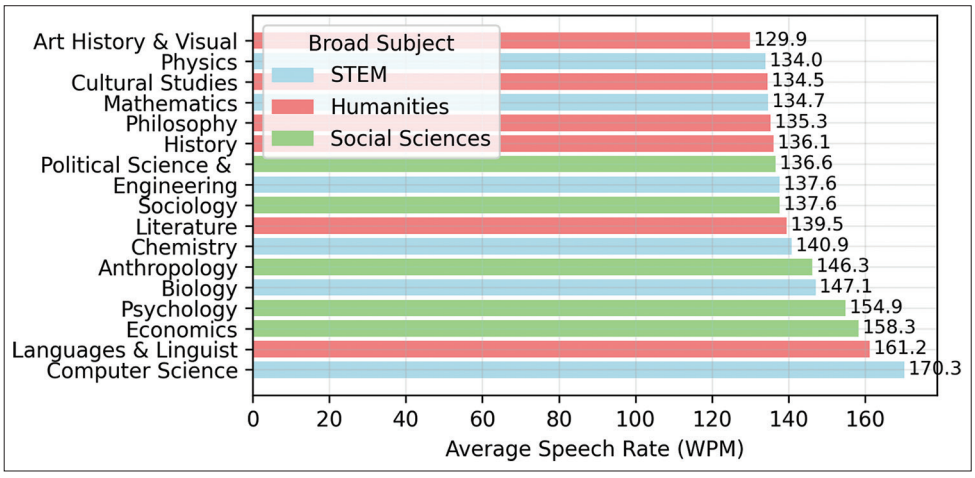


Fig. 5. Average speech rate by subject.

summaries generated by each of the two models.

The results of the A/B evaluation are shown in Fig. 10. As can be seen, the Gemini-1.5-Flash model outperformed the Gemini-1.5-Flash-8B model in terms of the number of wins, with a win-rate of 71.6% (68 wins out of 95 comparisons). The Gemini-1.5-Flash-8B model had a win-rate of 28.4% (27 wins out of 95 comparisons). An effect size of 43.16% was observed between the two models, with 95% confidence intervals for the win-rates being (0.6251, 0.8065) and (0.1935, 0.3749), respectively. The two-tailed binomial test was used to determine the statistical significance of the results with a null hypothesis of the win-rate being 50% (i.e., no difference between the two models). The *P*-value was found to be 0.00003114, which is statistically significant at the 0.05 level. Therefore, the null hypothesis can be rejected, and we conclude that the two models differ significantly in terms of zero-shot summarization task performance.

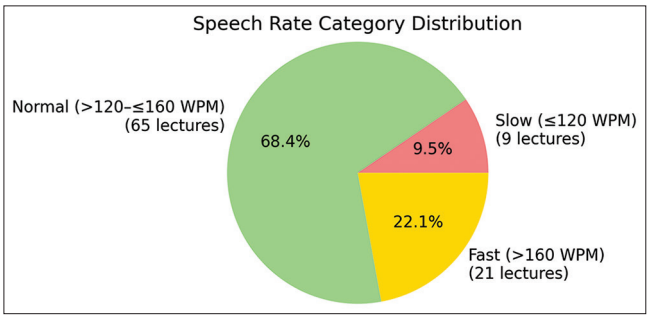


Fig. 6. Categorical breakdown of lectures by speech rate.

This all demonstrates that the dataset is capable of detecting statistically significant differences in performance between different techniques. In this case, far beyond the 20% effect size aimed for. The narrow confidence intervals further confirm the statistical significance of the results.

While BERTScore gave us F1 scores that could be used in a t-test to compare the magnitude of differences in average faithfulness, but instead the win-rate was favored as it looks at the consistency of a model's superiority over the other. We believe that the consistent superiority of a tool across lectures demonstrates more practical educational value than marginal average gains.

6. DISCUSSION

Here, the dataset's practical use cases and limitations are discussed, each of which provides valuable insights into avenues for independent research and future work.

6.1. Intended Use Cases

6.1.1. Evaluation of lecture processing techniques

The primary intended use case for this dataset is the evaluation of different lecture processing techniques in A/B settings. The dataset's diversity should allow for robust comparisons (e.g., testing technique performance across STEM vs. humanities lectures) while revealing the blind spots a technique might have in academic fields. Lecture processing techniques can include summarization, question generation, feedback generation, automated grading, and more.

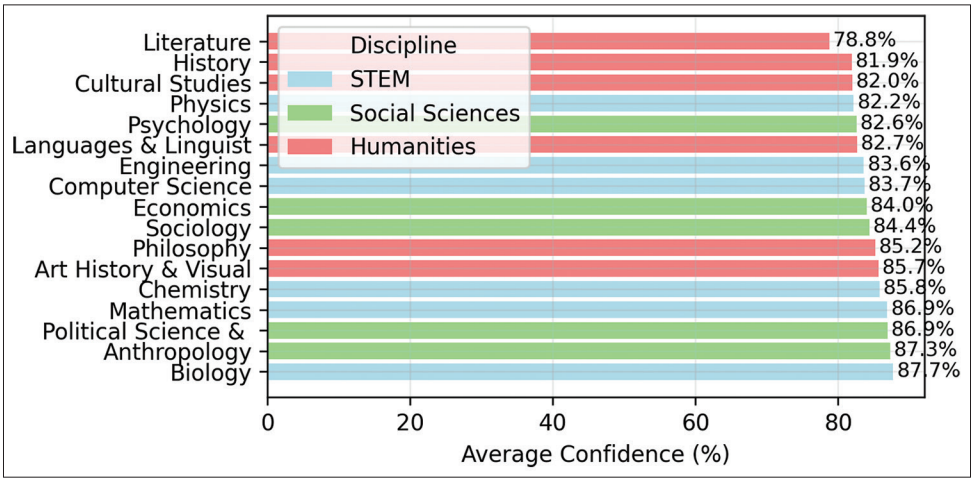


Fig. 7. Average confidence scores by subject.

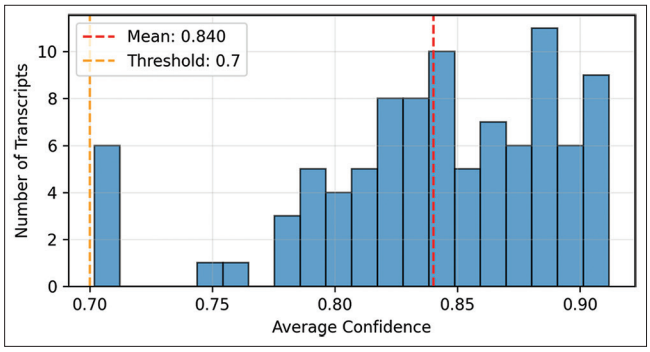


Fig. 8. Distribution of average confidence scores.

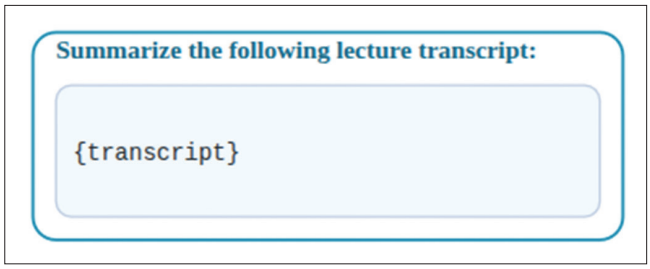


Fig. 9. Prompt template for lecture transcript summarization.

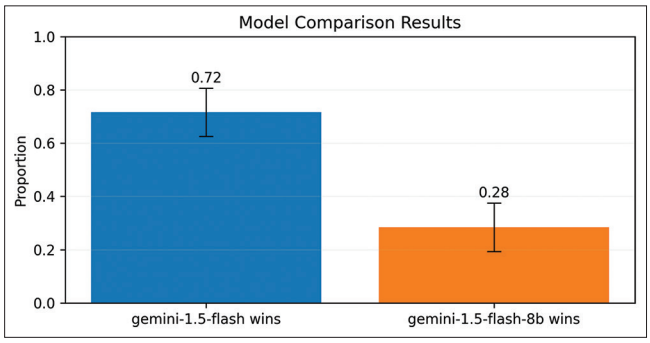


Fig. 10. Results of the toy A/B evaluation between the two LLM models on the task of zero-shot summarization of the transcripts.

6.1.2. Rapid and cheap proof-of-concepts

The dataset can be used for rapid prototyping and proof-of-concept implementations of new lecture processing techniques. Its modest yet statistically significant sample size gives it room for cheap experimentation and testing, especially in the emerging field of LLMs where API costs can add up quickly. All without the hassle of collecting and processing new data.

6.1.3. LLM post-training

The dataset can also be used with some LLM post-training techniques, such as fine-tuning, direct preference optimization, knowledge distillation, and prompt engineering research, to alter the behavior of the LLMs toward a certain task. For example, fine-tuning an LLM to generate realistic transcripts for classroom simulations or using segments of the transcript for few-shot prompts.

6.1.4. Research on educational dialogue analysis

The dataset can also be used for research on educational discourse analysis, such as analyzing the dynamics of teacher-student interactions, including questioning, prompting, and feedback. This can aid in understanding and improving instructional practices in higher education.

6.2. Dataset Limitations

6.2.1. Dataset size

While the current sample size of 95 transcripts is sufficient for most A/B evaluations with a decent effect size, it may not be enough for more nuanced analyses. Future work could benefit from a larger dataset (e.g., $n = 385$ for 10% effect size or even larger) to capture more subtle differences in performance, especially when comparing techniques with similar performance levels. All while ensuring diverse representation across different academic fields and universities. This would allow for more robust statistical analyses and generalizations.

6.2.2. Language and demographic diversity

The dataset currently only focuses on English lectures, primarily from Ivy League universities in North America, which may lead to applications with less generalizability to non-English speaking contexts or different demographics. Future work could benefit from expanding the dataset to include lectures in other languages and from a wider range of universities.

6.2.3. Detailed metadata

The dataset currently lacks detailed metadata about the lectures, such as the course name, professor's background, student demographics, and related attributes. This information could provide valuable context for any lecture processing techniques applied to the transcripts. For example, including the course level (e.g., undergraduate, graduate) would help a lecture rating model understand the expected complexity and depth of the content.

6.2.4. University bias

The dataset contains lectures from Ivy League universities such as Harvard and MIT, which may not always be representative of other universities or institutions. This may introduce a bias in the results of the evaluations. However, testing the techniques against one another will always yield the superior technique regardless of the university bias.

While the limitations might create obstacles for certain research undertakings, they reflect limitations either in terms of time or resources on our part, and deliberate choices to prioritize the A/B testing capabilities. Nonetheless, the dataset should provide a robust foundation for core lecture processing research, with identified gaps that can be addressed in future work by us or the community.

7. CONCLUSIONS AND FUTURE WORK

We present LectBench-95, a dataset of 95 university lecture transcripts from 3 different academic disciplines (STEM, Humanities, and Social Sciences) and 17 different subjects, with a total of 94 hours of transcribed lecture content. The dataset is curated with a focus on diversity (i.e., in the academic subjects, vocabulary used, and lecturer speech rates), quality, and student-lecturer interactivity as demonstrated. In addition, effort has been put into selecting a sample size that would always yield statistically significant results with a solid effect size. It is primarily intended for the evaluation and comparison of different lecture processing techniques (educational natural language processing) in A/B settings, as well as for rapid prototyping and proof-of-concept implementations. The dataset is released under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) license, and users are encouraged to respect the license and the original content creators. As for future work, any of the gaps identified in the limitations subsection (6.2) suggest complementary datasets that can be built on top of LectBench-95.

REFERENCES

[1] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan … & D. Amodei. Language models are few-shot learners. In: "*Advances in Neural Information Processing Systems*". Curran Associates, Inc., United States, pp. 1877-1901, 2020.

[2] A. I. Open, J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. S. Altman, S. Anadkat, R. Avila, I. Babuschkin …& S. Balaji. "*GPT-4 Technical Report*". Cornell University, United States, 2023.

[3] Y. Hicke, A. Agarwal, Q. Ma and P. Denny. "*AI-TA: Towards an Intelligent Question-Answer Teaching Assistant using Open-Source LLMs*". Cornell University, United States, 2023.

[4] B. Oh, Y. Lee and Y. Kim. Applicability of pretrained language models: Automatic Screening for children's language development level. In: L. Biester, D. Demszky, Z. Jin, M. Sachan, J. Tetreault, S. Wilson, L. Xiao and J. Zhao, editors. "*Proceedings of the Second Workshop on NLP for Positive Impact (NLP4PI)*". Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Hybrid), pp. 149-156. 2022.

[5] A. Xenakis, I. Dimos, M. Feidakis, D. Sotiropoulos, K. Kalovrektis and G. Nikolaou. An LLM-based smart repository platform to support educators with computational thinking, AI, and STEM Activities. In: "*Empowering STEM Educators With Digital Tools*". IGI Global Scientific Publishing, United States, pp. 107-136. 2025.

[6] C. Cohn, N. Hutchins, T. Le and G. Biswas. "A chain-of-thought prompting approach with LLMs for evaluating students' formative assessment responses in science". *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 21, pp. 23182-23190, 2024.

[7] D. Demszky and H. Hill. "*The NCTE Transcripts: A Dataset of Elementary Math Classroom Transcripts*". Association for Computational Linguistics, USA, 2023.

[8] T. J. Kane, D. F. McCaffrey, T. Miller and D. O. Staiger. "*Have we Identified Effective Teachers? Validating Measures of Effective Teaching using Random Assignment*". Bill and Melinda Gates Foundation, MET Project, [Research Paper], 2013.

[9] A. Suresh, J. Jacobs, C. Harty, M. Perkoff, J. H. Martin and T. Sumner. "*The TalkMoves Dataset: K-12 Mathematics Lesson Transcripts Annotated for Teacher and Student Discursive Moves*". Cornell University, United States, 2022.

[10] K. Stasaski, K. Kao and M. A. Hearst. CIMA: A large open access dialogue dataset for tutoring. In: J. Burstein, E. Kochmar, C. Leacock, N. Madnani, I. Pilán, H. Yannakoudakis, and T. Zesch, editors. "*Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications*". Association for Computational Linguistics, Seattle, WA, USA, pp. 52-64, 2020.

[11] A. K. Rai, S. D. Jaiswal and A. Mukherjee. "A deep dive into the disparity of word error rates across thousands of NPTEL MOOC videos". *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 18, pp. 1302-1314, 2024.

[12] A. Hernandez and S. Yang. "*Multimodal Corpus Analysis of Autoblog 2020: Lecture Videos in Machine Learning*". Springer, Berlin, pp. 262-270. 2021.

[13] A. Hernandez, P. Klumpp, B. Das, A. Maier and S. H. Yang. Autoblog 2021: The importance of language models for spontaneous lecture speech. In: "*Text, Speech, and Dialogue: 25th International Conference, TSD 2022, Brno, Czech Republic, September 6-9, 2022, Proceedings*". Springer-Verlag, Berlin, Heidelberg, pp. 291-300, 2022.

[14] M. Song, I. Aslan, E. Parada-Cabaleiro, Z. Yang, E. André, Y. Yamamoto and B. Schuller. Lecture Video Highlights Detection from Speech. In: "*2024 32nd European Signal Processing Conference (EUSIPCO)*". pp. 361-365. 2024.

[15] T. Lv, L. Cui, M. Vasilijevic and F. Wei. "*VT-SSum: A Benchmark Dataset for Video Transcript Segmentation and Summarization*". arXiv:2106.05606.

[16] J. Wright, M. Liberman, N. Ryant and J. Fiumara. "*Evaluating Speech-to-Text Systems with PennSound*". Cornell University, United States, 2025.

[17] D. S. Singh, A. Gupta, C. V. Jawahar and M. Tapaswi. Unsupervised Audio-Visual Lecture Segmentation. In: "*2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*". pp. 5221-5230, 2023.

[18] H. Wang, F. Yu, X. Shi, Y. Wang, S. Zhang and M. Li. "*SlideSpeech: A Large-Scale Slide-Enriched Audio-Visual Corpus*". Cornell University, United States, 2023.

[19] D. Ďurišková, D. Jurášová, M. Žilinec, E. Šubert and O. Bojar. Khan Academy Corpus: A Multilingual Corpus of Khan Academy Lectures. In: "N. Calzolari, M. Y. Kan, V. Hoste, A. Lenci, S. Sakti

and N. Xue, editors. "*Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*". ELRA and ICCL, Torino, Italia, pp. 9743-9752, 2024.

[20] A. Rousseau, P. Deléglise and Y. Estève. TED-LIUM: An Automatic Speech Recognition Dedicated Corpus. In: Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk and S. Piperidis, editors. "*Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, N". European Language Resources Association (ELRA), Istanbul, Turkey, 2012, pp. 125-129.

[21] F. Hernandez, V. Nguyen, S. Ghannay, N. Tomashenko and Y. Estève. "TED-LIUM 3: Twice as much data and corpus repartition for experiments on speaker adaptation. In: A. Karpov, O. Jokisch and R. Potapova, editors. "*Speech and Computer*". Springer International Publishing, Cham, pp. 198-208, 2018.

[22] D. W. Lee, C. Ahuja, P. P. Liang, S. Natu and L. P. Morency. "Lecture presentations multimodal dataset: Towards understanding multimodality in educational videos. In: "*2023 IEEE/CVF International Conference on Computer Vision (ICCV)*". IEEE, United States, pp. 20030-20041, 2023.

[23] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey and I. Sutskever. "*Robust Speech Recognition Via Large-Scale Weak Supervision*". Cornell University, United States, 2022.

[24] C. Graham and N. Roll. "Evaluating openAI's whisper ASR: Performance analysis across diverse accents and speaker traits". *JASA Express Letters*, vol. 4, no. 2, p. 025206, 2024.

[25] P. M. McCarthy and S. Jarvis. "MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment". *Behavior Research Methods*, vol. 42, no. 2, pp. 381-392, 2010.

[26] P. Wingrove. "How suitable are TED talks for academic listening?" *Journal of English for Academic Purposes*, vol. 30, pp. 79-95, 2017,

[27] W. Xiao and S. Sun. "Dynamic lexical features of PhD theses across disciplines: A text mining approach". *Journal of Quantitative Linguistics*, vol. 27, no. 2, pp. 114-133, 2020.

[28] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger and Y. Artzi. "*BERTScore: Evaluating Text Generation with BERT*". Cornell University, United States, 2019.

[29] M. Laouenan, P. Bhargava, J. B. Eyméoud, O. Gergaud, G. Plique and E. Wasmer. "A cross-verified database of notable people, 3500BC-2018AD". *Scientific Data*, vol. 9, no. 1, p. 290, 2022.

APPENDIX

A. Lexical Diversity Statistics

We report a more detailed lexical diversity descriptive statistics in Appendix Table 1. Including MTLD, TTR, and vocabulary statistics.

B. Speech and syllable rate statistics

We provide speech-rate (WPM) and syllable-rate (SPM) distributions in Appendix Table 2.

C. Transcript confidence statistics

Descriptive statistics for the Confidence scores (from Whisper's average token-level probabilities) are provided in Appendix Table 3.

D. Google Sheet Screenshot

Appendix Fig. 1. shows the Google sheet used to determine lecture interactivity via the storyboard images. As can be seen, 4 thumbnails are retrieved, each of which showcase different time points of the video lecture.

E. Anonymization and ethical considerations

Attempts were made to anonymize each segment by removing any Personally Identifiable Information (PII). The plan was to anonymize the following:

- Names: Replaced with the term "Person X" (e.g., "James" becomes "Person 1").
- Organizations: Replaced with the term "[REDACTED ORG]."
- Emails: Replaced with the term "[REDACTED EMAIL]."

This process was intended to ensure the privacy of any individuals mentioned in the transcripts. The Name and Organization replacements were done through Named Entity Recognition, specifically using the spaCy library[1]. The email replacement was done using a simple regex pattern that matches the standard email format.

The anonymization process, however, caused historical figures' names like "Henry Ford" to be replaced with "Person X." This is a problem as it removes crucial context from the transcripts. The magnitude of the impact this may have on the results is still not understood. To solve this problem, cross-checking with historical databases of notable people like the one by [29] was attempted, but was not feasible as sometimes only the last names were used when referencing such people (e.g., "James Madison" might be referred to by "Madison"), which could be common first names of other people. An attempt was also made to use the MediaWiki API to cross-check the names and used a scoring approach to only single out very historical and very popular people of a certain type but this method proved problematic as well. Another approach considered was to use the most popular baby names list of the United States for the last 5 years and use them as a roster list to anonymize the names, but to no avail. Thus, informed by the fact that all the lectures are public-domain YouTube videos, it was concluded that it may not be strictly necessary to anonymize the transcripts and therefore the idea of anonymization was abandoned altogether. Users are still encouraged to anonymize the dataset if their use case requires it.

1 https://spacy.io/—Specifically the en_core_web_lg model.

APPENDIX TABLE 1: Summary of measure of textual lexical diversity (MTLD), TTR, and vocabulary statistics

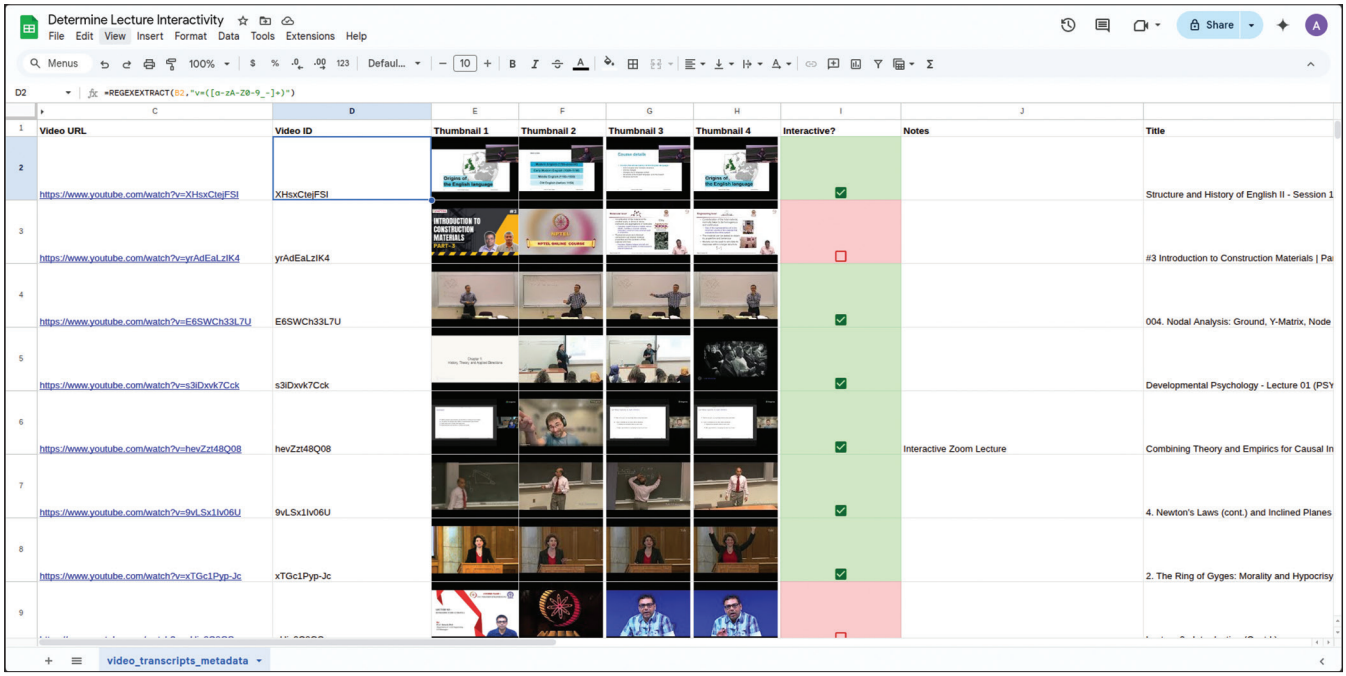
MTLD (Lexical diversity)	
Mean MTLD	54.49
Median MTLD	52.47
Standard deviation	11.56
Min MTLD	32.28
Max MTLD	90.35
25th percentile	46.42
75th percentile	60.34
Type-token ratio (TTR)	
Mean TTR	0.1578
Median TTR	0.1530
Range	0.0741–0.2822
Vocabulary statistics	
Average vocabulary size	1292 unique words
Vocabulary range	513–2381 unique words
Average hapax ratio	0.079
Average sophistication ratio	0.165

APPENDIX TABLE 2: Speech and syllable rate statistics

Speech rate (Words per minute [WPM])	
Mean speech rate	144.3 WPM
Median speech rate	141.7 WPM
Standard deviation	21.0 WPM
Min speech rate	96.9 WPM
Max speech rate	201.2 WPM
25th percentile	130.6 WPM
75th percentile	156.6 WPM
Syllable rate (Syllables per minute [SPM])	
Mean syllable rate	203.6 SPM
Median syllable rate	205.8 SPM
Range	127.6–266.6 SPM

APPENDIX TABLE 3: Transcript confidence statistics

Total segments	
Total segments across all transcripts	76,329
Overall confidence statistics	
Mean average confidence	0.8402
Median average confidence	0.8461
Standard deviation	0.0508
Min average confidence	0.7018
Max average confidence	0.9118
Segment-level statistics	
Mean segment confidence	0.8338
Median segment confidence	0.8508
Segments below 0.7	4,808 (6.30%)
Segments below 0.5	142 (0.19%)



Appendix Fig. 1. Screenshot of the Google Sheet interface for assessing video interactivity via storyboard images.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Hybrid U-Net Architectures with ResNet50 and VGG19 for Accurate CT-Based Kidney Disease and Stone Segmentation

Dlshad Abdalrahman Mahmood, Muhammad Amin Muhammadali

Department of Computer Science, Faculty of Science, Soran University, Kurdistan Region, Iraq

ABSTRACT

Kidney illness is a major worldwide health issue requiring prompt and precise diagnosis for optimal management. This paper presents a comprehensive evaluation of hybrid deep learning (DL) architectures that integrate U-Net with ResNet50 and VGG19 for the automatic segmentation of kidney stones and renal disorders from computed tomography (CT) images. We assembled a dataset of 118 individuals from a private hospital, comprising 13,035 kidney-specific CT scans, while also using the publicly accessible Kaggle Kidney Stone Segmentation Dataset. Three experimental situations were established: (1) Concurrent segmentation of kidney disease and stones, (2) segmentation of kidney stones alone, and (3) segmentation of kidney disease exclusively. The hybrid U-Net + ResNet50 model attained superior performance in stone-only segmentation, with an F1-score of 0.8653, an IoU of 0.7626, and an accuracy of 0.9998 at a resolution of 256 × 256. The U-Net + VGG19 model exhibited strong performance in all situations, attaining an F1-score and DC of 0.8663 for stone segmentation. Both models demonstrated exceptional generalization ability when evaluated on external datasets. The findings indicate that hybrid architectures markedly improve segmentation accuracy compared to conventional methods, providing dependable automated tools for clinical kidney pathology evaluation while ensuring computational efficiency with average processing durations below 0.05 s per scan.

Index Terms: Kidney stone, Renal disorders, DL, U-Net, VGG19, ResNet50.

1. INTRODUCTION

Kidney disease is a significant public health challenge to society today. Early diagnosis and awareness significantly reduce mortality. Disorders that interfere with the kidneys' typical function are referred to as kidney diseases [1]. Kidney disease classification requires dividing kidney images or patient data into multiple illness sorts, including chronic kidney disease (CKD) stages, polycystic kidney disease, nephrolithiasis, and renal tumors. A kidney stone is a solid formation that can result in damage to the kidneys, intense pain, and reduced quality of life due to urinary system blockages [2].

Numerous imaging techniques are employed in medical diagnosis, including sonography, computed tomography (CT), magnetic resonance imaging (MRI), and X-rays; however, these methods are not without issues, as they can be time-consuming, subjective, and susceptible to human error. The increase in incidence, coupled with technological advancements, imposes a substantial financial strain on healthcare facilities for managing kidney stone disease (KSD), with an estimated global expenditure of USD 5.3 billion in 2014, rendering it the second most expensive condition [3].

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp231-250

E-ISSN: 2521-4217

P-ISSN: 2521-4209

Copyright © 2025 Mahmood and Muhammadali. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Dr. Muhammad Amin Muhammadali, Computer Science Department, Faculty of Science, Soran University, Kurdistan Region, Iraq. E-mail: muhammad.amin@soran.edu.iq

Received: 13-06-2025 Accepted: 21-09-2025 Published: 13-10-2025

The literature on modern-day machine learning techniques in healthcare progressively focuses on their use in renal disease, particularly in the study of renal pathologies, which are common and affect a significant percentage of the population, leading to various complications, such as death in some instances [4]. Deep learning (DL)-based image segmentation is now firmly recognized as an effective method in the field of image segmentation. It has been widely used to delineate homologous areas as primary and fundamental elements of the diagnostic and therapeutic process [5].

The automated detection of renal abnormalities represents a critical objective in clinical practice, with medical imaging modalities including ultrasound (US), MRI, and CT serving as primary diagnostic tools [6]. The integration of whole slide images from histological samples in digital pathology algorithms for computer-aided assessment has gained significant momentum in recent years [7]. Beyond detection, CT imaging enables precise determination of stone dimensions and anatomical positioning, thereby facilitating comprehensive risk assessment for spontaneous stone passage and informing decisions regarding surgical intervention [8]. Images CT consistently provide the most accurate diagnosis. Conventionally, US has lower sensitivity and specificity than CT [9]. The difficulties of interpreting the challenges of complex image data in medical images can be addressed by applying ML and DL concepts [10]. Clinicians can therefore use DL approaches to automatically diagnose renal disorders. However, improving performance in the identification and challenge of renal disease remains difficult [11], [12].

An important element of artificial intelligence (AI) techniques is the training of appropriate images available to the public. Researchers can now use various datasets, including CT, US, and MRI. However, each dataset differs from the others in terms of quantity, illumination variation, dimensions, and image quality, which may require fine-tuning. That is why authors Kaur and Singh believe that image fusion plays a significant role in different computer vision applications. However, designing an efficient image fusion technique is still a challenging task [13]. The objective of the research is to augment diagnostic precision, alleviate the strain for radiologists, and boost the early identification of renal disorders using automated image segmentation. The following are major contributions mentioned:

1. Novel Hybrid Architecture Development U-Net with ResNet50 and U-Net with VGG19
2. Comprehensive Multi-Scenario Evaluation Framework
3. Clinical Dataset Creation and Expert Annotation
4. Outstanding Performance Achievement
5. Clinical Translation and Practical Implementation.

Although developments, including Transformer-based designs like Swin-Transformers, have produced cutting-edge outcomes in medical picture segmentation, their clinical utility is limited by substantial computing demands and the necessity for exceptionally large annotated datasets. These restrictions make it hard to use them in everyday clinical practice, especially in hospital settings where resources are limited. Hybrid U-Net models, on the other hand, are a better choice because they keep the original U-Net decoder's speed and reliability while adding powerful encoders such as ResNet50 and VGG19 to make feature extraction better. This hybrid design immediately solves the gap between accuracy and efficiency, giving models that are not only competitive with more sophisticated approaches but also practicable for real-world clinical integration.

The remainder of this paper is organized as follows: Section 2 presents a comprehensive review of related work in kidney pathology segmentation and DL applications in medical imaging. Section 3 details the methodology, including research design, data collection, preprocessing procedures, proposed hybrid U-Net architectures, model training protocols, and evaluation metrics. Section 4 presents the experimental results and provides an in-depth discussion of the performance evaluation across three distinct scenarios, comparing U-Net+ResNet50 and U-Net+VGG19 models under various configurations. Finally, Section 5 concludes the study with key findings, contributions, and future research directions for advancing automated kidney disease diagnosis through DL techniques.

2. RELATED WORK

The authors, Huang *et al.*, created a computer-aided diagnostic system for Kidney Ureter Bladder (KUB) imaging to help physicians correctly diagnose urinary tract stones [14]. Whereas Yildirim *et al.*, proposed an automatic detection system for kidney stones (stone or no stone) that uses coronary imaging CT, using DL techniques [15]. Moreover, Fitri *et al.*, using a convolutional neural network (CNN), have created an autonomous method for classifying urinary stones into the three categories based on micro-CT images [16]. Furthermore, Zhao *et al.*, introduced a multi-scale supervised 3D U-Net (MSS U-Net) for the segmentation of kidneys and renal tumors from CT scans [17].

However, researchers Alqahtani *et al.*, say sigmoid functions enhance prediction accuracy for binary outcomes. Finally, they apply classification using the proposed modified Extreme Gradient Boosting (XGBoost) for kidney stone prediction. The loss functions are modified to enhance the model's learning effectiveness and classification accuracy. Evaluate the proposed approach through internal comparison with the decision tree (DT) and Naive Bayes (NB) [18].

and Yang *et al.*, performed a retrospective analysis of the medical records of 358 patients who received shock wave lithotripsy for urinary stones (kidney and upper urinary tract stones), which includes patient demographic characteristics and urinary stone characteristics as depicted by non-contrast CT images. They used an 80% training set and a 20% test set to predict success, primarily using decision tree-based ML methods, including Random Forest (RF), XGBoost, and Light Gradient Boosting Method (LightGBM) [19].

However, Daniel *et al.* used computer learning, specifically 2D CNN, to accurately separate left and right kidneys from T2-weighted (MRI) data. The data set consisted of 30 HC volunteers and 30 CKD patients. The model was trained on 50 manually outlined HC and CKD kidney sections. The model was further evaluated using 50 test data sets, consisting of data from 5 healthy controls and 5 patients with CKD, each scanned 5 times in a session to facilitate comparison between the microscopic CNN and manual segmentation of the kidney [20].

Ma *et al.* enrolled 468 patients with kidney, bladder, and urinary stones at multiple sites at Peking Union Medical College Hospital. Urine metabolite profiling was used to discover markers for KSD using ML techniques. The total number of patients with renal stones was 148 (34.02%), bladder stones 34 (7.82%), and multisite stones 163 (34.83%). According to their analysis, the RF algorithm had the best prediction accuracy, with area under the curve (AUC) values for kidney stones of 0.809, urethral stones of 0.99, and multisite stones of 0.775 [21]. Aksakallı *et al.*, assessed multiple ML techniques, including DT, RF, support vector classifier, multilayer perceptron, K-nearest neighbors, NB (Bernoulli NB), and deep neural network utilizing CNN. The collection comprises 221 kidney X-ray images acquired from the Department of Urology at Atatürk University. The trials indicate that the DT yields the most favorable classification results. This method achieves the greatest F1 score, with a success rate of 85.3%, utilizing the S+U sampling technique [22].

Researchers, Fitri *et al.*, have devised an automated technique to categorize urinary stones into three categories utilizing micro-CT images through CNN. A total of 2,430 images were obtained from *in vitro* micro-CT scans of urinary stones in various patients. The validation accuracy of the devised method utilizing a CNN with optimized hyperparameters was 0.9852. The trained CNN algorithm attained a test accuracy of 0.9959 [16]. Furthermore, Parakh *et al.* used unenhanced CT scans of the abdomen and pelvis in 535 adults with suspected KSD. CNN's cascading model has a high accuracy AUC of 0.954 in detecting urinary tract stones on unenhanced CT scans [12].

Nonetheless, Elton *et al.* utilized a dataset of 91 CT colonography (CTC) images, including manually annotated kidney stones, alongside 89 CTC scans devoid of kidney stones. 50% of the data were allocated for training, while the remaining 50% was designated for testing. As an external validation set, 6185 patients' CTC scans from a separate institution were employed. A three-dimensional U-Net model was employed for kidney segmentation. A 13-layer CNN classifier was used to differentiate kidney stones from false-positive regions. The system attained an area under the receiver operating characteristic curve of 0.95 on an external validation set, with an AUC of 0.95, sensitivity of 0.88, and specificity of 0.91 at the Youden index [23].

Furthermore, Blau *et al.*, present a fully automated approach for renal cyst diagnosis, underpinned by a strong segmentation of the kidneys executed by a fully CNN. The evaluation of performance was conducted on 52 randomly selected abdominal CT scans from a genuine radiological process, which included more than 70 cysts annotated by a proficient radiologist. The program identified 59 out of 70 cysts (true-positive rate = 84.3%) while generating an average of 1.6 false positives per case [24].

Nonetheless, the authors Xiong *et al.* posited that ultrasonography is extensively utilized in the diagnosis of kidney tumors due to its widespread acceptance, affordability, and absence of radiation exposure. Consequently, they introduced a novel technique for segmenting renal tumors in US images utilizing the adaptive subregional diffusion level set model (ASSLSM). In comparison to conventional US segmentation techniques, ASLSM demonstrates superior accuracy in renal tumor segmentation. The test yielded a Hausdorff distance (HD) of (8.75 ± 4.21) mm, a mean absolute distance of (3.26 ± 1.69) mm, and a dice index (Dice) of 0.93 ± 0.03 [25].

To create their study sample, Gaikar *et al.*, examined the Pathology and Image Archiving and Communication System (PACS) database. Produced a set of MP-MRI scan data for individuals with kidney masses. T1W-NG image 3D volumes for 108 patients and T2W, T1W-IP, T1W-OP, T1W-PRE, and T1W-CM image 3D volumes for 50 patients made up the dataset. They also created a method based on TL to enhance kidney segmentation on the dataset. Two stages were taken to apply the created kidney segmentation algorithm to various mp-MRI data. Using a DL-based attention U-Net model, the kidney segmentation was first identified on T1W-NG images. In the subsequent phase, the pretrained T1W-NG kidney segmentation model was fine-tuned to distinguish kidneys in T2W, T1W-IP, T1W-OP, T1W-PRE, and T1W-CM MRI sequences. Increased average DSC T1W-IP from 83.64% to 85.42%, T1W-OP from 79.35% to 83.66%, T1W-PRE from 82.05% to 85.94%, T1W-CM from 85.65% to 87.64%, and T2W climbed from 87.19% to 89.90% as a result of the TL technique [26]. Table 1 summarizes the findings of published investigations. CT-based models, such as RDA-UNET, exhibit enhanced performance, whereas DeepLab, ResNet50, and Swin Transformer thrive across many modalities and tasks.

3. METHODOLOGY

The standard imaging modalities for assessing nephrolithiasis and renal pathology include US [32], CT [33], MRI [34], and KUB X-ray imaging [35]. This study used medical CT scan data to segment kidney stones and kidney disease after sequencing and image preparation procedures. Furthermore, the DL models used for training and testing are discussed.

3.1. Research Design

In the first step, a dataset of the kidney 3D images is collected. After collecting the data, the DL models are used to detect the kidney images. The results of the segmentation of kidney stones and kidney disease of the models are compared with several other methods based on the standard performance criteria. Thereafter, the optimal model is trained and evaluated using a publicly accessible dataset from Kaggle, which is explained diagrammatically in Fig. 1.

3.2. Data Collection and Preprocessing

The first objective was to obtain realistic data on kidney disease. It was achieved to collect CT scan data. CT scans were obtained from 118 patients at a private hospital in Ranya, consisting of subjects: 34 normal healthy, 29 with kidney stones only, 23 with kidney disease, and 32 with kidney stones and kidney disease, as shown in Table 2. Also, utilizing the publicly available Kidney Stone Segmentation Dataset from Kaggle.

The DICOMDIR of the 118 patients was then converted into images, using software (diVision Lite) to Joint Photographic Group (JPG), which contained 49,463 images. We isolated only 13,035 images dedicated to kidneys under the supervision of a nephrologist and urologist. Of these, there are 1131 CT scans of kidney stones and 1584 CT scans of kidney disease. Its data sets are shown in Table 3.

A preprocessing strategy was devised to prepare the dataset for efficient model training, encompassing numerous essential phases. Initially, all images were scaled to a standardized resolution of 512 × 512 pixels to maintain uniformity throughout the collection. All images were subsequently converted from 8-bit to 24-bit format to preserve greater color and detail information. Noise was mitigated by Gaussian filtering, which smooths images by averaging pixel values within the near vicinity. Furthermore, brightness was augmented to improve the visibility of essential features with linear brightness adjustment. The processed images were sorted into a separate folder for improved accessibility and efficient use in later modeling steps.

A specialized open-source Python application was utilized for annotating medical CT scans. The labeling process was executed through labelImg.py, a commonly employed annotation tool incorporated within the Anaconda environment. CT scans illustrating nephropathy (Fig. 2) and kidney stones (Fig. 3) were meticulously examined and documented under the direct supervision of a qualified nephrologist to guarantee clinical precision. Every image was carefully annotated, and the labels were stored in three distinct formats: JavaScript Object Notation (JSON), TXT, and Extensible Markup Language (XML). For the purpose of creating precise segmentation masks unique to each annotated image, the chosen formats were meant to improve compatibility with later processes. In medical image analysis, the manual way ensured better annotations, which are necessary for supervised learning models to work.

The JSON file serves as the primary reference for producing the appropriate segmentation mask. Automation of this procedure was achieved with Python scripts, which guarantee uniform and accurate extraction of mask regions. Fig. 4 shows an example of this procedure.

TABLE 1: Summary of published investigation findings comparing model performance across different approaches

Study	Method	Objective	challenge limitations	Result
Fu *et al.* [27]	RDA-UNET model	Automatic segmentation of renal cysts	Selecting the best model for renal cyst segmentation in medical image CT	Left kidney DSC=96.25 Precision=96.34 Recall=96.88 Right kidney DSC=94.22 Precision=95.34 Recall=94.61
Goyal *et al.* [28]	R-CNN	Automatic segmentation of the kidneys	The automatic segmentation of the kidneys in MRI	Dice=0.890 IoU=0.816
Sharma *et al.* [29]	InceptionV3	To analyze images of renal pathologies, including stones, tumors, and cysts.	For optimal performance in diagnosing kidney disorders, including cysts, tumors, and stones, by CT imaging.	Accuracy=97.76 Recall of=98.28 F1 score=0.95
Göker [30]	ResNet101	Classifying kidney stone illnesses	Compare transfer learning models from CT images	Overall accuracy=0.981 Recall=0.991 Specificity=0.974 Precision=0.966 F1-measure=0.978 Kappa statistic=0.962
Islam *et al.* [31]	Swin VGG16 ResNet50 Inceptionv3	Renal cyst, tumor, and stone detection.	To achieve the best model in the diagnosis of renal diseases, such as cysts and tumors, using CT imaging.	Swin Accuracy=99.30% VGG16 Accuracy=98.20% ResNet50 Accuracy=61.60% Inceptionv3 Accuracy=73.80%

MRI: Magnetic resonance imaging, CT: Computed tomography

TABLE 2: Comprehensive distribution of patients by renal health status

Patients total	Healthy	Only has kidney stones	Only has kidney disease	Kidney stones and kidney disease
118	34	29	23	32

TABLE 3: The distribution of CT scans by kidney condition

Total CT scans	CT scans of the kidneys	CT scans of kidney stones	CT scans of kidney disease	CT scans contain kidney stones and kidney disease
49463	13035	1131	1584	308

CT: Computed tomography

From the collected dataset, three distinct datasets were ultimately constructed, each consisting of CT scans and their corresponding segmentation masks. The first dataset included images and masks for both kidney disease and kidney stones. The second dataset focused exclusively on kidney stone cases, while the third contained only instances related to renal illness. All datasets were randomly split into three subsets: 70% for training, 20% for validation, and 10% for testing. The entire process of dataset construction and splitting was automated using Python to ensure consistency and efficiency. The datasets are available at this link in Kaggle: https://kaggle.com/datasets/b65b0abc924a6aa91f894ac91839c6cb52894b4092767770c5dd34e7eb09d210

Furthermore, the study utilized the publicly accessible Kidney Stone Segmentation Dataset from Kaggle (https://www.kaggle.com/datasets/bemorekgg/kidney-stone-segmentation-dataset) to train and assess the efficacy of the suggested segmentation algorithm. This dataset includes CT scans of kidney stones together with related segmentation masks developed by the Segment Anything Model (SAM) utilizing YOLO bounding box annotations. The CT images are supplied in Joint Photographic Experts Group (JPEG) format, while the segmentation masks are presented as binary Portable Network Graphics (PNG) images. The dataset consists of 923 image-mask pairings, randomly assigned to training (70%, 646 images + 218 rotated images), validation (20%, 184 images), and testing (10%, 93 images). Using the same model and evaluation subsets, Fig. 5 shows examples of Cagle images and masks.

3.3. Proposed Methods

DL is achieving success and garnering interest across various disciplines, including computer vision, speech recognition,

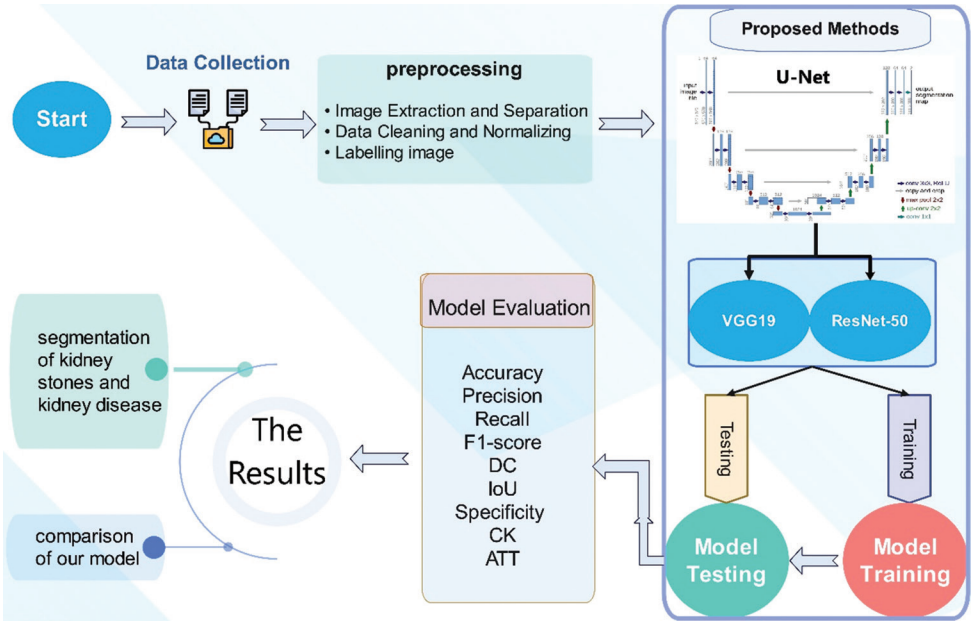


Fig. 1. The proposed structure.

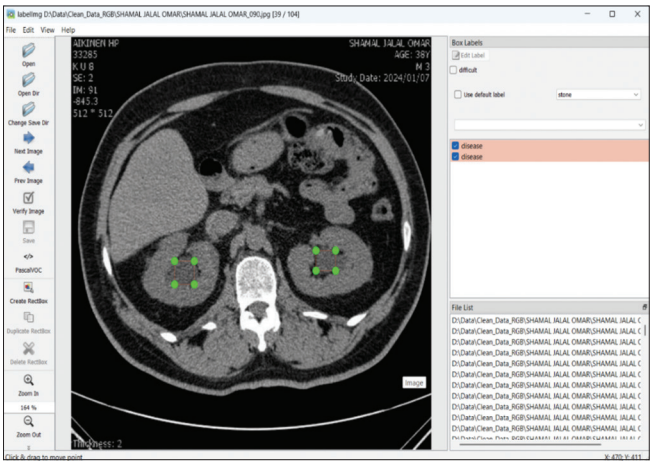


Fig. 2. Labelling kidney disease.

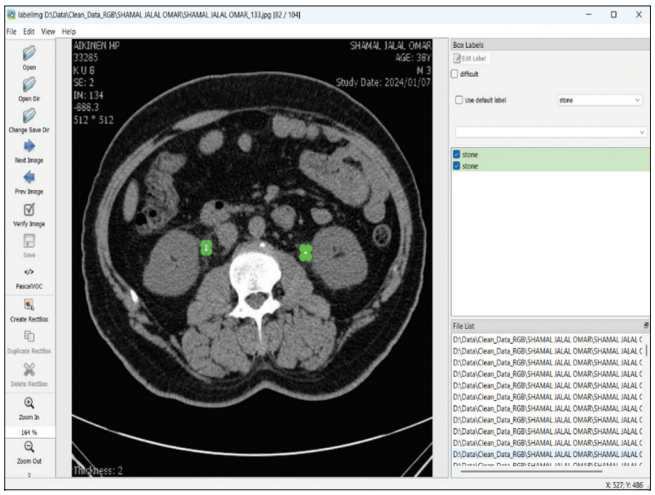


Fig. 3. Labelling a kidney stone.

natural language processing, and gaming. DL techniques generate a correspondence between raw inputs and target outputs (e.g., image classifications) [36]. The use of DL-based medical image analysis in computer-aided detection (CAD) offers decision support to clinicians, enhancing the accuracy and efficiency of diagnostic and treatment processes while stimulating new research and development initiatives in CAD [37].

3.3.1. U-Net model

U-Net, as shown in Fig. 6, is a CNN architecture designed for image segmentation tasks, introduced by Olaf Ronneberger, Philipp Fischer, and Thomas Brox in 2015. The network derives its name from its U-shaped architecture, comprising a contracting path (encoder) and an expanded path (decoder) [38].

The U-Net is a CNN designed for biological image segmentation. The architecture, featuring an encoder–decoder framework, exhibits remarkable stability and can achieve accurate segmentation with a reduced number of training images. The network consists of three 3 × 3 convolutional layers, with a maximum of two 2 × 2 layers following each pooling layer, utilizing the ReLU activation function. A 1 × 1 convolutional layer is appended at

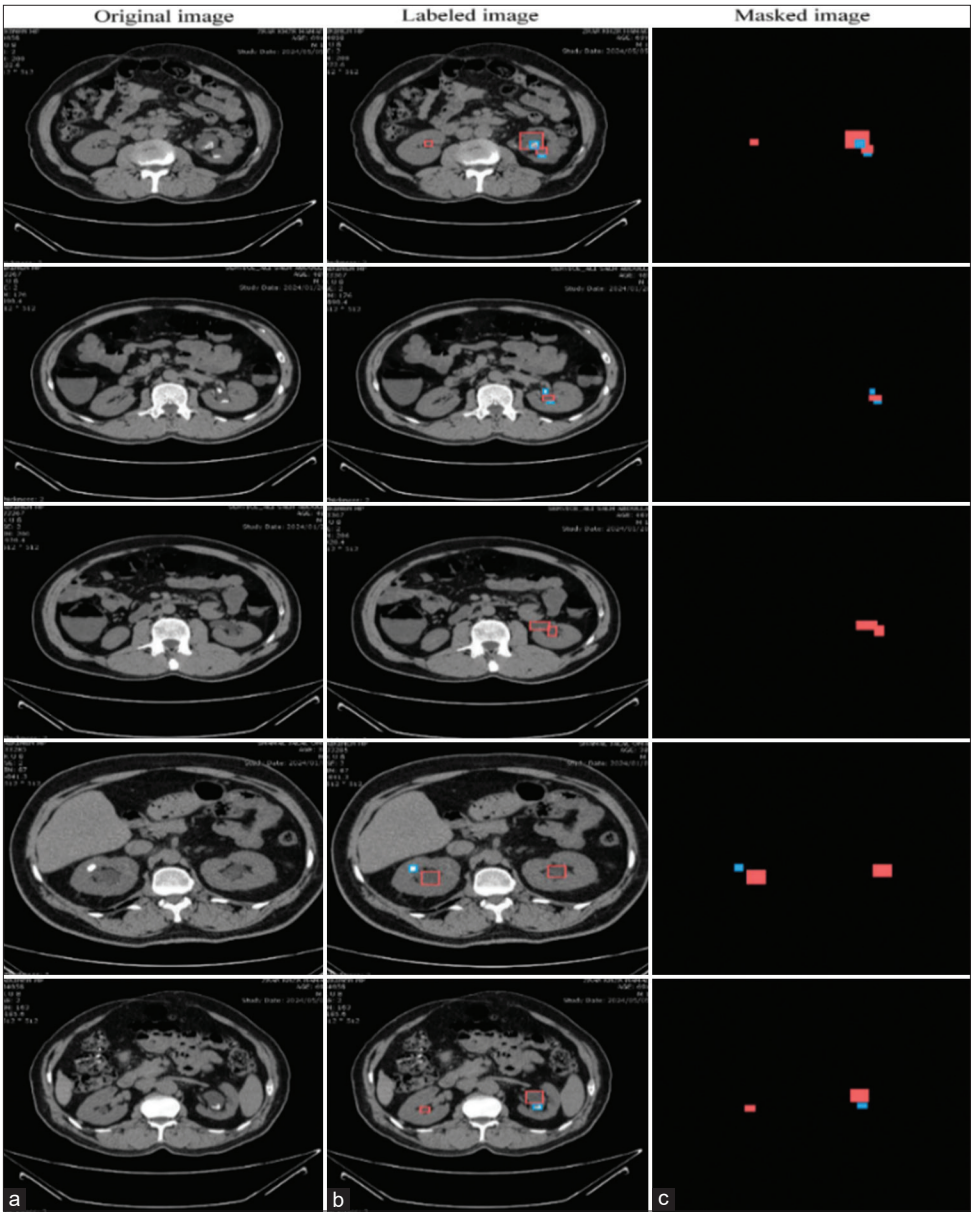


Fig. 4. (a) Some original images; (b) labels the images; (c) creates masks from the JSON file; the red color represents kidney disease, and the blue color represents kidney stones.

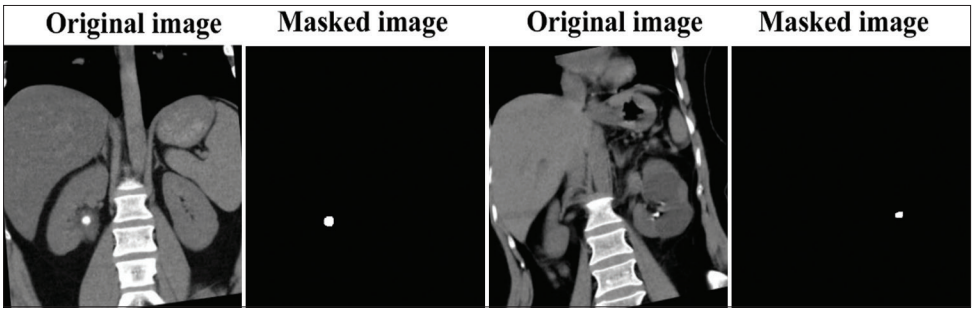


Fig. 5. A sample dataset on kidney stones and masks is available on Kaggle.

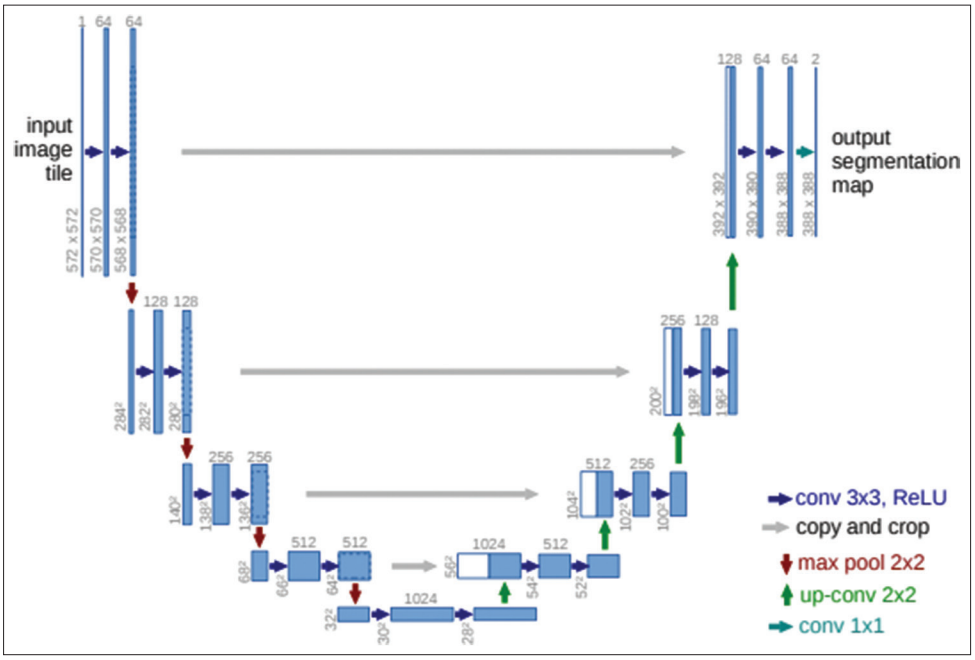


Fig. 6. U-Net design (example for 32 × 32 pixels at the lowest resolution) [39].

the conclusion. The U-Net comprises a contracting path and a symmetric expanding path, which are utilized for accurate localization and feature extraction. The U-Net design relies significantly on data augmentation methods. The network is capable of segmenting a 512 × 512 image on a contemporary GPU in under one second. The U-Net has been effectively utilized for medical image segmentation; nevertheless, 3D convolution is recommended to fully leverage the spatial information in 3D images, including CT and MRI scans [39].

In the domain of binary classification, the binary cross-entropy loss function is one of the most widely used metrics to evaluate the performance of a model. It measures the dissimilarity between the predicted probabilities and the actual binary labels (0 or 1), penalizing incorrect predictions more heavily as they diverge from the ground truth. Mathematically, the binary cross-entropy loss is defined as:

$$\text{Loss} = \frac{1}{N}\sum_{i=1}^{N}\left[y_i \log\left(\hat{y}_i\right) + \left(1 - y_i\right)\log\left(1 - \hat{y}_i\right)\right] \quad (1)$$

where y_i represents the actual label and $\hat{y}_i$ signifies the expected probability for the i_th sample. This function guarantees that the loss converges to zero with accurate forecasts and escalates markedly with inaccuracies. Moreover, it is essential in gradient-based optimization by offering explicit and comprehensible feedback throughout model training.

3.3.2. Comparison of hybrid U-Net methods

Hybrid U-Net methodologies augment the conventional U-Net design by substituting the encoder with superior DL models, such as ResNet50 and VGG19. These adjustments aim to enhance feature extraction and increase the model's accuracy, especially in intricate tasks such as medical image segmentation. By retaining the original U-Net decoder, these hybrid models sustain spatial detail reconstruction while leveraging enhanced feature representations. Comparative assessments generally quantify performance through metrics such as segmentation pixel-wise accuracy, precision, recall, Dice coefficient (DC), and Intersection over Union (IoU), with findings frequently indicating that hybrid models surpass the original U-Net in both precision and generalization. In this study, advanced segmentation frameworks were used by combining VGG19 and ResNet50 with U-Net to accurately segment kidney stones and related kidney diseases from medical images. This approach enhances the automated analysis of kidney conditions, supporting more effective diagnosis and treatment planning.

3.4. Model Training and Evaluation

To thoroughly assess the efficacy of the hybrid models utilized on the three specified scenarios, an extensive array of performance metrics was applied, including accuracy, precision, recall, F1-score, DC, IoU, specificity, Cohen's kappa (CK), and average test time (ATT). The selection of these measures was guided by their ability to evaluate several dimensions of model

performance, particularly in medical image segmentation. Accuracy provides a broad assessment of overall correctness by juxtaposing the total number of accurate predictions against all predictions produced; precision denotes the ratio of accurately predicted positives to the total predicted positives, hence regulating the false-positive rate, whereas recall measures the ratio of correctly detected actual positives, emphasizing the model's sensitivity. Their harmonic mean, the F1-score, offers a balanced metric, especially advantageous in cases of uneven class distribution. The DC and IoU function as metrics for spatial overlap between predicted and ground truth segmentation masks, essential for assessing model performance at the pixel level. Specificity enhances recall by evaluating the model's capacity to accurately detect negative instances, hence minimizing false positives. CK provides a chance-corrected metric for assessing agreement between forecasts and actual outcomes, particularly significant in multi-class tasks. The ATT serves as a pragmatic metric for the model's computational efficiency, underscoring its viability for real-world applications where processing speed is critical. These measures collectively establish a comprehensive framework for evaluating the segmentation accuracy, reliability, and efficiency of hybrid models in complicated, clinically relevant datasets.

3.5. Ethical Considerations

This research followed ethical guidelines to make sure that data and technology were used in a responsible way. To protect patient privacy, all CT scan data were made anonymous. To avoid bias and overfitting, a diverse dataset was employed. The AI models were created just for research and are not meant to be used in clinical settings without more testing. The whole process was guided by ethical principles to make sure that it was fair, safe, and respectful of human dignity.

4. RESULTS AND DISCUSSION

This section delineates the experimental setup, preprocessing section, training procedure, and assessment of the suggested DL models. The emphasis is on evaluating several hybrid U-Net topologies to determine the most efficient model for the particular segmentation problem. Various models, U-Net with ResNet50 and VGG19, were trained and evaluated utilizing the prepared datasets. According to the evaluation standards described in Section 3.4, the models were assessed using various performance metrics.

4.1. Experimental Setup Recap

This study evaluated the performance of DL models for kidney CT scan segmentation using U-Net + ResNet50 and U-Net + VGG19 architectures. Experiments were conducted in a consistent environment with an NVIDIA GTX 1080 Ti GPU, an Intel Core i7 CPU, and 16 GB RAM, using Python 3, TensorFlow 2.x, Keras, and OpenCV on Windows 10 through Anaconda and Jupyter Notebook.

4.2. Model Performance Evaluation

The models were assessed utilizing various performance indicators. All models were trained and validated under uniform experimental settings to guarantee an equitable comparison. The experiments were conducted using different image resolutions (128 × 128 and 256 × 256), batch sizes (8 and 24), and epoch settings (100, 150, and 200) with a learning rate of 0.0001 and the Adam optimizer, and a 5-fold cross-validation (CV) strategy.

This section presents a quantitative assessment of each DL model employed for segmenting CT scans associated with kidney stones and renal illnesses. The models were assessed and trained over three distinct scenarios within the dataset to gauge their segmentation efficacy. The scenarios were as follows: In the first scenario, the data set consisted of CT scans with their respective segment masks for kidney disease with kidney stones. The dataset was split into 70% for training (1675 scans), 20% for validation (479 scans), and 10% for testing (240 scans). In the second scenario, the dataset consisted exclusively of renal stone-related CT scans and masks. We used data augmentation to enhance the training set. The training set consisted of 1582 CT scans (70%), the validation set included 226 scans (20%), and the test set included 114 scans (10%). The third scenario focused the dataset exclusively on CT scans and masks associated with renal disease. The data set consisted of 1108 scans from the training set (70%), 317 from the validation set (20%), and 159 from the test set (10%). Table 4 provides a summary of all three scenarios along with their corresponding distributions. Samples were randomly assigned to each category to guarantee unbiased assessment.

The study thoroughly assessed two hybrid models, U-Net + ResNet50 and U-Net + VGG19, for the multiclass segmentation of renal disease and kidney stones, utilizing CT scans. To ensure equitable comparisons and minimize residual learning effects, each model was trained from inception using varying image resolutions (256 × 256 and 128 × 128), batch sizes (8 and 24), and training epochs (100, 150, and 200), as shown in Table 5. The U-Net + ResNet50 used deep feature extraction to make boundaries more accurate, whereas the U-Net + VGG19 used its deep hierarchical structure to keep spatial features. There were three tests for each model:

TABLE 4: Splitting of the training, validation, and testing in three diagnostic scenarios

n	Data sets	Train (70%)	Validation (20%)	Test (10%)	Total of images
1	CT scans and masks related to kidney disease and kidney stones.	1675	479	240	2394
2	Only CT scans and kidney stone-related masks.	791×2 rotation	226	114	1922
3	Only CT scans and masks for kidney disease.	1108	317	159	1584

CT: Computed tomography

TABLE 5: Experimental configuration parameters for U-Net-based model architectures

Models	Image size	Batch size	Epochs
U-Net+ResNet50	• 256×256 • 128×128	• 8 • 24	• 100 • 150 • 200
U-Net+VGG19	• 256×256 • 128×128	• 8 • 24	• 100 • 150 • 200

the first scenario was segmentation of kidney disease and kidney stones, the second scenario was only kidney stone segmentation, and the third scenario was only kidney disease segmentation. Figs. 7 and 8 show the input images, ground truth masks, and anticipated outputs for all configurations. This methodical approach yielded significant insights into the advantages and drawbacks of each hybrid architecture across various training scenarios.

Figs. 7 and 8 display segmentation outcomes produced by two models: U-Net combined with ResNet50 and U-Net combined with VGG19. Each image comprises the actual CT scans, ground truth masks, and anticipated segmentation masks. Subfigures (a-c) present findings for a picture dimension of 256 × 256 with a batch size of 8 at 100, 150, and 200 epochs, respectively. Subfigures (d-f) present findings for an image dimension of 128 × 128 with a batch size of 8, whereas subfigures (g-i) depict outcomes for the identical image dimension with a batch size of 24. These visualizations illustrate the models' efficacy across various training configurations and segmentation contexts.

4.3. Discussion

The evaluation criteria were used across two different U-Net-based models, hybrid models comprising U-Net with each model, ResNet50 and VGG19, for three separate scenarios. This multi-scenario methodology provides a detailed examination of how model performance fluctuates based on task complexity and data characteristics.

4.3.1. Hybrid U-Net + ResNet50

The U-Net+ResNet50 model was evaluated across multiple configurations of batch size, image dimensions, and training epochs. Initial testing with 256 × 256 images and batch size 8 showed optimal performance at 150 epochs (accuracy: 0.9992, precision: 0.8514, recall: 0.786, F1-score: 0.8174), with performance declining at 200 epochs due to overfitting. Reducing image size to 128 × 128 maintained robust segmentation performance while significantly decreasing inference time from 0.0275s to 0.0149s. The best overall results were achieved using 128 × 128 images, a batch size of 24, and 200 epochs, yielding an F1-score of 0.8238 and an IoU of 0.7000 with an inference time of 0.0157s. This setup achieved an optimal balance between segmentation accuracy and computational efficiency, demonstrating that the model works well with less computing power by optimizing parameters in Scenario 1.

In the second scenario, the hybrid U-Net + ResNet50 model was evaluated exclusively on CT kidney stone images with consistent accuracy (0.9998) and specificity (0.9999) across all configurations. The model was tested across three parameter combinations: 256 × 256 image resolution with a batch size of 8, 128 × 128 resolution with a batch size of 8, and 128 × 128 resolution with a batch size of 24, each trained for 100, 150, and 200 epochs. Optimal performance was achieved using 256 × 256 images, a batch size of 8, and 200 epochs, yielding a precision of 0.879, a recall of 0.8521, an F1-score/DC of 0.8653, an IoU of 0.7626, CK of 0.8652, and an average training time of 0.0259 s. Reducing image dimensions to 128 × 128 generally decreased performance metrics, while increasing batch size from 8 to 24 showed mixed results with slightly improved precision but reduced recall.

For the third scenario, focusing on renal disease segmentation using the hybrid U-Net + ResNet50 model, experiments were conducted with varying training epochs (100, 150, 200) and image resolutions (256 × 256 and 128 × 128) at different batch sizes. At 256 × 256 resolution with a batch size of 8, optimal performance was achieved after 100 epochs, yielding an accuracy of 0.9987, a precision of 0.8142, a recall of 0.7951, an F1-score of 0.8046, and an IoU of 0.673. Extended training to 150 and 200 epochs showed gradual performance degradation across most metrics. When image resolution was reduced to 128 × 128, similar patterns emerged

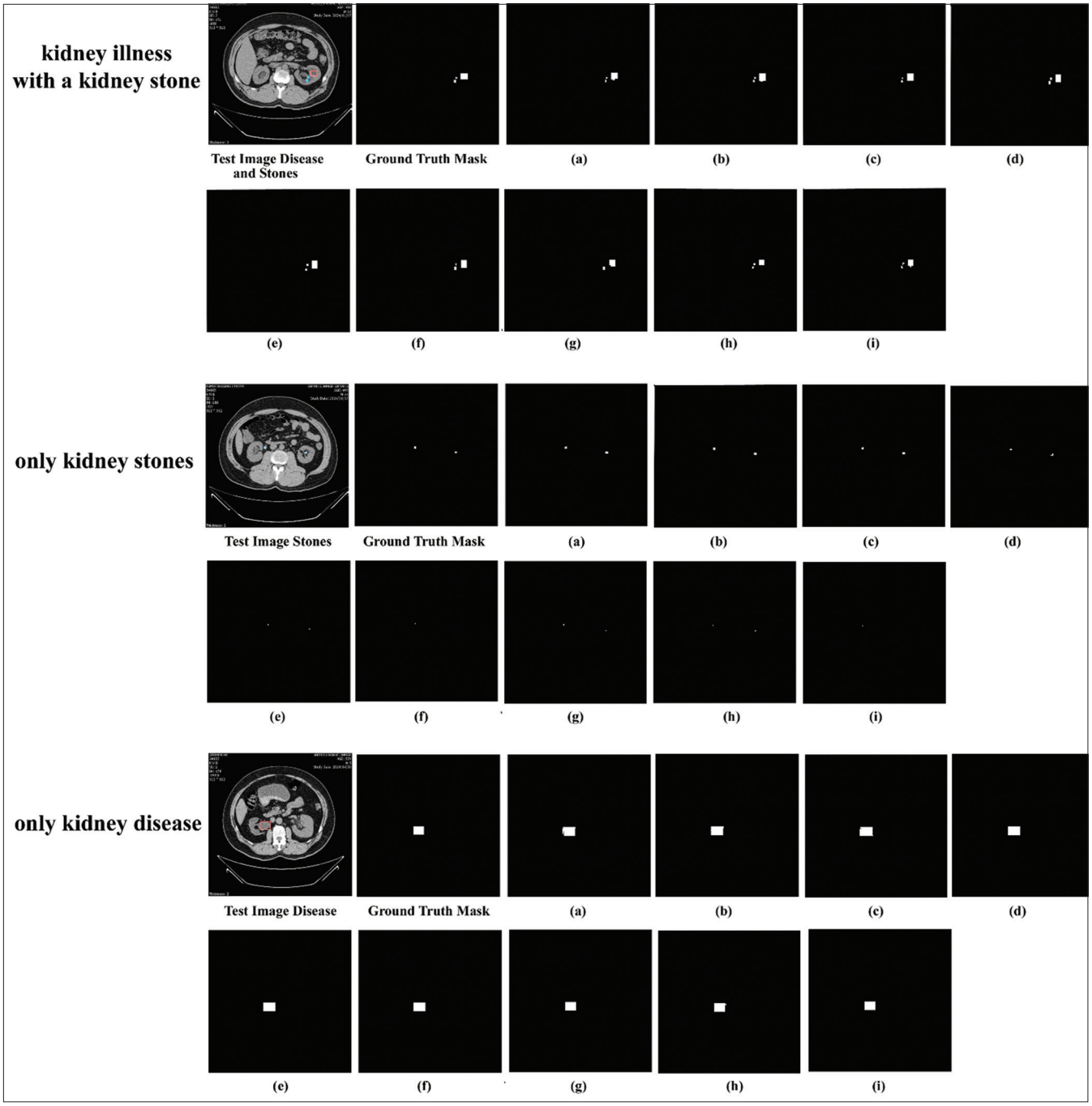


Fig. 7. Illustrates segmentation results for three scenarios: (top) renal disease with a kidney stone, (middle) a kidney stone only, and (bottom) renal disease only. Each set displays the CT test image, the ground truth mask, and the predicted masks ((a)–(i)) generated by the proposed U-Net + ResNet50 model.

with peak performance at 100 epochs, albeit overall metrics were slightly lower. Increasing the batch size from 8 to 24 while maintaining a 128 × 128 resolution produced marginal improvements but failed to surpass the original configuration. The analysis demonstrated that optimal segmentation quality was achieved using the hybrid model trained at 256 × 256 resolution for 100 epochs with a batch size of 8.

Fig. 9 presents the optimal performance of assessment metrics attained by the hybrid U-Net+ResNet50 model for

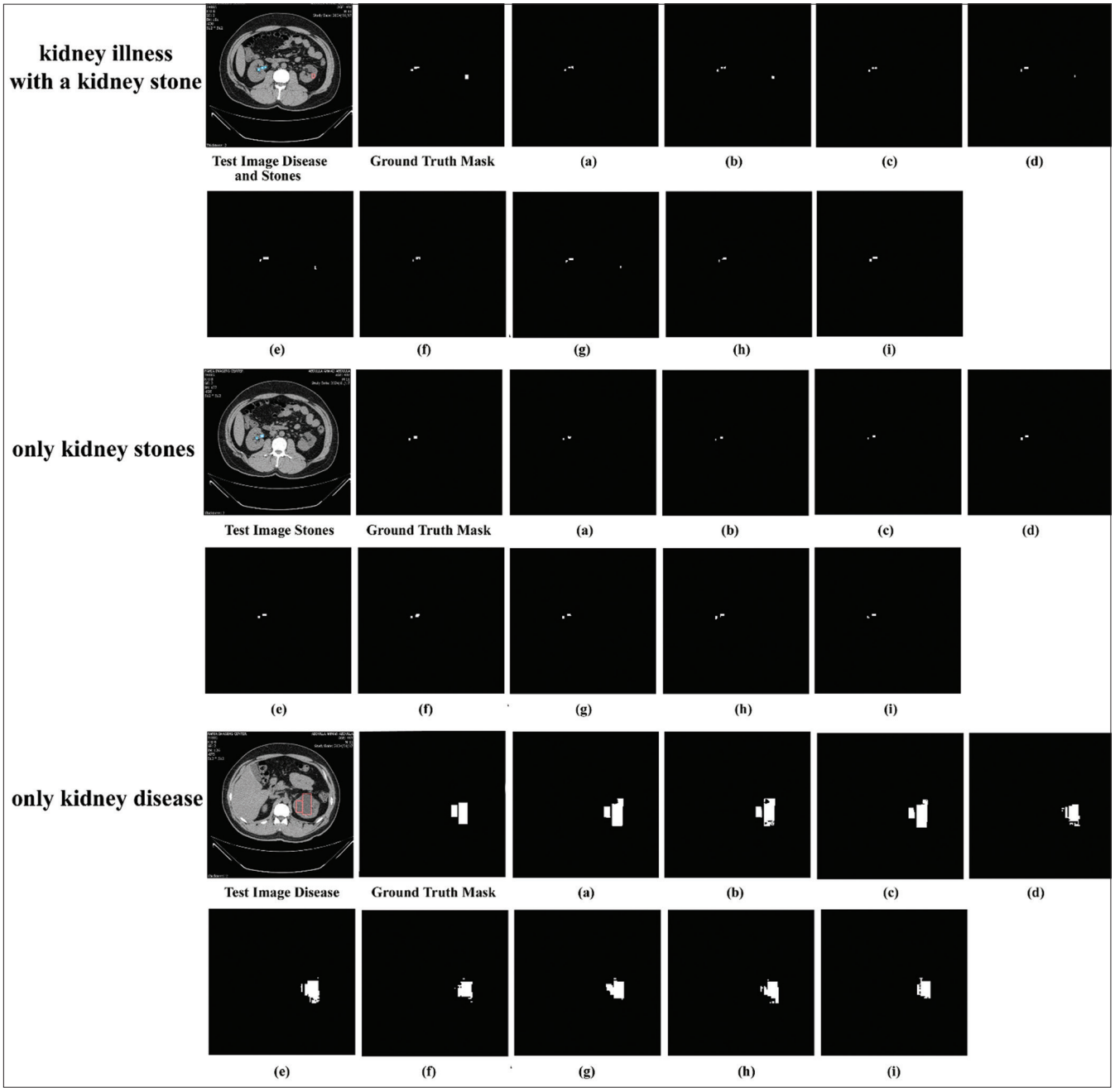


Fig. 8. Illustrates segmentation results for three scenarios: (Top) renal disease with a kidney stone, (middle) a kidney stone only, and (bottom) renal disease only. Each set displays the CT test image, the ground truth mask, and the predicted masks (a-i) generated by the proposed U-Net + VGG19 model.

segmenting renal disease and kidney stones across all three experimental scenarios, depicted in chart form. This graphic depiction highlights the model's efficacy across several parameters, including batch size, image size, and training epochs.

Fig. 10 illustrates the CM for the three U-Net + ResNet50 combinations across various circumstances. Scenario 2 (utilizing a 256 × 256 input size) exhibits the most equitable performance, recording the fewest false positives (655) and false negatives (826), underscoring its proficient

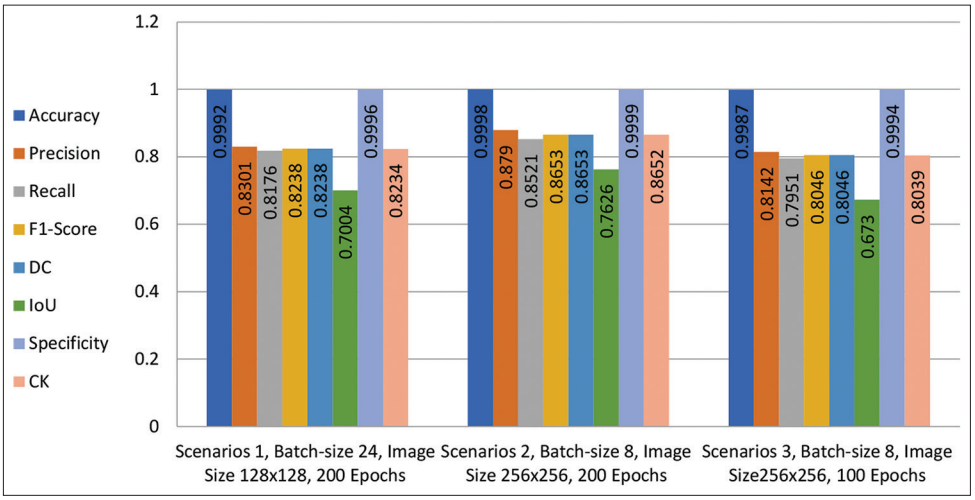


Fig. 9. Presents the highest evaluation criteria for the hybrid U-Net+ResNet50 model in all three scenarios.

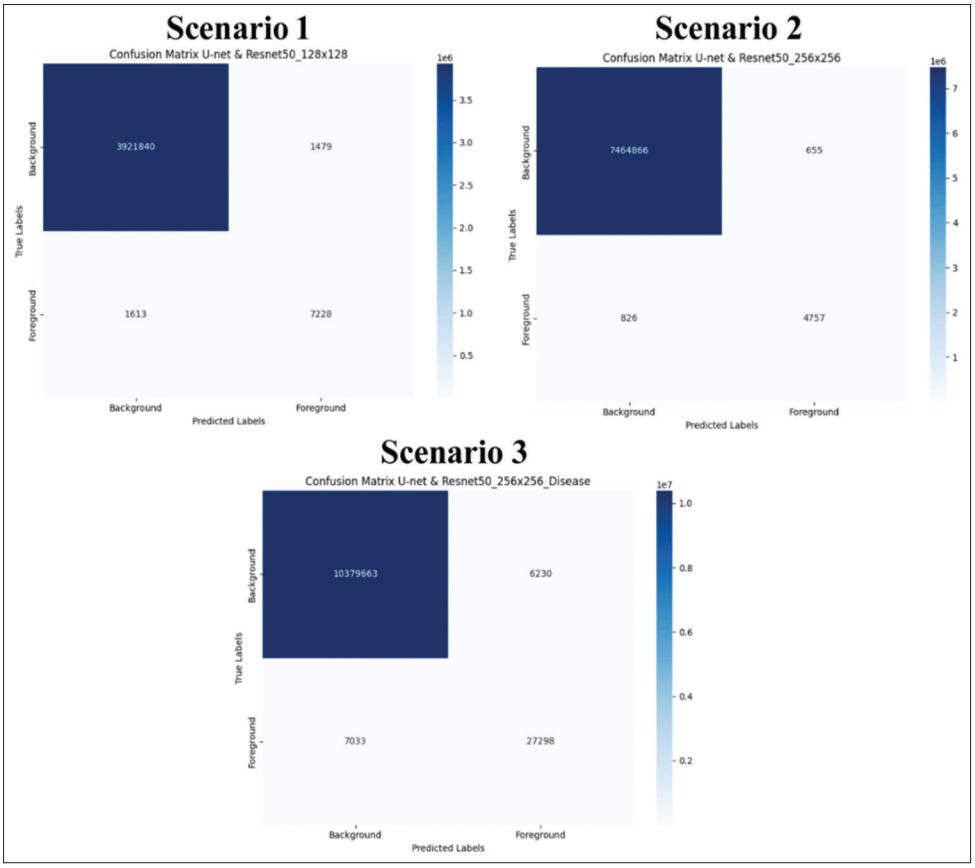


Fig. 10. Depicts the confusion matrices for the three distinct U-Net+ResNet50 model configurations employed in each scenario.

differentiation between background and foreground classes. Scenario 1 demonstrates robust segmentation performance but exhibits a marginally elevated incidence of misclassifications, suggesting potential enhancements in precision. Conversely, Scenario 3 shows the highest incidence of false positives (6,230), suggesting a propensity for over-segmentation, although it yields a substantial number of real positives. These findings underscore the enhanced

reliability of Scenario 2 in producing accurate and consistent segmentation results.

4.3.2. Hybrid U-Net + VGG19

In Scenario 1, the hybrid U-Net+VGG19 architecture was systematically evaluated by varying batch size, image resolution, and training epochs to assess their impact on kidney disease and kidney stone segmentation. The model demonstrated robust performance across all configurations, with optimal results achieved using 256 × 256 image resolution, a batch size of 8, and 200 training epochs. Under these conditions, the model achieved an accuracy of 0.9992, an F1-score of 0.8193, an IoU of 0.6939, and a CK of 0.8189, while maintaining computational efficiency with an average training time of 0.0310 seconds. Alternative configurations with reduced image resolution (128 × 128) and increased batch size (24) demonstrated comparable performance, with F1-scores ranging from 0.8066 to 0.8189, indicating the model's stability across different parameter combinations. These findings confirm the robustness and effectiveness of the U-Net+VGG19 hybrid model for kidney pathology segmentation.

In Scenario 2, the U-Net+VGG19 hybrid model was trained from scratch on kidney stone CT scans using various configurations of image dimensions, batch sizes, and training epochs to systematically evaluate the impact of each parameter on model performance. The optimal configuration, achieved with a 256 × 256 image size, batch size of 8, and 150 epochs, delivered the best balance of segmentation accuracy (F1-score = 0.8663), robustness (CK = 0.8662), and spatial overlap (IoU = 0.7641). While reducing image size to 128 × 128 improved processing time (~0.0123 s), it compromised segmentation quality with lower F1-scores peaking at 0.8439. Increasing the batch size to 24 showed minimal performance variations, and extending training beyond 150 epochs led to overfitting, with decreased recall and F1-scores, indicating that the 150-epoch threshold represents the optimal training duration for this model configuration.

In Scenario 3, the hybrid U-Net + VGG19 model was trained exclusively on CT scans of diseased kidneys to evaluate segmentation performance with reduced class diversity. The model demonstrated sensitivity to hyperparameters, achieving optimal results with a batch size of 8, image dimensions of 256 × 256, and 150 training epochs, yielding a consistent accuracy of 0.9987, an F1-score of 0.8032, and an IoU of 0.6712. Reducing the image size to 128 × 128 improved computational efficiency (ATT ~0.0120 seconds), but slightly compromised segmentation metrics. In contrast, increasing the batch size to 24 maintained accuracy but reduced the F1-score. Training beyond 150 epochs showed diminishing returns with potential overfitting, indicating that the optimal configuration balances spatial accuracy, overlap consistency, and computational cost. The results confirm the model's robust generalization capabilities in simplified classification contexts.

Fig. 11 illustrates a comparative chart that emphasizes the superior performance of the hybrid U-Net + VGG19 model in segmenting kidney stones and renal illnesses across the three experimental scenarios. This visual summary illustrates how critical parameters, such as batch size, image resolution, and training epochs, affect the model's efficacy, providing a concise understanding of its segmentation abilities across various configurations.

Fig. 12 shows the confusion matrices for the U-Net+VGG19 model across the three scenarios, providing an extensive overview of its segmentation efficacy in various medical imaging contexts. In Scenario 2 (stone alone), the model has good precision and recall, evidenced by the low counts of false positives (691) and false negatives (789), indicating a robust capacity to identify different stone features. Scenario 1 (stone + disease) and Scenario 3 (disease only) have consistently elevated accurate negative rates, indicating dependable background classification. Both situations indicate an increase in false negatives (6,674 and 6,643, respectively), suggesting a minor reduction in sensitivity when illness characteristics are evident, particularly in more intricate or nuanced pathological areas. The model demonstrates strong performance, highlighting the versatility and reliability of the U-Net + VGG19 architecture in addressing various image segmentation difficulties in the biomedical field.

Table 6 reveals distinct performance patterns across the hybrid architectures and segmentation scenarios. Both models achieve their highest F1-scores (0.8653 and 0.8663) in stone-only segmentation, indicating that isolated kidney stone identification represents the least complex task among the three scenarios. The U-Net+VGG19 model demonstrates superior consistency, maintaining F1-scores above 0.80 across all scenarios, while U-Net+ResNet50 shows more variable performance with a notable drop to 0.8046 in disease-only segmentation. Configuration analysis reveals that optimal performance typically requires higher resolution (256 × 256) images, contradicting the common assumption that computational efficiency necessarily improves outcomes. The processing time differences are substantial, with U-Net+ResNet50 achieving nearly twice the speed in the

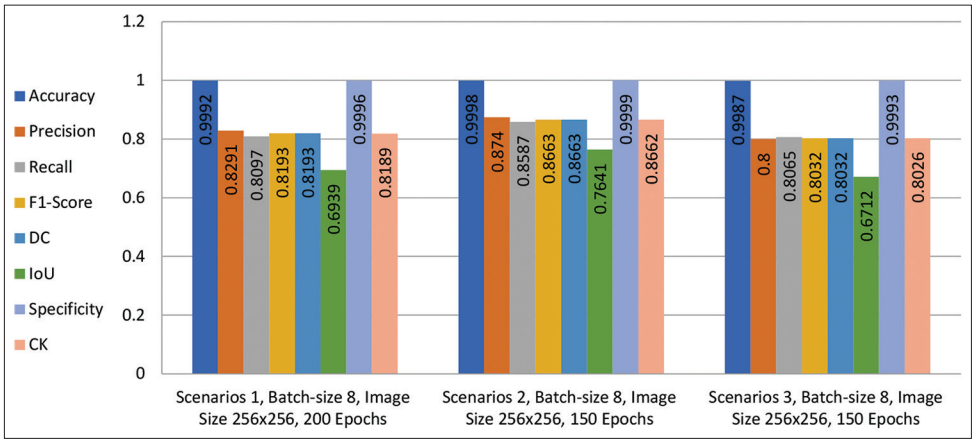


Fig. 11. Presents the highest evaluation criteria for the hybrid U-Net + VGG19 model in all three scenarios.

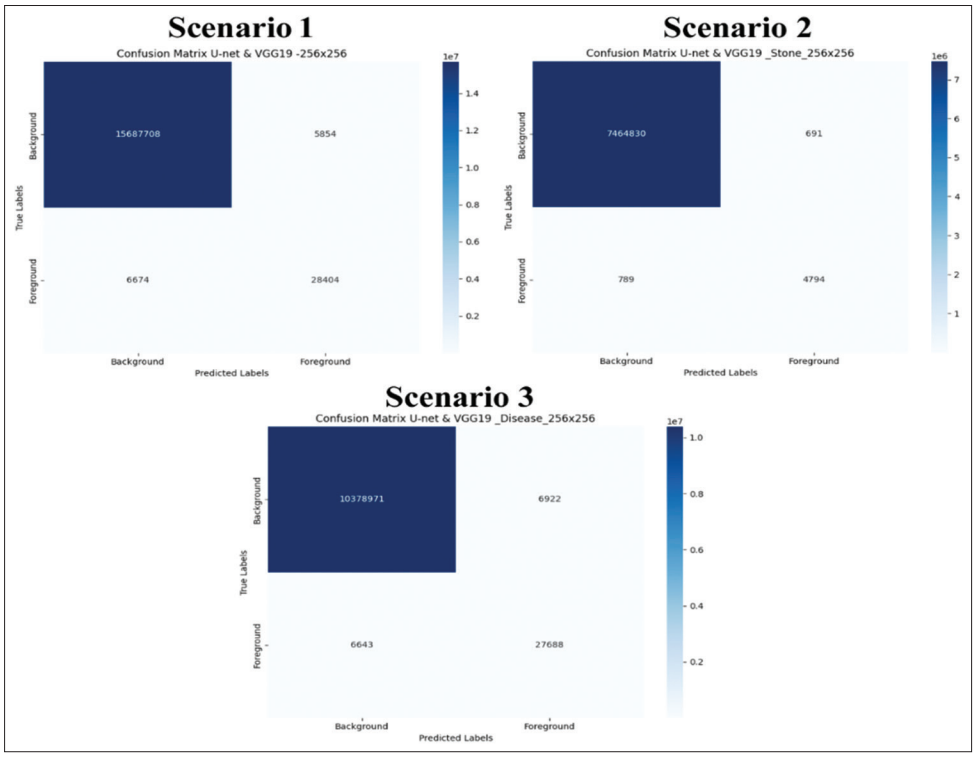


Fig. 12. Depicts the confusion matrices for the three distinct U-Net + VGG19 model configurations employed in each scenario.

TABLE 6: Optimal performance summary by model and scenario

Model	Scenario	Batch size	Image size	Epochs	Accuracy	Precision	Recall	F1-score	IoU	ATT (s)
U-Net+ResNet50	Disease+Stone	24	128×128	200	0.9992	0.8301	0.8176	0.8238	0.7004	0.0157
U-Net+ResNet50	Stone Only	8	256×256	200	0.9998	0.879	0.8521	0.8653	0.7626	0.0259
U-Net+ResNet50	Disease Only	8	256×256	100	0.9987	0.8142	0.7951	0.8046	0.673	0.0302
U-Net+VGG19	Disease+Stone	8	256×256	200	0.9992	0.8291	0.8097	0.8193	0.6939	0.031
U-Net+VGG19	Stone Only	8	256×256	150	0.9998	0.874	0.8587	0.8663	0.7641	0.0343
U-Net+VGG19	Disease Only	8	256×256	150	0.9987	0.8	0.8065	0.8032	0.6712	0.0296

disease+stone scenario (0.0157s vs. 0.0310s), suggesting that ResNet50's residual connections provide computational advantages beyond accuracy improvements. The IoU values consistently lag behind F1-scores by approximately 0.10-0.15 points, indicating that while both models achieve good pixel-wise classification, precise boundary delineation remains challenging in kidney pathology segmentation.

Table 7 exposes the architectural trade-offs between the two hybrid models and their clinical implications. U-Net+VGG19 emerges as the superior choice for stone segmentation, achieving the highest F1-score and IoU values, which translates to more accurate stone boundary detection crucial for treatment planning and surgical guidance. However, U-Net+ResNet50 demonstrates faster processing across most scenarios, with its computational efficiency being particularly pronounced in complex multi-class segmentation tasks. The consistent performance metric reveals U-Net+VGG19's architectural stability, maintaining more uniform results across diverse pathological conditions, while U-Net+ResNet50 shows task-specific optimization that favors certain scenarios over others. From a clinical deployment perspective, the processing time differences, though measured in milliseconds, could compound significantly in high-throughput environments where thousands of scans require analysis daily. The performance gap between stone and disease segmentation across both models (F1-score difference of ~0.06) suggests that kidney stone identification benefits from clearer radiological contrast compared to the more subtle tissue changes associated with kidney disease, highlighting the inherent complexity differences in pathological feature recognition.

Fig. 13 presents a comprehensive comparison of F1-score versus inference time, clearly demonstrating the performance-efficiency trade-offs between models. The chart reveals that U-Net+ResNet50 (blue) consistently achieves faster inference times across all scenarios, while U-Net+VGG19 (red) provides marginally higher F1-scores in certain configurations. Both models peak in Scenario 2 (kidney stones only), with minimal performance differences but significant efficiency variations, enabling clinicians to select the optimal model based on specific operational requirements.

The performance disparities between U-Net+ResNet50 and U-Net+VGG19 stem from their architectural attributes. ResNet50's residual connections enhance feature extraction for high-contrast structures like kidney stones, whereas VGG19's hierarchical depth more effectively maintains fine spatial details, resulting in marginally superior outcomes in complex pathological segmentation. In practical applications, U-Net+ResNet50 is advantageous for rapid and reliable stone identification, while U-Net+VGG19 delivers more uniform performance across various renal states, offering doctors adaptability in model selection based on the diagnostic context.

Table 8 provides a comparative examination of two hybrid models, U-Net integrated with ResNet50 and U-Net integrated with VGG19, across several scenarios. Scenario 2 of both models, utilizing a batch size of 8, an image dimension of 256 × 256, and a substantial epoch count, attained the highest overall performance. Significantly, the U-Net + VGG19 model in Scenario 2 marginally surpassed its equivalent, attaining an F1-score of 0.8663 and an IoU of 0.7641. The results demonstrate that the U-Net combined with VGG19 is a competitive alternative to ResNet50 regarding segmentation accuracy and efficiency.

Furthermore, the comparative assessment of the proposed hybrid designs reveals significant differences that elucidate their relative performance in the experimental circumstances. The U-Net+ResNet50 model attained optimal performance in stone-only segmentation, attributable to the residual connections that improve gradient propagation and facilitate more efficient extraction of distinguishing characteristics from high-contrast areas. Conversely, the U-Net+VGG19

TABLE 7: Comparative performance analysis

Performance aspect	Best Model	F1-Score	IoU	Processing Time (s)	Interpretation
Overall Performance	U-Net+VGG19 (Stone)	0.8663	0.7641	0.0343	Highest segmentation accuracy
Stone Segmentation	U-Net+VGG19	0.8663	0.7641	0.0343	Superior stone boundary detection
Disease Segmentation	U-Net+ResNet50	0.8046	0.673	0.0302	Better disease feature extraction
Computational Efficiency	U-Net+ResNet50 (Disease+Stone)	0.8238	0.7004	0.0157	Fastest processing time
Consistent Performance	U-Net+VGG19	0.8296*	0.7097*	0.0316*	Most stable across scenarios

*Average across all three scenarios

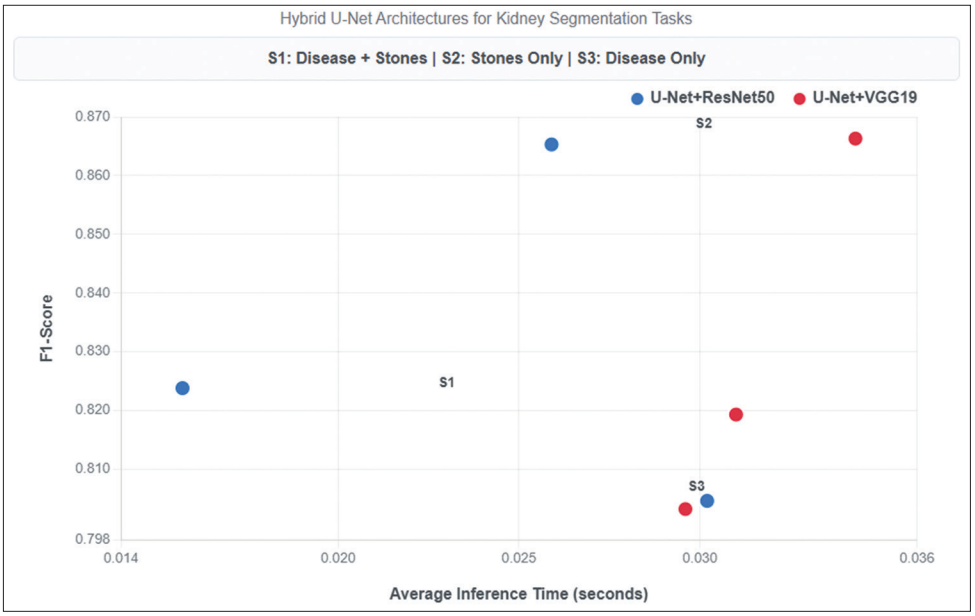


Fig. 13. Comparative chart of F1-score versus inference time for both models.

TABLE 8: A comparative analysis of the employed models across three scenarios, U-Net in conjunction with ResNet-50 and VGG-19

M	S	B-S	IS	Epo	Acc	Pre	Rec	F1-S	DC	IoU	Spe	CK	ATT
U-Net	1	24	128×128	200	0.9992	0.8301	0.8176	0.8238	0.8238	0.7004	0.9996	0.8234	0.0157 s
and	2	8	256×256	200	0.9998	0.879	0.8521	0.8653	0.8653	0.7626	0.9999	0.8652	0.0259 s
ResNet50	3	8	256×256	100	0.9987	0.8142	0.7951	0.8046	0.8046	0.673	0.9994	0.8039	0.0302 s
U-Net	1	8	256×256	200	0.9992	0.8291	0.8097	0.8193	0.8193	0.6939	0.9996	0.8189	0.031 s
and	2	8	256×256	150	0.9998	0.874	0.8587	0.8663	0.8663	0.7641	0.9999	0.8662	0.0343 s
VGG19	3	8	256×256	150	0.9987	0.8	0.8065	0.8032	0.8032	0.6712	0.9993	0.8026	0.0296 s

M: Model, S: Scenario, B-S: Batch-size, IS: Image Size, Epo: Epochs, Acc: Accuracy, Pre: Precision, Rec: Recall, F1-S: F1-Score, DC: Dice coefficient, Spe: Specificity, CK: Cohen's kappa, ATT: Average time test

model demonstrated unwavering stability throughout all three scenarios, a consequence presumably attributable to its hierarchical feature representation, which maintains intricate spatial information essential for precisely defining complicated renal diseases. The architectural differences explain the observed performance variances and have practical consequences for clinical practice. Enhanced segmentation of kidney stones and renal diseases might augment diagnostic reliability, enable precise quantification for treatment planning, and alleviate the manual workload on radiologists, hence promoting expedited and consistent clinical decision-making.

In addition, to evaluate the generalizability of the models to novel data, we re-examined the models, which excelled on the benchmark dataset, using the Kaggle dataset. The results showed that both hybrid models exhibited excellent performance, demonstrating their robustness across various datasets. The U-Net + VGG19 hybrid model, trained with a batch size of 8, an image size of 256 × 256, and for 150 epochs, attained an accuracy of 0.9997, a precision of 0.8495, a recall of 0.8502, an F1-score and DC of 0.8498, an IoU of 0.7389, a specificity of 0.9998, a CK of 0.8497, and an ATT of 0.0443 s. In the same way, the U-Net + ResNet50 hybrid model, employing the identical training configuration, achieved an accuracy of 0.9996, precision of 0.8499, recall of 0.8067, F1-score and DC of 0.8277, IoU of 0.7061, specificity of 0.9998, CK of 0.8276, and ATT of 0.0257 seconds. These results validate the efficacy and robustness of both models across various datasets.

Currently, the suggested hybrid U-Net models can be included in hospital (PACS) Picture Archiving and Communication Systems to autonomously segment kidney stones and

renal diseases from CT data. These outputs function as decision-support instruments, assisting radiologists in swiftly and reliably validating findings, while also supplying urologists with quantitative data for treatment formulation. Consequently, the models alleviate workload, enhance diagnostic reliability, and facilitate coordinated decision-making in authentic clinical workflows.

Several limitations should be acknowledged in this study. The relatively small dataset of 118 patients from a single institution, while supplemented with public data, may limit the generalizability of our findings across diverse populations, imaging protocols, and clinical settings. The lack of multimodal validation using other imaging modalities, such as MRI or US, restricts our ability to assess model performance across different diagnostic approaches commonly used in clinical practice. Furthermore, while our models demonstrate computational efficiency with processing times under 0.05 s per scan, the deployment of these DL architectures may pose significant challenges in resource-limited hospital environments where access to high-performance GPUs and adequate computational infrastructure is constrained. The memory requirements and initial setup costs for implementing such systems could create barriers for widespread clinical adoption, particularly in developing countries or smaller healthcare facilities. In addition, the requirement for specialized technical expertise to maintain and troubleshoot these AI systems may further limit their practical implementation in settings with limited IT support resources.

5. CONCLUSION

This research advances the field of automated medical image analysis by demonstrating that hybrid U-Net architectures can bridge the critical gap between segmentation accuracy and computational efficiency in kidney pathology detection. Our primary contribution lies in establishing a practical framework where established encoder architectures (ResNet50, VGG19) can be strategically integrated with U-Net decoders to achieve clinically viable performance without the prohibitive computational demands of transformer-based approaches. However, significant gaps remain in translating these technical achievements to diverse clinical environments, particularly regarding model robustness across different imaging protocols, patient demographics, and healthcare systems with varying computational resources. The geographical and institutional limitations of our validation highlight the urgent need for federated learning approaches that can leverage multi-institutional datasets while preserving patient privacy. Future research should prioritize three critical directions: developing domain adaptation techniques to ensure model generalizability across different CT scanner types and imaging protocols, integrating explainable AI methodologies to build clinician trust and facilitate clinical adoption, and establishing standardized evaluation frameworks that can assess not only technical performance but also clinical utility in real-world healthcare workflows. The ultimate success of automated kidney pathology detection will depend on creating systems that complement rather than replace radiological expertise, requiring interdisciplinary collaboration between computer scientists, radiologists, and healthcare administrators to address both technical and implementation challenges.

6. ACKNOWLEDGMENT

The assistance of numerous individuals was essential to the accomplishment of this endeavor. Many thanks to my advisor, the College of Science, Soran University, and many thanks to the Department of Computer Science, as well as to my friend Jafar Majidpour, who helped me a lot, and my friend Dr. Dyari Awla Awla, a specialist in kidney and urology. And finally, thanks to my brother Karokh and my wife.

REFERENCES

[1] N. A. Almansour, H. F. Syed, N. R. Khayat, R. K. Altheeb, R. E. Juri, J. Alhiyafi, S. Alrashed and S. O. Olatunji. "Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study". *Computers in Biology and Medicine*, vol. 109, pp. 101-111, 2019.

[2] F. Ahmed, S. Abbas, A. Athar, T. Shahzad, W. A. Khan, M. Alharbi, M. A. Khan and A. Ahmed. "Identification of kidney stones in KUB X-ray images using VGG16 empowered with explainable artificial intelligence". *Scientific Reports,* vol. 14, no. 1, p. 6173, 2024.

[3] M. Akram, V. Jahrreiss, A. Skolarikos, R. Geraghty, L. Tzelves, E. Emilliani, N. F. Davis and B. K. Somani. "Urological guidelines for kidney stones: Overview and comprehensive update". *Journal of Clinical Medicine*, vol. 13, no. 4, p. 1114, 2024.

[4] R. Magherini, E. Mussi, Y. Volpe, R. Furferi, F. Buonamici and M. Servi. "Machine learning for renal pathologies: An updated survey". *Sensors (Basel)*, vol. 22, no. 13, pp. 4989, 2022.

[5] M. H. Hesamian, W. Jia, X. He and P. Kennedy. "Deep learning techniques for medical image segmentation: Achievements and challenges". *Journal of Imaging Informatics in Medicine*, vol. 32, no. 4, pp. 582-596, 2019.

[6] Y. Wu and Z. Yi. "Automated detection of kidney abnormalities using multi-feature fusion convolutional neural networks". *Knowledge-Based Systems*, vol. 200, p. 105873, 2020.

[7] T. A. Rashid, J. Majidpour, R. Thinakaran, M. Batumalay, D. A. Dewi, B. A. Hassan, H. Dadgar and H. Arabi. "NSGA-II-DL: Metaheuristic optimal feature selection with deep learning framework for HER2 classification in breast cancer". *IEEE Access*, vol. 12, pp. 38885-38898, 2024.

[8] K. K. Patro, A. Jayaprakash, U. Rajendra Achary, M. Hammad, O. Yildirim and P. Pławiak. "Application of kronecker convolutions in deep learning technique for automated detection of kidney stones with coronal CT images". *Information Sciences*, vol. 640, p. 119005, 2023.

[9] W. Brisbane, M. R. Bailey and M. D. Sorensen. "An overview of kidney stone imaging techniques". *Nature Reviews Urology,* vol. 13, no. 11, pp. 654-662, 2016.

[10] D. A. Mahmood and S. A. Aminfar. "Efficient machine learning and deep learning techniques for detection of breast cancer tumor". *BioMed Target Journal*, vol. 2, no. 1, pp. 1-13, 2024.

[11] S. Asif, M. Awais and S. U. R. Khan. "IR-CNN: Inception residual network for detecting kidney abnormalities from CT images". *Network Modeling Analysis in Health Informatics and Bioinformatics,* vol. 12, no. 1, p. 35, 2023.

[12] A. Parakh, H. Lee, J. H. Lee, B. H. Eisner, D. V. Sahani and S. Do. "Urinary stone detection on CT images using deep convolutional neural networks: Evaluation of model performance and generalization". *Radiology: Artificial Intelligence*, vol. 1, no. 4, p. e180066, 2019.

[13] M. Kaur and D. Singh. "Fusion of medical images using deep belief networks". *Cluster Computing*, vol. 23, no. 2, pp. 1439-1453, 2020.

[14] Z. H. Huang, Y. Y. Liu, W. J. Wu and K. W. Huang. "Design and validation of a deep learning model for renal stone detection and segmentation on kidney-ureter-bladder images". *Bioengineering*, vol. 10, no. 8, 970.

[15] K. Yildirim, P. G. Bozdag, M. Talo, O. Yildirim, M. Karabatak and U. R. Acharya. "Deep learning model for automated kidney stone detection using coronal CT images". *Computers in Biology and Medicine*, vol. 135, p. 104569, 2021.

[16] L. A. Fitri, A. Koudonas, G. Langas, S. Tsiakaras, D. Memmos, I. Mykoniatis, E. N. Symeonidis, D. Tsiptsios, E. Savvides, I. Vakalopoulos, G. Dimitriadis and J. De la Rosette. "Automated classification of urinary stones based on microcomputed tomography images using convolutional neural network". *Physical Medicine*, vol. 78, pp. 201-208, 2020.

[17] W. Zhao, D. Jiang, J. Peña Queralta and T. Westerlund. "MSS U-Net: 3D segmentation of kidneys and tumors from CT images with a multi-scale supervised U-Net". *Informatics in Medicine Unlocked*, vol. 19, p. 100357, 2020.

[18] A. Alqahtani, S. Alsubai, A. Binbusayyis, M. Sha, A. Gumaei and Y. D. Zhang. "Optimizing kidney stone prediction through urinary analysis with improved binary particle Swarm optimization and extreme gradient boosting". *Mathematics*, vol. 11, no. 7, p. 1717.

[19] S. W. Yang, Y. K. Hyon, H. S. Na, L. Jin, J. G. Lee, J. M. Park, J. Y. Lee, J. H. Shin, J. S. Lim, Y. G. Na, K. Jeon, T. Ha, J. Kim and K. H. Song. "Machine learning prediction of stone-free success in patients with urinary stone after treatment of shock wave lithotripsy". *BMC Urology,* vol. 20, no. 1, p. 88, 2020.

[20] A. J. Daniel, C. E. Buchanan, T. Allcock, D. Scerri, E. F. Cox, B. L. Prestwich and S. T. Francis. "Automated renal segmentation in healthy and chronic kidney disease subjects using a convolutional neural network". *Magnetic Resonance in Medicine*, vol. 86, no. 2, pp. 1125-1136, 2021.

[21] L. Ma, Y. Qiao, R. Wang, H. Chen, G. Liu, H. Xiao and R. Dai. "Machine learning models decoding the association between urinary stone diseases and metabolic urinary profiles". *Metabolites*, vol. 14, no. 12, p. 674.

[22] I. Aksakallı, S. Kaçdıoğlu and Y. S. Hanay. "Kidney x-ray images classification using machine learning and deep learning methods". *Balkan Journal of Electrical and Computer Engineering*, vol. 9, no. 2, pp. 144-151, 2021.

[23] D. C. Elton, E. B. Turkbey, P. J. Pickhardt and R. M. Summers. "A deep learning system for automated kidney stone detection and volumetric segmentation on noncontrast CT scans". *Med Phys,* vol. 49, no. 4, pp. 2545-2554, 2022.

[24] N. Blau, E. Klang, N. Kiryati, M. Amitai, O. Portnoy and A. Mayer. "Fully automatic detection of renal cysts in abdominal CT scans". *International Journal of Computer Assisted Radiology and Surgery,* vol. 13, pp. 957-966, 2018.

[25] X. Xiong, Y. Guo, Y. Wang, D. Zhang, Z. Ye, S. Zhang and X. Xin. "Kidney tumor segmentation in ultrasound images using adaptive sub-regional evolution level set models". *Sheng Wu yi Xue Gong Cheng Xue Za Zhi = Journal of Biomedical Engineering = Shengwu Yixue Gongchengxue Zazhi,* vol. 36, no. 6, pp. 945-956, 2019.

[26] R. Gaikar, F. Zabihollahy, M. W. Elfaal, A. Azad, N. Schieda and E. Ukwatta. "Transfer learning-based approach for automated kidney segmentation on multiparametric MRI sequences". *Journal of Medical Imaging*, vol. 9, no. 3, p. 036001, 2022.

[27] X. Fu, H. Liu, X. Bi and X. Gong. "Deep-learning-based CT imaging in the quantitative evaluation of chronic kidney diseases". *Journal of Healthcare Engineering,* vol. 2021, p. 3774423, 2021.

[28] M. Goyal, J. Guo, L. Hinojosa, K. Hulsey and I. Pedrosa. "Automated kidney segmentation by mask R-CNN in T2-weighted magnetic resonance imaging". In: *Medical Imaging 2022: Computer-Aided Diagnosis*, vol. 12033. Springer, Berlin, pp. 803-808, 2022.

[29] G. Sharma, V. Anand, R. Chauhan, H. S. Pokhariya, S. Gupta and G. Sunil. "Transfer Learning Empowered Multi-Class Classification of Kidney Diseases: A Deep Learning Approach". In: *2024 2nd International Conference on Advancement in Computation and Computer Technologies (InCACCT)*, pp. 240-245, 2024.

[30] H. Göker. "Transfer Learning-Based Classification of Kidney Stone From Computed Tomography Images". In: *2024 8th International Artificial Intelligence and Data Processing Symposium (IDAP)*, pp. 1-7, 2024.

[31] M. N. Islam, M. Hasan, M. K. Hossain, M. G. R. Alam, M. Z. Uddin and A. Soylu. "Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from CT-radiography". *Scientific Reports*, vol. 12, no. 1, p. 11440, 2022.

[32] F. J. N. Mactina and S. Neduncheliyan. "Multi-classification of kidney abnormalities in sonography using the LOA-MFO and long-term recurrent convolutional network". *Multimedia Tools and Applications*, vol. 83, no. 5, pp. 13577-13612, 2024.

[33] A. Odenrick, N. Kartalis, N. Voulgarakis, F. Morsbach and L. Loizou. "The role of contrast-enhanced computed tomography to detect renal stones". *Abdominal Radiology*, vol. 44, pp. 652-660, 2019.

[34] M. S. B. Islam, M. S. I. Sumon, R. Sarmun, E. H. Bhuiyan and M. E. H. Chowdhury. "Classification and segmentation of kidney MRI images for chronic kidney disease detection". *Computers and Electrical Engineering*, vol. 119, p. 109613, 2024.

[35] U. Kilic, I. Karabey Aksakalli, G. Tumuklu Ozyer, T. Aksakalli, B. Ozyer and S. Adanur. "Exploring the effect of image

enhancement techniques with deep neural networks on direct urinary system X-Ray (DUSX) images for automated kidney stone detection". *International Journal of Intelligent Systems*, vol. 2023, no. 1, p. 3801485, 2023.

[36] G. Chartrand, P. M. Cheng, E. Vorontsov, M. Drozdzal, S. Turcotte, C. J. Pal and S. Kadoury, A. Tang. "Deep learning: A primer for radiologists". *Radiographics*, vol. 37, no. 7, pp. 2113-2131, 2017.

[37] H. P. Chan, R. K. Samala, L. M. Hadjiiski and C. Zhou. "Deep learning in medical image analysis". *Deep learning in Medical Image Analysis: Challenges and Applications*. Berlin: Springer, pp. 3-21, 2020.

[38] A. Abdelrahman and S. Viriri. "FPN-SE-ResNet model for accurate diagnosis of kidney tumors using CT images". *Applied Sciences*, vol. 13, no. 17, p. 9802, 2023.

[39] O. Ronneberger, P. Fischer and T. Brox. "U-net: Convolutional Networks for Biomedical Image Segmentation. In: *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18"*. Springer, pp. 234-241, 2015.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Cascading Overtopping-Induced Dam Failures in a Transboundary Basin: Hydrodynamic Modeling from Gawshan to Darbandikhan Dam

Fryad Hama Ibrahim, Kawa Zaidan Abdulrahman

Department of Water Resources Engineering, College of Engineering, University of Sulaimani, Sulaymaniyah, Iraq

ABSTRACT

This study investigates the cascading failure potential of a multi-dam system in the transboundary Diyala (Sirwan) River basin, spanning Iran and Iraq. Using a two-dimensional hydrodynamic model developed in Hydrologic Engineering Center's River Analysis System 6.6, the study simulates overtopping- and piping-induced breach scenarios at Gawshan Dam and assesses their propagation through downstream reservoirs including Zhave, Daryan, and Hirwa, culminating at Darbandikhan Dam. The results highlight the formation of a compound flood wave that exceeds crest elevations, particularly under overtopping scenarios, leading to overtopping failure at Darbandikhan Dam despite emergency discharge operations. Scenario 1, representing full reservoir conditions and sequential overtopping failure, was identified as the most severe. The model outputs show a rise in water surface elevation from 485.0 m to 495.5 m within 412 min at Darbandikhan Reservoir, causing structural overtopping. Extensive inundation impacts both Iranian upstream settlements and Iraqi downstream communities, submerging over 20 villages and 80 km^2 of agricultural land. The findings underline the need for transboundary coordination, reservoir reoperation protocols, and updated Emergency Action Plans to mitigate the risk of cascading dam failures in seismically active regions.

Index Terms: HEC-RAS, Dam Failure, Inundation Map, Flood Hazard Map, Flood Disaster

1. INTRODUCTION

A dam is an engineered barrier constructed across a river to retain and supply water for various uses, including irrigation, hydropower, flood control, water supply, recreation, and navigation [1]. As noted by M. Zagonjolli [2], the global count of dams exceeds 800,000, built primarily to serve these functions. However, the accumulation of vast water volumes behind these structures can pose significant hazards to populations living downstream.

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp251-265

E-ISSN: 2521-4217

P-ISSN: 2521-4209

Copyright © 2025 Ibrahim and Abdulrahman This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Although dams offer substantial advantages to society, they also introduce serious risks. A dam's structural failure may lead to uncontrolled water release, potentially resulting in catastrophic flooding and extensive environmental destruction. Such dam-break floods tend to be more devastating than natural flood events, often causing fatalities, economic losses, ecological damage, sediment displacement, terrain deformation, and psychological trauma in affected communities. Worldwide, floods account for approximately 40% of deaths resulting from natural disasters.

Historical data indicate that dam failures affect all structural types. According to Costa [3], embankment dams have failed primarily due to piping (38%), overtopping (35%), foundation failure (21%), and other causes (6%). In addition, failure rates are high during the first 1–5 years and between 20 and

Corresponding author's e-mail: Fryad Hama Ibrahim, Department of Water Resources Engineering, College of Engineering, University of Sulaimani, KRG, Iraq. E-mail: Fryad.ibrahim@univsul.edu.iq

Received: 20-07-2025 Accepted: 25-09-2025 Published: 13-10-2025

50 years after dam commissioning, accounting for 31% and 22% of failures, respectively [4].

Research based on over 1065 failures of earthen dams highlights overtopping and piping as the leading causes. Spillways, downstream slopes, and foundations are key points vulnerable to overtopping, whereas piping threatens the dam body as a whole [5]. Most dam failures are preceded by observable warning signs such as the development of a breach. Once a breach forms, it triggers a flood wave originating from the water stored upstream [6]. Therefore, it is essential for downstream residents to be informed about such threats, as early awareness can prevent costly damages or enable mitigation efforts [7]. The prediction of breach outflow hydrographs and their routing is central to dam breach studies [8]. Critical elements in such analyses include the shape of the flood hydrograph and the length of the river section affected [7]. Reliable flood prediction supports the development of emergency preparedness plans to reduce risks to life and property [9]. Accurate modeling of dam-break floods and downstream propagation is typically carried out using hydraulic simulation tools [10], [11].

Common modeling tools include DAMBRK [4], MIKE developed by DHI [12], and the Hydrologic Engineering Center's River Analysis System (HEC-RAS) from USACE [13]. Moreover, the physically based model Dam Break-Institute of Water Resources and Hydropower Research has also been utilized in breach studies [14]. Among these, HEC-RAS stands out for its widespread use in dam failure research [15]-[24]. For instance, HEC-RAS was employed to calculate flood depths downstream of Um Al-Khair Dam and compared against observed water levels during the breach event [25].

Combining hydraulic modeling, GIS tools, and risk assessment frameworks enables the identification of flood-prone regions following dam breaches. These methods integrate spatial data, flood depths, and velocities to evaluate hazard levels [22], [26], [27]. HEC-RAS is often used to simulate dam failure timing and geometry, incorporating peak discharge values. To achieve this, cross-sectional data are imported through digital elevation models (DEMs) and HEC-GeoRAS. For example, Sharma and Mujumdar [19] applied HEC-RAS, HEC-GeoRAS, and GIS to study the Ajwa Dam breach after heavy rainfall, deriving critical outputs such as peak water levels, downstream discharge, flow velocities, and inundation mapping.

The border between Iran and Iraq is seismically active, housing major hydraulic structures such as Gawshan and Daryan Dam (Iran) and Darbandikhan Dam (Iraq) within the Diyala River basin. A Mw 7.3 earthquake on November 12, 2017, caused structural impairments to Darbandikhan Dam [28], revealing its susceptibility to seismic events and underscoring the risk of cascading upstream dam failures. Consequently, Darbandikhan Dam is now operated below its rule curve to minimize potential damage from sudden upstream inflows.

This study focuses on modeling the hydrodynamic consequences of hypothetical upstream dam failures specifically Gawshan, Zhave, Daryan, and Hirwa on Darbandikhan Dam using HEC-RAS 6.6. The analysis investigates flood wave movement, peak discharge values, and inundation extent across a range of breach scenarios. The primary objectives are as follows: (1) To evaluate Darbandikhan Dam's structural response to incoming flood waves, and (2) to identify scenario-specific vulnerabilities to guide the development of Emergency Action Plans (EAPs) and support adaptive strategies in reservoir management.

2. MATERIALS AND METHODS

2.1. Study Area

The study area is situated within the Diyala (Sirwan) River Basin, encompassing a transboundary river system shared between Iran and Iraq. This region includes a cascade of major dams Gawshan, Zhave (Java), Daryan, Hirwa, and Darbandikhan, each contributing critically to water resource management, hydropower production, and flood control. The interconnected nature of these structures makes the basin highly sensitive to cascading dam failure scenarios. A general overview of the study area and dam locations is presented in Fig. 1. The Darbandikhan Dam (Fig. 2), in particular its considerable height, substantial reservoir volume, and critical location upstream of major population centers including the cities of Kalar and Diyala, greatly amplifies the potential consequences of failure. This positioning greatly amplifies the consequences of a potential dam failure, underscoring the necessity for comprehensive risk evaluation and effective emergency response strategies. The dam is located roughly 65 km southeast of Sulaimani City and around 230 km northeast of Baghdad. Built between 1956 and 1961, Darbandikhan is an embankment dam serving multiple functions such as irrigation, hydroelectric power generation, and flood control.

The reservoir is mainly supplied by four tributaries: The Tanjaro and Zalm rivers from Iraq, and the Sirwan and

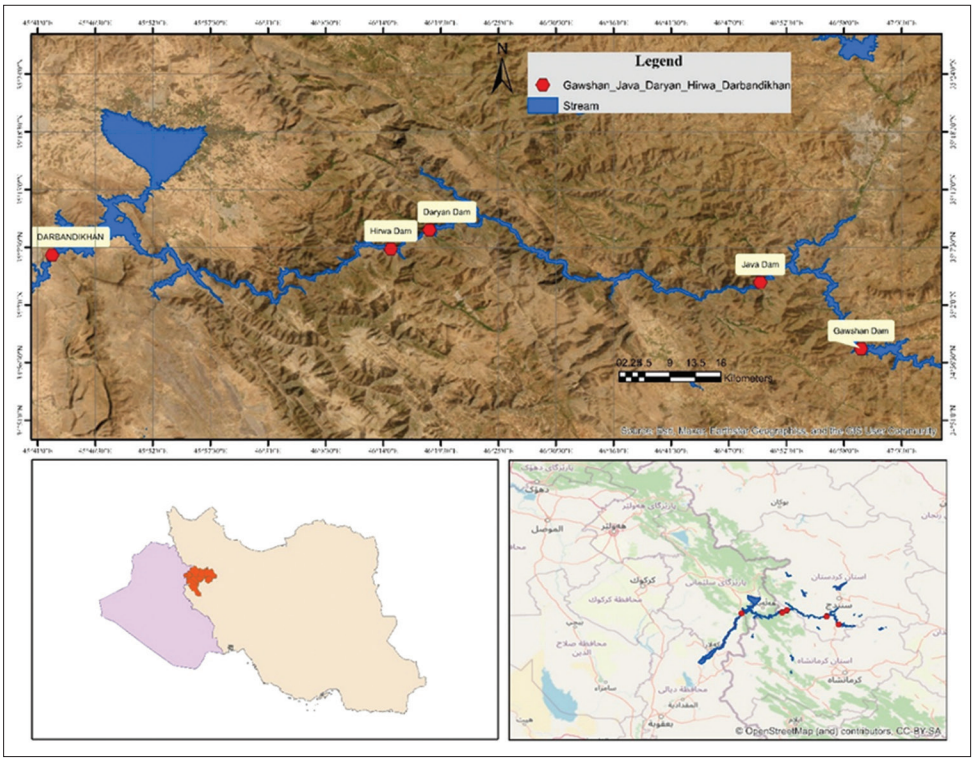


Fig. 1. Map of Stream Gawshan to Darbandikhan.

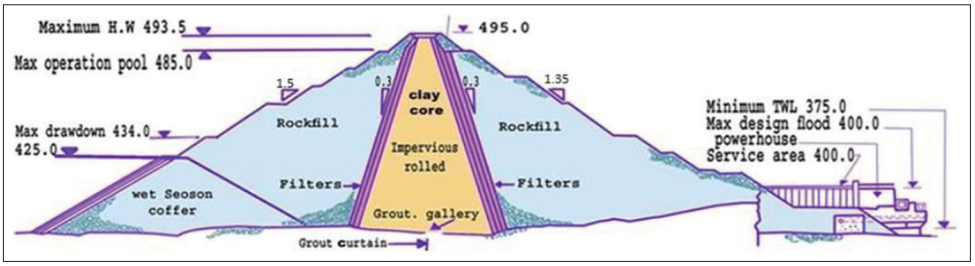


Fig. 2. Maximum cross section of the Darbandikhan Dam (Adopted from the Darbandikhan Dam Directorate).

Zmkan rivers flowing from Iran. The dam rises to a height of 128 m and spans 445 m along its crest. Originally, its reservoir had a total capacity of (3.0×10^9 m^3), which has since declined to (2.55×10^9 m^3) due to sediment buildup over six decades of operation. This includes (2.05×10^9 m^3) of usable (active) storage and (0.50×10^9 m^3) categorized as inactive storage. The spillway system comprises a gated chute with three outlets, collectively capable of handling up to 11,400 m^3/s during peak flow events. The reservoir itself spans an area of 113 km^2 and is part of a catchment that drains 17,850 km^2. At its highest operational level, the water surface can reach an elevation of 485 m, ensuring optimal water retention and distribution. The dam's power station uses three Francis turbines, each with a maximum discharge capacity of 113 m^3/s, and a minimum flow release of 63 m^3/s through the turbine gates.

Gawshan Dam is an embankment-type structure positioned near Kamyaran in Iran's Kurdistan Province, constructed to regulate the flow of the Gaveh River, a tributary of the Sirwan River Fig. 3. Its precise location is marked by the geographic coordinates (34°57'48"N) and (46°59'40"E), as illustrated in Fig. 4. The dam's construction commenced in 1992 and was completed in 2004 [29]. It is currently managed by the Iran Water and Power Resources Development Company. With a total reservoir storage capacity of 550 million cubic meters, Gawshan ranks among the largest hydraulic infrastructure initiatives in western Iran. It plays a key role in the region's

water and energy management by supplying 183 million cubic meters annually for agricultural irrigation and 63 million cubic meters to meet the municipal water needs of Kermanshah. The dam is also equipped with hydroelectric generation facilities, producing up to 11 megawatts of electricity. In addition, a 21-km-long diversion tunnel was constructed to deliver water at a rate of 30 cubic m/s, enabling irrigation across 31,000 ha of productive farmland. Due to its location and multifunctional design, Gawshan Dam serves as a key asset in regional water allocation, agricultural development, and energy production [30].

Zhave (Java) Dam located in Iran's Kurdistan Province Fig. 5. Zhave Dam sits at the geographic coordinates of (35°04'N) and (46°50'E), with its crest reaching an elevation of 1320 m above sea level. It is situated approximately 6 km downstream from the junction where the Gheshlagh and Gavehrood rivers meet two rivers that originate in the eastern highlands of the Sirwan River Basin. Their confluence forms the Zhave River, which flows southwestward through the region [32]. The Zhave Dam Reservoir stretches upstream into the main channels of both contributing rivers, thus offering substantial hydrological regulation over the sub-basin. Further upstream, the Gavehrood River is also regulated by Gawshan and Soleyman Shah dams, which contribute to managing the flow entering the Zhave system. By controlling inflow from these upstream sources, Zhave Dam plays a key role in regional flood mitigation and water resource storage [33].

Fig. 3. Gawshan Dam.

As illustrated in Fig. 6, Daryan Dam is positioned upstream of Darbandikhan Dam along the Sirwan River in Paveh County, Kermanshah Province, Iran. The dam lies approximately 55.8 km away from Darbandikhan Dam in a straight-line distance. Construction activities commenced in 2009, and the dam became operational by the end of 2015 [34]. With a structural height of 169 m, the primary objective of the dam's development was the generation of hydroelectric power, with a capacity of 210 megawatts. The reservoir created by Daryan Dam spans an area of 10 km^2, reaching a maximum width of 800 m. Its total storage volume is 316.3 million cubic meters, of which 281 million cubic meters are designated as active storage.

Shown in Fig. 7, Hirwa Dam is a concrete diversion structure built in 2018. It is located around 8 km downstream of the

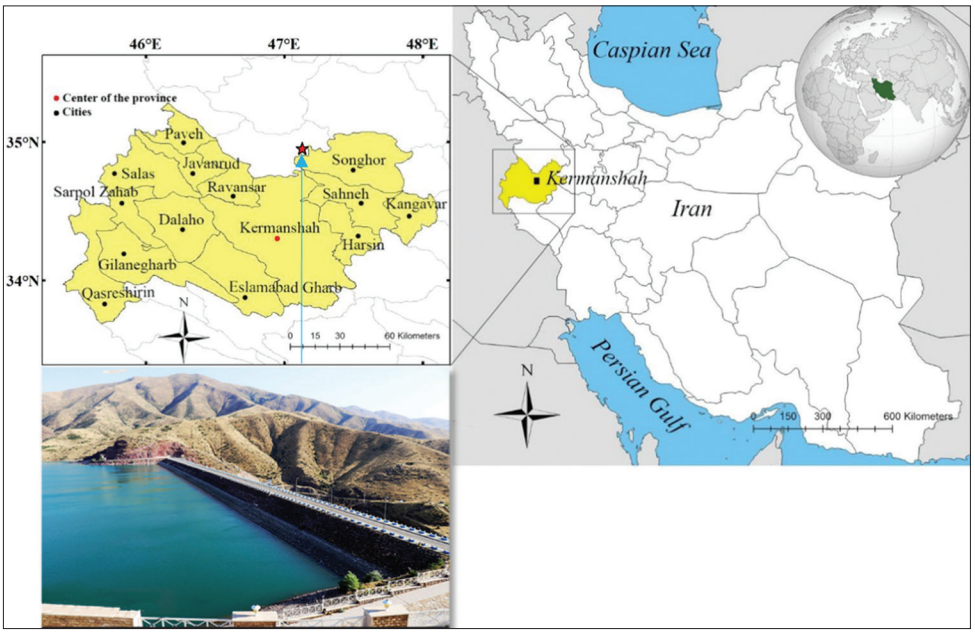


Fig. 4. Iran map and the study area (Kermanshah province, Gawshan dam) [31].

Fig. 5. Zhave Dam.

Fig. 6. Upstream view of Daryan dam [31].

Fig. 7. Hirwa Dam looking upstream.

Daryan Dam, in Hirwa Village, Paveh County, Iran. This dam plays a key role in managing the outflow from Daryan Dam before its diversion into the Nowsud Tunnel, thereby ensuring consistent water delivery for both irrigation and hydroelectric power generation. Hirwa Dam has a height of 45 m and a total storage capacity of 12 million cubic meters [35].

Considering the strategic importance of the interconnected dam system in the region, this research concentrates on the cascading failure risks associated with the Gawshan, Zhave, Daryan, and Hirwa dams. It specifically models breach scenarios to evaluate their downstream hydrodynamic consequences on the Darbandikhan Dam. A sequential failure of these upstream dams could severely compromise the structural stability, flood management capability, bank protection, and overall operational performance of the Darbandikhan Dam.

2.2. Problem Statement

Gawshan Dam, situated on a principal tributary of the Sirwan River, is the largest and most critical storage reservoir in the upper Sirwan basin. Due to its considerable impoundment capacity, a hypothetical failure, particularly one triggered by overtopping or pipping due to seismic disturbance, would generate an exceptionally high-energy flood wave that no downstream reservoir, including Zhave, Daryan, or Hirwa dams, could effectively withstand. The scale of this outflow would likely initiate a cascading sequence of structural failures, amplifying the flood wave as it moves downstream and ultimately threatening the stability of Darbandikhan Dam, one of many vital infrastructures for hydropower, irrigation, and water supply in Iraq.

This hydrodynamic threat is compounded by the fact that the region lies within a seismically active zone as seismic map, as shown in Fig. 8.

The Mw 7.3 earthquake of November 12, 2017, which caused notable structural damage to Darbandikhan Dam [28], underscores the potential for seismic events to trigger or exacerbate dam failures. Earth-fill dams such as Gawshan are particularly vulnerable to combined seismic loading and elevated reservoir levels, which may lead to slope instability, foundation movement, or overtopping-induced breach. In response to these risks, Darbandikhan Dam is currently operated below its designated rule curve to provide buffer capacity for absorbing sudden upstream inflows. While this precautionary strategy enhances flood resilience, it also introduces trade-offs in terms of reduced water storage for hydropower generation, agricultural irrigation, and municipal supply.

These operational limitations highlight the urgent need to determine a safe and sustainable reservoir level that balances hazard mitigation with resource utilization. Despite the severity of these scenarios, there remains a notable lack of quantitative research on the cascading impacts of Gawshan Dam failure on the broader Sirwan River system. This study aims to fill that gap by employing 2D hydrodynamic modeling using HEC-RAS 6.6 to simulate breach scenarios at Gawshan Dam and assess their downstream effects, particularly on

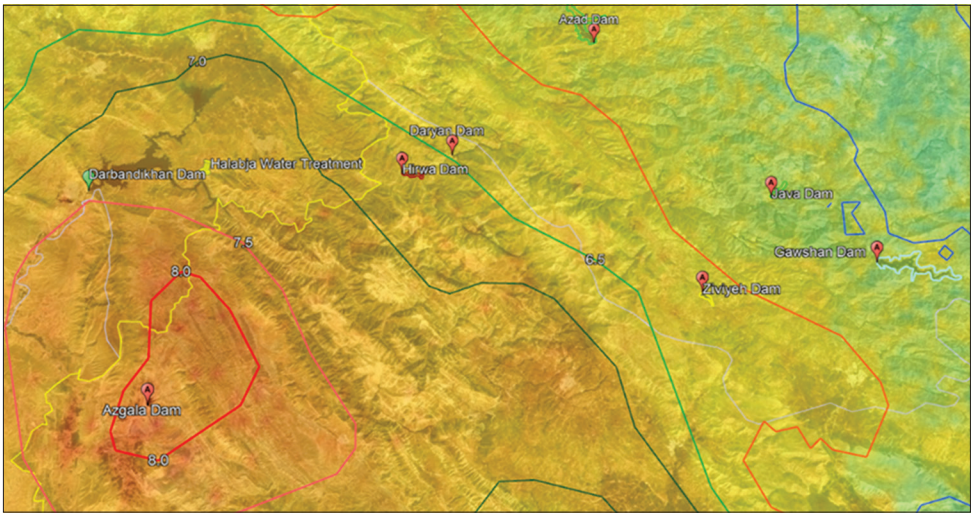


Fig. 8. Seismic map of the study area.

Darbandikhan Dam. The research focuses on flood wave routing, peak discharge propagation, and structural loading under various breach conditions. The outcomes are intended to support the development of EAPs, enhance cross-border risk communication, and inform adaptive reservoir management under extreme climate and seismic uncertainties.

2.3. HEC-RAS

This study employed HEC-RAS version 6.6, along with its updated 6.7 Beta release, to perform detailed hydrodynamic modeling. Both versions were sourced from the official platform of the U.S. Army Corps of Engineers (www.hec.usace.army.mil). HEC-RAS is a robust and extensively adopted hydraulic simulation tool capable of modeling both one-dimensional and two-dimensional unsteady flow conditions [36]. Its intuitive graphical interface supports a variety of functions, including dam breach analysis, sediment transport, water quality modeling, and the hydraulic design of in-channel and lateral infrastructure such as spillways, culverts, sluice gates, and weirs.

In this research, the breach scenario of Gawshan Dam was simulated using the 2D flow module of HEC-RAS, which provides a more spatially detailed view of flood wave behavior across complex terrain. While offering greater precision in modeling flood propagation dynamics, 2D simulations demand more advanced computational resources and meticulous preprocessing of topographic and boundary condition data. The technical specifications and structural properties of the modeled dams essential for defining breach parameters and flow hydraulics are summarized in Table 1.

2.4. Terrain Data

To support the hydraulic modeling in this study, a DEM with a spatial resolution of approximately 30 m (1 arc-second) was acquired from the USGS Earth Explorer portal (https://earthexplorer.usgs.gov). This elevation data, derived from the Shuttle Radar Topography Mission and made publicly available by the USGS on September 23, 2014, was processed within ArcMap after being converted into a Triangulated Irregular Network. All geospatial processing was conducted using the (WGS 1984 UTM Zone 38N) coordinate system to ensure consistency in horizontal referencing. Covering the geographic range between (34°N to 35°N) latitude and (45°E to 46°E) longitude, the dataset offers a reliable elevation profile suitable for detailed floodplain delineation and hydrodynamic modeling. Since the DEM predates the construction of the Gawshan Dam, a fully dynamic wave routing approach was adopted to simulate the reservoir's hydraulic response with greater accuracy than static storage-area methods, following guidelines from [37].

2.5. Breach Prediction Models

Analyzing dam failure scenarios is inherently complex due to the uncertain nature of key parameters such as breach location, breach dimensions, and the time required for breach formation, all of which significantly influence the resulting flood hydrograph [38]. To address these uncertainties, researchers have proposed numerous empirical and regression-based models aimed at estimating breach characteristics.

Among the most frequently utilized models are the equations introduced by Von Thun and Lawrence [39], which have been cited and adapted in several studies, including those by

Brunner [36], Froehlich [40], [41], Xu and Zhang [42], and Pierce [43]. These models have gained widespread application in dam breach assessments [38], [44], [45], especially due to their compatibility with diverse dam types and failure conditions. The Gawshan Dam, when evaluated for its structural profile and hydrological setting, demonstrates characteristics consistent with the assumptions underlying these regression models. Therefore, the breach equations from these established studies are considered reliable and suitable for application in the present simulation. Additional studies, such as those by Wahl [7], Wu *et al.* [8], also emphasize the importance of selecting breach parameter equations that reflect dam-specific conditions and site characteristics.

2.6. Assessment of the Flood Hazard

In flood modeling, water depth and flow velocity are regarded as the two most critical parameters influencing flood hazard evaluation. According to the classification developed by Mihu-Pintilie *et al.* [46], these variables serve as the primary indicators for assessing flood severity, as outlined in Table 2. This classification system has been adopted in the current study to support the analysis and categorization of flood hazard levels across the affected region.

2.7. Dam Failure Scenarios

To assess the downstream consequences of potential upstream dam failures on Darbandikhan Dam, four distinct failure scenarios were formulated and analyzed. These scenarios differ based on the mode of failure, initial reservoir levels, and the resulting cascading impacts on downstream hydraulic structures. The initial reservoir volumes and elevations used in the modeling are summarized below:

- Gawshan Dam (overtopping): 550 million m^3 plus an additional 80 million m^3 from overtopping inflow
- Gawshan Dam (piping failure): 550 million m^3
- Daryan Dam (overtopping): 316.3 million m^3
- Zhave Dam (at full capacity): 172 million m^3
- Hirwa Dam (in all cases): 12 million m^3.

Darbandikhan Dam – full condition: 2.55 billion m^3 at an elevation of 485 m, according to the most recent bathymetric survey conducted by the Darbandikhan Dam Directorate, the current live storage at the normal operation elevation of 485 m a.s.l. is 2,550 million m^3.

Darbandikhan Dam – semi-full condition: 2.05 billion m^3 at an elevation of 480 m. According to the operational rule curve Fig. 9, during wet seasons, the reservoir water level is maintained at 480 m, and based on the most recent survey conducted by the Dam Directorate, the corresponding storage at this elevation is 2,050 MCM.

The simulated failure scenarios are defined as follows:

- Scenario 1: An overtopping-induced failure at Gawshan Dam causes the collapse of Zhave, Daryan, and Hirwa dams, while Darbandikhan remains at full capacity (elevation 485 m)
- Scenario 2: Initiation of piping failure at Gawshan Dam triggers successive failures of Zhave, Daryan, and Hirwa dam, with Darbandikhan reservoir at full storage (elevation 485 m)
- Scenario 3: An overtopping event at Gawshan results in the failure of the same three downstream dams, while

TABLE 1: Gawshan, Zhave, Daryan, Hirwa, and Darbandikhan dams' properties

Items	Gawshan Dam	Zhave Dam	Daryan Dam	Hirwa Dam	Darbandikhan Dam
Dam type	Rock-fill dam	Concrete dam	Rockfill dam	Concrete dam	Rockfill dam
Height	123	86	169 m	45 m	128 m
Crest width	15 m	4.5 m	15 m	4.5 m	17 m
Reservoir capacity	550 million m^3	172 million m^3	316.3 million m^3	12 million m^3	2.55 billion m^3

TABLE 2: Classification flood hazard evaluation

Hazard level	Flood depth (m)	Flow velocity (m/s)	Hazard vulnerability classification	Description
Low	<0.5	0–2	H1	Flood does not pose hazard to people and on-foot evacuation is not difficult.
Medium	0.5–1	0–2	H2	Flood water poses hazard for infants and on-foot evacuation of adults becomes dicult; evacuation becomes more complicated.
High	1–2	0–2	H3	Flood depth can drown people; people may be safe inside their homes.
Crisis	2–5	0–2	H4	People are exposed to flood hazard even inside their homes and evacuate toward the roof of their homes is suggested.
Catastrophic	>5	2–4	H5	Built-up structures like homes may get covered by the flood; people may get drowned even if they evacuate toward the roof of their homes.

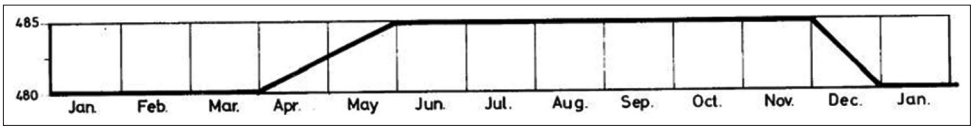


Fig. 9. Darbandikhan reservoir operational rule curve (adopted from Darbandikhan dam directorate).

Darbandikhan is operating below maximum storage (elevation 480 m)

- Scenario 4: Piping failure at Gawshan leads to the cascading failure of Zhave, Daryan, and Hirwa dams, under semi-full conditions at Darbandikhan (elevation 480 m).

These scenarios aim to capture the variation in downstream flood dynamics and structural response under different breach mechanisms and storage levels, thereby supporting the development of robust emergency response plans and reservoir operation strategies.

3. RESULTS AND DISCUSSION

3.1. Scenario Characterization

To analyze the cascading dam failure risks within the study area, four initial breach scenarios were constructed involving Gawshan, Zhave (also referred to as Java), Daryan, and Hirwa dams. These scenarios were formulated to capture two key failure modes: Overtopping and piping across two storage states: Full supply level (FSL) and semi-full level. Among these, two representative scenarios were selected for detailed hydraulic evaluation based on their severity, relevance to real-world dam management practices, and their potential to cause extreme downstream flooding and instability at Darbandikhan Dam.

- Scenario 1: Overtopping Cascade at Full Capacity

Sequence: Gawshan (overtopping) → Zhave (overtopping) → Daryan (overtopping) → Hirwa (overtopping) → Darbandikhan (at FSL: 2.55 billion m^3, Water Surface Elevation (WSE) = 485 m)

This scenario reflects the worst-case cascade, with all upstream dams breaching due to overtopping while Darbandikhan remains at full reservoir capacity.

It is designed to examine peak inflow volumes, minimum flood arrival times, and the structural load limits of the downstream dam.

- Scenario 2: Mixed Failure Under Full Storage

Sequence: Gawshan (piping) → Zhave (overtopping) → Daryan (overtopping) → Hirwa (overtopping) → Darbandikhan (at FSL: 2.55 billion m^3, WSE = 485 m)

This scenario combines internal erosion at Gawshan Dam with overtopping failures at the downstream dams. The objective is to assess how variation in initial breach mechanisms influences the shape of the flood hydrograph, wave travel time, and the resulting stress imposed on the Darbandikhan reservoir and dam body.

3.2. Peak WSE Analysis

Table 2 presents a comparison of peak WSEs at six strategically important locations during flood propagation: The Iraq–Iran border, marking the point where breach outflows begin entering Darbandikhan Reservoir. The immediate upstream zone of Darbandikhan Dam reflects the attenuated wave impact after storage buffering. By analyzing the differences in peak WSE at these locations, the reservoir's capacity to absorb and dampen incoming flood waves can be evaluated. Higher values recorded at the border indicate the initial magnitude of the flood wave, whereas lower elevations near the dam body reveal how much energy and volume were mitigated by the reservoir's geometry and storage.

This comparative analysis is essential for determining:

Emergency drawdown requirements, design thresholds for structural reinforcement, and response trigger in the EAP under varying failure conditions.

3.3. WSE Rise and Flood Arrival Dynamics

A comparison between the breach hydrographs produced for piping and overtopping failures at Gawshan Dam revealed only slight variations in both peak discharge and total flood volume. As a result, Scenario 1, which simulates the overtopping failure, has been selected for further hydraulic analysis due to its potential to generate the most critical downstream flood impacts and heightened risk of structural stress on Darbandikhan Dam.

In Scenario 1, the cascading failure sequence begins with the overtopping-induced breach of Gawshan Dam, which occurs while all upstream reservoirs Gawshan, Zhave (Java), Daryan, and Hirwa are operating at full storage capacity. The downstream Darbandikhan Reservoir is similarly assumed to be at its maximum level, with an initial WSE of 485.0 m.

Following the initial breach, the flood wave of Gawshan Reservoir after failure propagates rapidly through the cascade system. It arrives at Zhave Dam approximately 52 min after the failure of Gawshan Dam, inducing structural failure at Zhave 67 min post Gawshan breach. The compound flood surge then continues its path, reaching Daryan Dam at 172 min and causing overtopping failure at 187 min.

The wave subsequently reaches Hirwa Dam at 190 min, where it also results in overtopping-induced failure due to the dam's limited retention capacity. The cumulative floodwaters from the four failed dams, Gawshan, Zhave, Daryan, and Hirwa then traverse the transboundary river system, arriving at the Iraqi border approximately 270 min after the initial breach event.

On entering Iraq, the flood wave impacts the Darbandikhan Reservoir within 280 min, where rapid inflow begins elevating the reservoir level. Modeling results indicate that the WSE at Darbandikhan rises from the initial 485.0 m to 495.5 m by 412 min after Gawshan Dam's failure. This significant elevation surpasses the dam crest level (495.0 m), resulting in overtopping by approximately 0.5 m. Such an event poses a critical threat to the structural integrity of the dam, with the potential to trigger a catastrophic downstream flood wave unless immediate emergency response and reservoir drawdown measures are enacted. This chain-reaction failure scenario highlights the severe transboundary hydrodynamic risks posed by overtopping-induced cascading dam breaches and underscores the urgent need for coordinated reservoir operation strategies and real-time flood management protocols.

3.4. Submergence of Rural Settlements and Agricultural Lands Due to Cascading Dam Failures from Gawshan to Hirwa in Iran

The cascading overtopping scenario results in severe upstream flooding, particularly in the valleys located above Daryan and Hirwa dams. The initial failure of Gawshan dam due to overtopping generates a high-velocity flood wave that rapidly propagates through the narrow river valleys, overwhelming the storage capacity of Zhave dam and causing its subsequent failure. The combined discharge from both dams substantially elevates upstream water surface levels and increases flow velocities throughout the interconnected catchments. The flood hazard classification of these impacted villages is shown in Table 3, while the downstream flood depth and hazard vulnerability of affected villages and roads are detailed in Table 4.

TABLE 3: WSE results for each scenarios

Location	Scenario 1 WSE (m) Overtopping	Scenario 2 WSE (m) Pipping
Downstream of Gawshan Dam	1470.57	1466.76
Downstream of Zhave Dam	1290.21	1290.0
Downstream of Daryan Dam	759.8	759.6
Downstream of Hirwa Dam	704.19	704.01
At the Iran-Iraq Border	518.60	516.53
Near Darbandikhan Dam body	495.50	494.66

Hydraulic modeling indicates that over 10 upstream villages situated between Gawshan, Zhave, Daryan, and Hirwa reservoirs are critically affected by the flood wave. In particular, the stretch between Gawshan dam and Hirwa dam located upstream of both Hirwa and Daryan dams contains a concentration of rural settlements where over 2,000 houses are expected to be submerged or severely damaged by the cascading flood wave. Notably, Kwana and Faqia Sleman and Taza Abad Sarpil and Askaran and Palangan which is located and Dewaznaw and Jolandah and Hirwa Villages, Palangan village located just upstream of Daryan Dam Fig. 10, Palangan village is projected to be completely submerged, with approximately 194 houses and an estimated 821 residents inundated under flood depths exceeding 19 m., widespread inundation also impacts surrounding agricultural lands, Dewaznaw and Jolandah village affected by flood submerged under flood water over 30 m, as shown in Fig. 11. Due to the steep topography and narrow shape of the river valleys, the velocity of floodwaters accelerates rapidly during propagation.

In addition, the floodwaters submerge key regional roads that connect upstream villages, effectively severing transport and communication links. This loss of connectivity poses major challenges to emergency response, aid delivery, and long-term recovery, further compounding the social and economic vulnerability of the impacted population.

3.5. Impact on Downstream Areas in Iraq

Beyond Hirwa Dam, the cascading flood wave crosses into northeastern Iraq, triggering severe transboundary hydrodynamic consequences. HEC-RAS simulation results indicate that more than 12 Iraqi villages fall within the projected inundation zone. These predominantly agricultural communities face a high risk of socioeconomic disruption, infrastructure damage, and population displacement. The flood also impacts over 80 km^2 of farmland, particularly in the low-lying plains near Halabja city and across the Said Sadiq and Sirwan districts, posing a serious threat to food security and rural livelihoods in this agriculturally vital region.

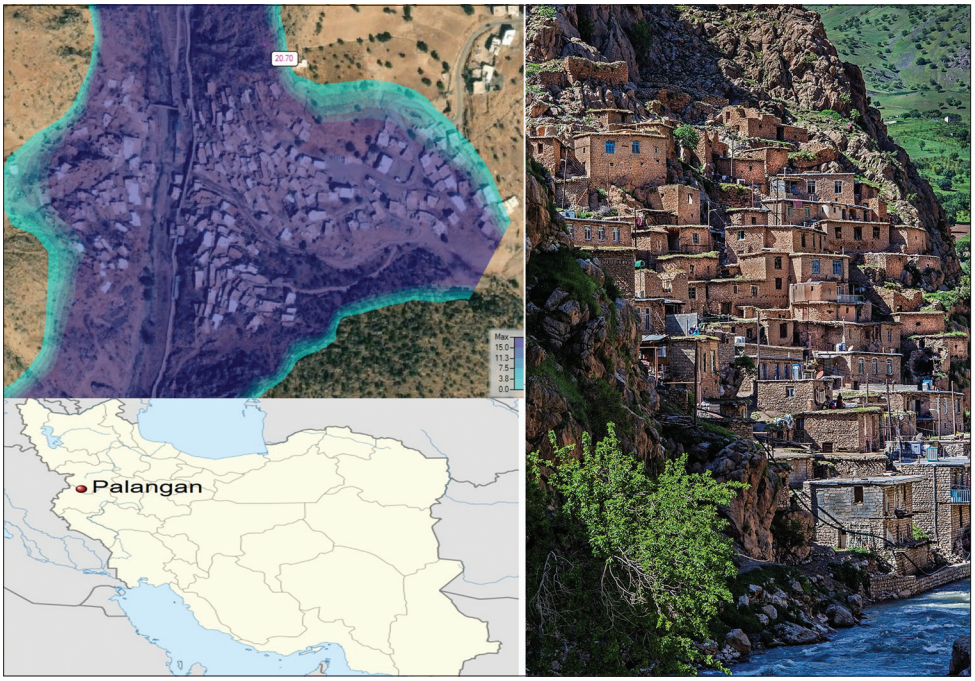


Fig. 10. Palangan Village after flood, submerged.

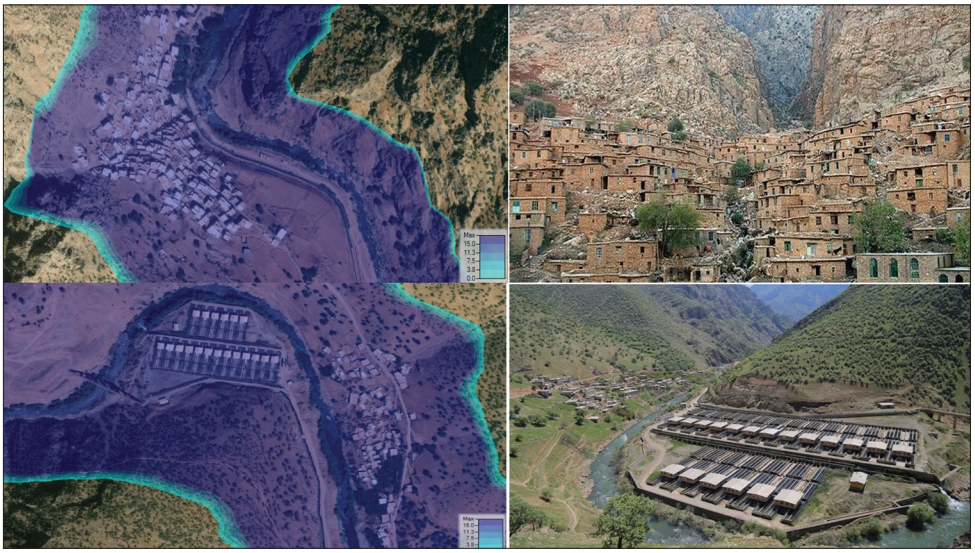


Fig. 11. Dewaznaw and Jolandah Village after flood, submerged.

Critical infrastructure is similarly threatened. The newly constructed highway linking Halabja to surrounding districts is projected to be overtopped by floodwaters, cutting off essential transportation routes and severely hindering emergency response and post-disaster recovery efforts.

Under the Gawshan Dam overtopping scenario, the situation deteriorates significantly. Following the sequential collapse of upstream dams including Zhave, Daryan, and Hirwa, the flood inflow to Darbandikhan Reservoir intensifies rapidly. As a result, the reservoir level rises from 485.0 m to 495.5 m, exceeding the dam's crest elevation of 495.0 m. Despite the full operation of spillway gates (crest at 485.0 m) and bottom outlets, the discharge capacity proves insufficient. This leads to uncontrolled overtopping and the eventual structural failure of Darbandikhan Dam.

The catastrophic breach of Darbandikhan Dam generates an overwhelming downstream flood wave, exacerbating hydrodynamic risks along the lower Sirwan River corridor. Several populated areas including Kampi Xwarw Fig. 12, Azadi Fig. 13, Eisay Fig. 14, Karashala, Hamid Awa, Panamini Mala Sleman, parts of Kalar Fig. 15, and Kulajo Fig. 16 are projected to be inundated. These regions face immediate threats to life, widespread infrastructure destruction, and long-term disruption to community stability and regional development.

4. DISCUSSION

The cascading failure of dams within a transboundary river basin poses a unique and severe hydrodynamic threat, as evidenced by the simulation results of this study.

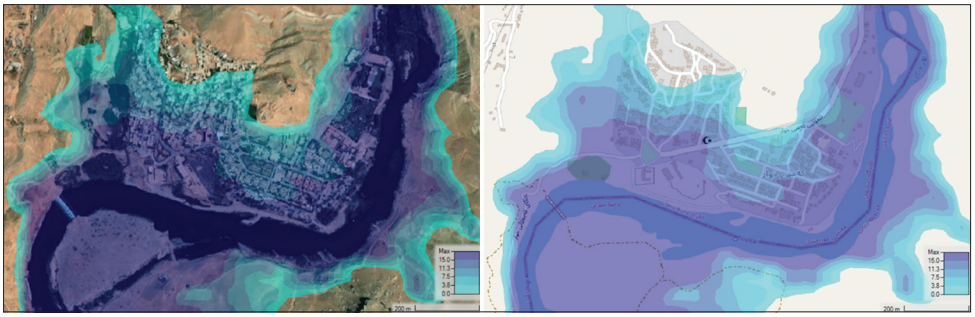


Fig. 12. Kampi Xwarw.

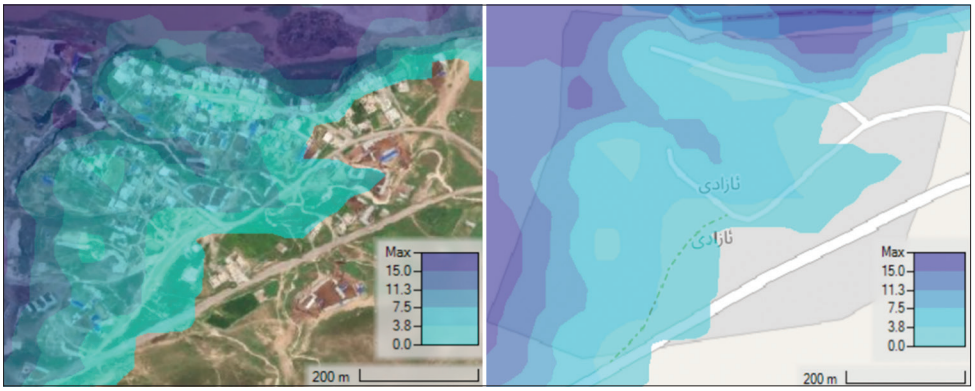


Fig. 13. Azadi.

Fig. 14. Eisay.

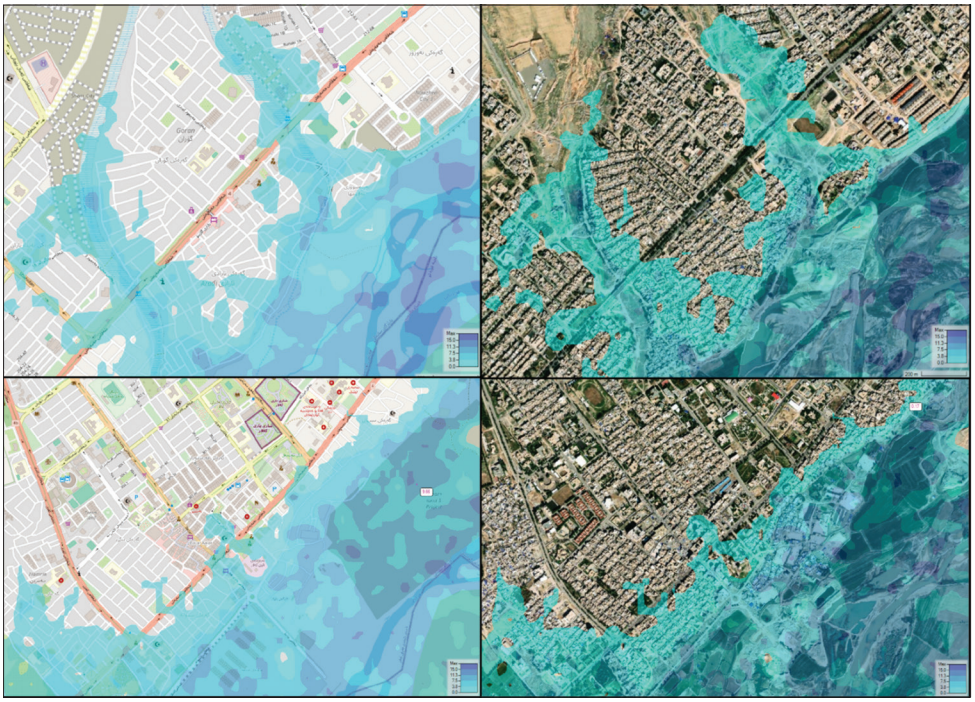


Fig. 15. Kalar.

TABLE 4: Flood depth and hazard vulnerability classification of affected villages and roads in the downstream region

No.	Village name	Depth of water	Hazard vulnerability classification	Description
1	*Kwana*	Over 20 m	H5	Unsafe for vehicles and people. All building types are vulnerable to structural damage.
2	*Faqia Sleman*	Over 16 m	H5	
3	*Taza Abad Sarpil*	Over 25 m	H5	
4	*Askaran*	Over 6 m	H5	
5	*Sanandaj – Kamiaran Road*	From 17 m to 40 m	H5	
	Mariwan – Kamiaran Road	From 10 m to 78 m	H5	
6	*Palangan*	Over 19.60 m	H5	
7	*Dewaznaw*	Over 30 m	H5	
8	*Jolandah*	Over 50 m	H5	
9	*Dolamrz*	Over 55.40 m	H5	
10	*Zum*	Over 9.0 m	H5	
11	*Kani pl*	Over 12 m	H5	
12	*Pawa – Nowsud Road*	Over 40 m	H5	
13	*Hirwa*	Over 30 m	H5	

The sequential breach of Gawshan, Zhave, Daryan, and Hirwa dams triggered by overtopping or piping produces a compound flood wave that intensifies as it descends through the Sirwan River system. Notably, this flood wave reaches Darbandikhan Dam with minimal attenuation, despite the significant distances involved. The modeling outcomes clearly indicate that the proximity of dams, their full operational levels, and the confined nature of the upstream valleys collectively reduce the capacity for wave dispersion, resulting in the convergence of peak flows at critical downstream locations.

The hydraulic behavior of Darbandikhan Reservoir under these scenarios reveals significant vulnerability. In the most extreme case Scenario 1, which assumes full upstream and downstream reservoir levels the WSE at Darbandikhan rises from 485.0 m to 495.5 m within less than seven hours. This exceeds the dam's crest elevation, leading to overtopping and potential structural failure. Even under less severe breach mechanisms, such as the piping-induced scenario

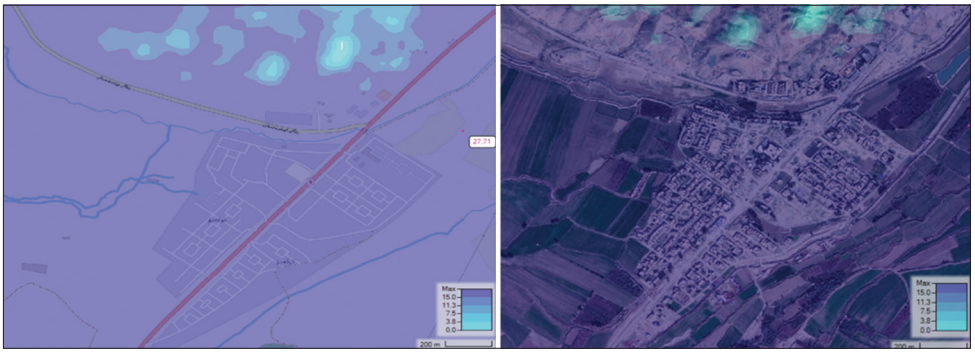


Fig. 16. Kulajo.

(Scenario 2), the WSE still reaches 494.66 m, dangerously close to the overtopping threshold. These findings suggest that the reservoir's current operational rule curve does not provide sufficient buffer capacity for high-magnitude, compound inflow events.

The limitations of Darbandikhan Dam's emergency discharge infrastructure are also evident. Despite full engagement of bottom outlets and spillway gates, the inflow from the cascading breaches overwhelms its outflow capacity. This indicates that the dam's design parameters, while effective under isolated failure or high rainfall conditions, are inadequate in the context of transboundary, compound flood events. The modeling indicates that these systems were never intended to absorb flood waves of this scale and velocity, which originate from multiple upstream sources in rapid succession.

Beyond the hydraulic implications, the human and ecological impacts are profound. In the Iranian portion of the basin, over 10 villages are directly affected by upstream flooding, with flood depths exceeding 30 m in several locations. Entire communities including Palangan, Dewaznaw, and Jolandah are projected to be submerged. The steep terrain and poorly connected road systems limit evacuation options and emergency access. In addition to residential damage, extensive agricultural areas and lifeline infrastructure such as the Mariwan–Kamyaran road are inundated, isolating populations and interrupting essential services. The flood wave's transboundary journey into Iraq further exacerbates the crisis. On arrival, it submerges more than 80 km^2 of agricultural land and affects villages and infrastructure near Halabja, Said Sadiq, and Kalar. The secondary failure of Darbandikhan Dam leads to further destruction along the lower Sirwan River. Urban centers and transportation routes downstream are not designed to withstand this scale of hydraulic load, exposing the systemic fragility of water infrastructure and civil protection systems in both nations.

The overall findings from this study reinforce the importance of cooperative, anticipatory planning. The cascading failure risk in the Sirwan Basin is not a theoretical construct it is a plausible, data-supported scenario that demands immediate attention. Both structural adaptations (e.g., auxiliary spillways and crest raising) and non-structural measures (e.g., early warning systems, revised operational protocols, and bilateral emergency coordination) are required to prevent catastrophic losses in the event of extreme natural or human-induced triggers.

5. CONCLUSION

This study provides the first integrated hydrodynamic simulation of cascading overtopping and piping failures across a transboundary dam network from Gawshan Dam in Iran to Darbandikhan Dam in Iraq. The results reveal that under full reservoir conditions, a sequential overtopping cascade leads to catastrophic consequences downstream, including the overtopping and potential breach of Darbandikhan Dam. Scenario 1, involving maximum water storage and overtopping-driven failures, generated the most critical conditions, with rapid flood wave propagation and insufficient structural mitigation capacity. Both upstream Iranian and downstream Iraqi communities face substantial human, agricultural, and infrastructural losses under these scenarios. These findings underscore the urgency for coordinated risk assessments, early warning systems, and bilateral flood risk governance.

6. ACKNOWLEDGMENT

The authors express greatest appreciation to the Darbandikhan Dam Directorate and the supervisory committee for

their significant assistance in supplying crucial data and information necessary to this study.

7. FUNDING

The authors declare that this research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

8. DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author on reasonable request.

9. CONSENT TO PARTICIPATE

Not applicable. This review article does not involve human participants or personal data requiring consent.

10. CONSENT TO PUBLISH

All authors have read and approved the final version of the manuscript and consent to its submission and publication in UHD journal of science and technology.

11. ETHICS

This article does not contain any studies with human participants or animals performed by any of the authors.

12. COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

[1] A. Bharath, A. V. Shivapur, C. G. Hiremath and R. Maddamsetty. "Dam break analysis using HEC-RAS and HEC-GeoRAS: A case study of Hidkal dam, Karnataka State, India". *Environmental Challenges*, vol. 5, p. 100401, 2021.

[2] M. Zagonjolli. "*Dam Break Modelling, Risk Assessment and Uncertainty Analysis for Flood Mitigation: IHE-PhD Thesis, Unesco-IHE, Delft, the Netherlands*". Routledge and CRC Press. Available from: https://www.routledge.com/dam/break/modelling/risk/assessment/and/uncertainty/analysis/for/flood/mitigation/ihe/phd/thesis/unesco/ihe/delft/the/netherlands/zagonjolli/p/book/9780415455947 [Last accessed on 2025 Jul 20].

[3] J. E. Costa. "*(PDF) Floods from Dam Failures (1985). J.E. Costa. 388 Citations*". SciSpace - Paper. Available from: https://scispace.com/papers/floods/from/dam/failures/2g8xboybj8 [Last accessed on 2025 Jul 20].

[4] D. L. Fread. "*DAMBRK: The NWS Dam-Break Flood Forecasting Model*". Hydrologic Research Laboratory, National Weather Service, NOAA, USA, 1984.

[5] L. M. Zhang, Y. Xu and J. S. Jia. "Analysis of earth dam failures: A database approach". *Georisk Assessment and Management of Risk for Engineered Systems and Geohazards*, vol. 3, no. 3, pp. 184-189, 2009.

[6] Y. Xiong. "A dam break analysis using HEC-RAS". *Journal of Water Resource and Protection*, vol. 3, no. 6, 2011.

[7] T. Wahl. "*Predicting Embankment Dam Breach Parameters - A Needs Assessment*". U. S. Bureau of Reclamation, United States, 1997.

[8] W. Wu, M. S. Altinakar, M. Al-Riffai, N. Bergman, S. F. Bradford, Z. Cao, Q. J. Chen, G. Constantinescu, J. Duan, D. M. Gee, B. P. Greimann, G. Hanson, Z. He, P. Hegedus.,… & L. M. Zhang. "Earthen embankment breaching". *Journal of Hydraulic Engineering*, vol. 137, no. 12, pp. 1549-1564, 2011.

[9] M. M. A. Mohamed. "Overtopping breach peak outflow approximation of embankment dam by using Monte Carlo method". *Beni-Suef University Journal of Basic and Applied Sciences*, vol. 7, no. 4, pp. 724-732, 2018.

[10] M. Sowiński. "An uncertainty analysis of the flood-stage upstream from a bridge". *Water Science and Technology*, vol. 53, no. 1, pp. 77-84, 2006.

[11] B. Wang, T. Zhang, Q. Zhou, C. Wu, Y. Chen and P. Wu. "A case study of the Tangjiashan landslide dam-break". *Journal of Hydrodynamics*, vol. 27, no. 2, 2015, pp. 223-233.

[12] DHI. "*MIKE 11 - River and Channel Modelling*". DHI, Denmark, 2017.

[13] USACE. "*National Inventory of Dams*." Available from: https://geo.usace.army.mil/pgis/p?p14397:5:0:no [Last accessed on 2025 Sep 03].

[14] S. Yu, Q. Zhang, Z. Chen, J. Hao, L. Wang, P. Li and Q. Zhong. "Study of the sheyuegou dam breach - experience with the post-failure investigation and back analysis". *Engineering Failure Analysis*, vol. 125, p. 105441, 2021.

[15] M. J. Butt, M. Umar and R. Qamar. "Landslide dam and subsequent dam-break flood estimation using HEC-RAS model in Northern Pakistan". *Natural Hazards*, vol. 65, no. 1, pp. 241-254, 2013.

[16] I. Haltas, G. Tayfur and S. Elci. "Two-dimensional numerical modeling of flood wave propagation in an urban area due to Ürkmez dam-break, İzmir, Turkey". *Natural Hazards*, vol. 81, no. 3, pp. 2103-2119, 2016.

[17] O. S. Balogun and H. O. Ganiyu. "*(PDF) Study and Analysis of Asa River Hypothetical Dam Break Using Hec-Ras*". ResearchGate. Available from: https://www.researchgate.net/publication/352045561_study_and_analysis_of_asa_river_hypothetical_dam_break_using_hec-ras [Last accessed on 2025 May 28].

[18] M. M. Joshi and S. S. Shahapure. "Two dimensional dam break flow study using HEC-RAS for Ujjani dam". *International Journal of Engineering and Technology*, vol. 9, no. 4, pp. 2923-2928, 2017.

[19] P. Sharma and S. Mujumdar. "*(PDF) Dam Break Analysis using HEC-RAS and HEC-GeoRAS - A Case Study of Ajwa Reservoir*." ResearchGate. Available from: https://www.researchgate.net/publication/314273997_dam_break_analysis_using_hec/ras_and_hec-georas_-_a_case_study_of_ajwa_reservoir [Last accessed on 2025 May 28].

[20] B. P. Yakti, M. B. Adityawan, M. Farid, Y. Suryadi, J. Nugroho and I. K. Hadihardaja. "*(PDF) 2D Modeling of Flood Propagation due to the Failure of Way Ela Natural Dam*". ResearchGate. Available from: https://www.researchgate.net/publication/322638804_2d_modeling_of_flood_propagation_due_to_the_failure_of_way_ela_natural_dam [Last accessed on 2025 May 28].

[21] L. M. Albu, A. Enea, M. Iosub and I. G. Breaban. "Dam breach size comparison for flood simulations. A HEC-RAS based, GIS approach for Drăcșani Lake, Sitna River, Romania". *Water*, vol. 12, p. 1090, 2020.

[22] L. M. Albu, A. Enea, C. C. Stoleriu and L. Niacşu. "(PDF) G.I.S. Implementation on Dam-Break Flood Vulnerability Analysis - a Case Study of Cătămărăști Dam, Botoșani, Romania". ResearchGate. Available from: https://www.researchgate.net/publication/334063213_gis_implementation_on_dam/break_flood_vulnerability_analysis/_a_case_study_of_catamarasti_dam_botosani_romania [Last accessed on 2025 May 28].

[23] S. Kilania and B. R. Chahar. "*A Dam Break Analysis Using HEC-RAS*". Proceedings, 2019. Available from: https://ascelibrary.org/doi/10.1061/9780784482353.036 [Last accessed on 2025 May 28].

[24] T. Mulia Aribawa and Farid M. "*Assessment of Flood Propagation Due to Several Dams Break in Banten Province*, 2021". Available from: https://www.academia.edu/99747382/assessment_of_flood_propagation_due_to_several_dams_break_in_banten_province [Last accessed on 2025 May 28].

[25] O. Azeez. "*Dam Break Analysis and Flood Disaster Simulation in arid Urban Environment: The Um Al-Khair Dam Case Study*". Jeddah, Saudi Arabia, 2020. Available from: https://www.academia.edu/85030851/dam_break_analysis_and_flood_disaster_simulation_in_arid_urban_environment_the_um_al_khair_dam_case_study_jeddah_saudi_arabia [Last accessed on 2025 Jul 20].

[26] V. B. Ghindaoanu, E. Hutanu and A. Urzica. "*The Gis Modeling of the Terrain Favorability for the Placement of Constructions in the Areas with Hydro-Geomorphological Risk*". [Conference Paper], pp. 21-30, 2019.

[27] H. Qi and M. S. Altinakar. "*GIS-Based Decision Support System for Dam Break Flood Management under Uncertainty with Two-Dimensional Numerical Simulations*". 2012. Available from: https://ascelibrary.org/doi/10.1061/%28asce%29wr.1943-5452.0000192 [Last accessed on 2025 May 28].

[28] K. Zaidan, O. S. Q. Yousif, Y. Alshkane, A. Khani and S. K. Hama. "*Performance of Darbandikhan Dam during a Major Earthquake on November 12, 2017*." In: 3rd Meeting of EWG Dams and Earthquakes an International Symposium, 2019.

[29] M. Rashidi and S. M. Haeri. "Evaluation of behaviors of earth and rockfill dams during construction and initial impounding using instrumentation data and numerical modeling". *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 9, no. 4, pp. 709-725, 2017.

[30] Talashgaran co., 2015. Available from: https://talashgaran.co/fa/project-details/148/[Last accessed on 2025 May 01].

[31] M. Tatar, A. Papzan and M. Ahmadvand. "Farmers' lived experiences of water crisis: Lessons learned from a phenomenology study". *Environmental Management*, vol. 72, pp. 1204-1215, 2023.

[32] J. Jamshidi and M. Shourian. "Hedging rules-based optimal reservoir operation using bat algorithm". *Water Resources Management*, vol. 33, no. 13, pp. 4525-4538, 2019.

[33] Amini, A., and Goftari, Z. "Assessment of the impacts of Zhawa Dam on adjacent rural areas: The case of Sanandaj County". *Geography and Regional Developmen*t, vol. 15, pp. 213-236, 2017.

[34] K. Chomani and T. Bijnens. "*THE Impact of the Daryan Dam on the Kurdistan Region of Iraq*". Fertiliser Association of India, New Delhi, 2016.

[35] A. Jamshidi, M. Sedaghatnia and K. Mirbiyk Sabzevari. "The role of structural factors in directing methane and carbon dioxide to the Nosud water transmission tunnel (Azgeleh - Northwest of Kermanshah)". *Journal of New Findings in Applied Geology*, vol. 15, pp. 28-38, 2021.

[36] G. W. Brunner. "*HEC-RAS River Analysis System*". Ver. 5.0. Hydraulic Reference Manual, Davis, CA, 2016.

[37] Brunner. "*HEC-RAS 5.0 Users Manual. PDF Computing Computing and Information Technology*". Available from: https://www.scribd.com/document/339335302/hec/ras/5/0/users/manual [Last accessed on 2025 May 13].

[38] D. M. Gee. "*Dam Breach Modeling with HEC-RAS using Embankment Erosion Process Models1347-1356"*, 2012. Available from: https://ascelibrary.org/doi/10.1061/41114%28371%29144 [Last accessed on 2025 May 01].

[39] Von Thun and J. Lawrence. "*Guidance on Breach Parameters. Internal Memorandum, U.S. Department of the Interior, Bureau of Reclamation, Denver, 17*". Vol. 17. Internal Memorandum, U.S. Department of the Interior, Bureau of Reclamation, Denver, 1990.

[40] D. Froehlich. "Predicting peak discharge from gradually breached embankment dam". *Journal of Hydrologic Engineering*, vol. 21, p. 04016041, 2016b.

[41] D. Froehlich. "Embankment dam breach parameters and their uncertainties". *Journal of Hydraulic Engineering*, vol. 134, p. 1708, 2008.

[42] Y. Xu and L. M. Zhang. "*Breaching Parameters for Earth and Rockfill Dams*". Available from: https://ascelibrary.org/doi/10.1061/%28asce%29gt.1943-5606.0000162 [Last accessed on 2025 May 01].

[43] M. Pierce, C. Thornton and S. Abt. "Predicting peak outflow from breached embankment dams". *Journal of Hydrologic Engineering*, vol. 15, pp. 338-349, 2010.

[44] K. Zaidan Abdulrahman. "Case study of the Chaq-Chaq dam failure: Parameter estimation and evaluation of dam breach prediction models". *Journal of Engineering Research and Applications*, vol. 4, pp. 109-116, 2014.

[45] E. Psomiadis, L. Tomanis, A. Kavvadias, K. Soulis, N. Charizopoulos and S. Michas. "Potential dam breach analysis and flood wave risk assessment using HEC-RAS and remote sensing data: A multicriteria approach". *Water*, vol. 13, p. 364, 2021.

[46] A. Mihu-Pintilie, C. I. Cîmpianu, C. C. Stoleriu, M. N. Pérez and L. E. Paveluc. "*Using High-Density LiDAR Data and 2D Streamflow Hydraulic Modeling to Improve Urban Flood Hazard Maps: A HEC-RAS Multi-Scenario Approach*". Available from: https://www.mdpi.com/2073/4441/11/9/1832 [Last accessed on 2025 May 30].

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

An Efficient Hybrid Framework U-Net-based deep learning Technique of Early Detection of Pulmonary Nodules using LIDC-IDRI Dataset

Hataw Jalal Mohammed[1], Fawzi Abdul Azeez Salih[2], Tofiq Ahmed Tofiq[3], Shaniar Tahir Mohammed[3]

[1]Charmo Center for Research, Training, and Consultancy, University of Charmo, Sulaimani, Iraq, [2]Department of Computer Science, College of Basic Education, University of Sulaimani, Sulaimani, Iraq, [3]Department of Computer Science, College of Science, University of Sulaimani, Sulaimani, Iraq

ABSTRACT

Radiologists still face difficulties and mistakes while screening lung computed tomography (CT) images for pulmonary nodules, particularly small and inconspicuous malignant lesions. Frequent radiation exposure, the complexity of radiomic characteristics in low-dose CT scans, and the high cost of imaging therapy are some of the challenges. Our technique proposed a novel automated computer-aided diagnostic technique to address these issues by increasing the accuracy of early lung nodule diagnosis. We suggested a four-step technique that involves (1) a preprocessing step consisting of contrast-limited adaptive histogram equalization to refine the contrast of contribution inputs, followed by extracting and combining texture and shape features in parallel using a gray level co-occurrence matrix for the first features and region of interest (ROI) properties for the second. In addition, we suggested a hybrid U-Net-based deep learning architecture for categorization that successfully blends automatically learned features with manually created features. This integration improves the precision and resilience of pulmonary nodule classification by utilizing the convolutional neural networks (CNN)'s capacity to capture spatial hierarchies. We implemented our proposed technique on the overtly accessible Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) dataset, employing the Python programming language for the implementation. Experimental results confirmed that our technique attained segmentation and classification results of 95.35% accuracy, 95.33% sensitivity, 94.23% specificity, and 95.44% AUC rates, outperforming several state-of-the-art methods. This high-performance approach offers a reliable solution for early detection, potentially reducing lung cancer mortality rates through timely diagnosis.

Index Terms: U-Net, Convolutional Neural Networks, Pulmonary Nodule Detection, Lung Image Database Consortium

1. INTRODUCTION

Pulmonary nodules cancer in its primary phases is not evident with obvious symptoms, which impede early diagnosis [1]. Published statistics indicate that lung nodule cancer is an important role of the leading reasons of destruction of the living in the world, accounting for 18.75% of wholly cancer deaths [2]. Furthermore, according to a study conducted in 2023, more than 238,000 people in the United States are expected to be exposed and diagnosed with lung cancer, and nearly 127,000 lung cancer patients will perish [3] and according to a study conducted in the following year, the facts indicated that about (50%) of smokers suffer from a lung nodule, and (25%) suffer from two or more [4]. Since pulmonary nodules are small spots in the lung tissue

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp266-275

E-ISSN: 2521-4217

P-ISSN: 2521-4209

Copyright © 2025 Mohammed, *et al*. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Hataw Jalal Mohammed, Charmo Center for Research, Training, and Consultancy, University of Charmo, Sulaimani, Iraq. Email: hataw.jalal@chu.edu.iq

Received: 15-05-2025 Accepted: 25-09-2025 Published: 01-11-2025

that are solider than the surrounding lung, they are usually sphere-shaped or oval, furthermost of these nodules are smaller than 1cm in size, while they can grow up to 3cm in size, trials by organizations such as the National Lung Screening Trial (NLST) have indicated the high sensitivity of (CT) which can help stakeholders detect lung cancer early and thus reduce lung cancer deaths by (20%) [2]. Since it can effectively help doctors recognize the size, shape, and nature of pulmonary nodules more accurately by making it easier to identify important features present in each imaging slice such as texture and shape [4], the early diagnosis of nodules is very necessary for the controlling of lung cancer affected [5]. Fig. 1 illustrates how pulmonary nodules from a well-known open-source image dataset which they may be spherical opacities or irregular lung lesions ranging in size (3–30) mm and can be hermitic or multiple.

Nodules usually have various kinds and morphologies [6] such as (with/without) vasculature attachment (surrounded/not surrounded) by vasculature and lung membrane, (Far/near) to the diaphragm, (regular/irregular) formed, (low/high) contrast nodule, (has/has not) vasculature attachment, (single/multiple) in number, boundaries (smooth/rough), density (solid/semi solid/ground glass) nodule, and position (specified/unspecified)) as shown in Fig. 2.

Simultaneously, computer-aided diagnostic (CAD) is a typical technique as a domain of digital image processing for recognizing and detecting the many-sidedness of diseases, as well as different kinds of cancer. Moreover, it's systems also provide appropriate support for clinical decision-making [7]. The majority of pulmonary nodule detection systems perform two tasks: Lung nodule detection techniques achieve two missions: (1) Detect candidate nodules and (2) reduce false positives (FPs). Any nodule not determined at initial stage will no far be retrieved later. Thus, the first stage typically permits to be actual very sensitivity in being evident with the rate of numerous FP cases. Consequently, the second phase aims to reduce the number of FPs [8]. In recent years, CAD systems have mainly depended on machine learning (ML) and deep learning (DL) technologies [9], which introduce a major role as a second interpreter in the interpretation of CT images [4]. The aim of choosing DL is to eliminate the major drawbacks of using traditional ML methods, especially when it depends on feature extraction [9]. In spite of DL requiring large amounts of computing power and data for training, it has been widely researched due to the accuracy of its estimations [7]. DL has proven itself as a major breakthrough not only in the fields of medicine but also in another numerous scientific fields, making far-reaching changes in the way we learned and interpreted complex data forms. Enclosed by the dominance of medical imaging, mainly in the analysis of CT lungs, there is an eminent tendency where DL models are being combined for nodule segmentation. These models involve numerous network architectures such as convolutional neural networks [CNNs] which are the furthermost broadly used, VGG16, ResNet [10], dual-branch residual networks, Mask R-CNN, and U-Net [3].

A full decade has passed since the invention of U-Net and its assistance in the field of medical image processing [11] as a semantic segmentation network built on a fully CNN [12]. Its manner was developed as a product of the development and enhancement of CNNs. In addition of allowing computing the measurable attribute of damaged (texture, perimeter,

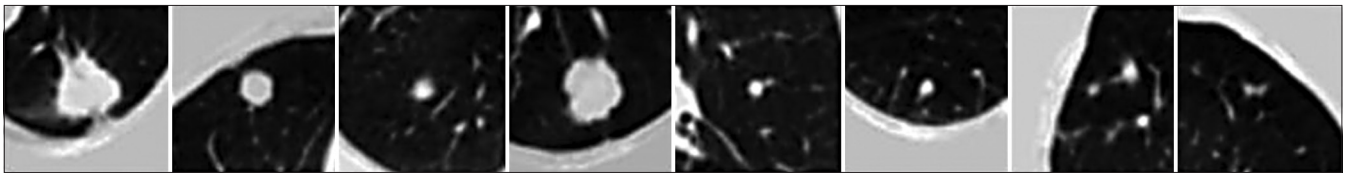


Fig. 1. Computed tomography image examples of the dataset (lung image database consortium).

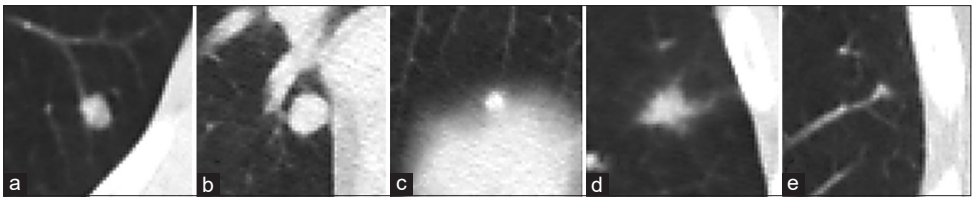


Fig. 2. Various kinds of pulmonary nodules. (a) With vasculature attachment. (b) Between vasculature and lung membrane. (c) Small and near the diaphragm. (d) Irregular shape with low contrast. (e) Has not vasculature.

direction, area, etc.), which facilitates the identification of the nature of the damage and improves and reduces the treatments required to exclude it [11].

The basic U-Net neural network structure is a typical convolutional network construction consists of repeated application of convolutions, each followed by a rectified linear unit (ReLU) and a max pooling, gives a symmetric U-shaped architecture as indicated in Fig. 3, as noted that it consists of dual parts: The left is the compression (encoder) and the right is the expansion (decoder). Throughout the reduction, the spatial information is decreased while feature information is increased. The extensive pathway collects the feature and spatial information through a sequence of up-convolutions and concatenations with high-resolution features from the constraining path.

At the termination of each up-sampling, convolution is applied consuming a (3×3) kernel and a ReLU activation occupation. During the up-sampling process, new pixels are increased until the input is molded as the desired dimension, allowing the last layer to utilize a (1×1) convolution which aligns feature vectors to the aim classes.

Building a hybrid U-Net-based DL framework for the lung image database consortium (LIDC)-IDRI dataset's early pulmonary nodule identification and classification is the goal of this research. Through the integration of strong segmentation and classification approaches designed to identify nodules early on, the goal is to increase diagnostic accuracy by utilizing DL's advantages in medical image processing.

2. RELATED WORK

Over the earlier years, cancer handling has been the serious emphasis of medical research on human health over the world [13]. In this section, we presented a comprehensive review of the research conducted over the last decade that has focused on the same subject as our study. Sun *et al.* in 2016 [14], introduced another study using CNNs. The authors implemented three distinct DL approaches, utilizing cropped (52 × 52) pixel covers from the LIDC dataset for training and testing. These approaches included CNN, deep belief networks (DBN), and a stacked denoizing autoencoder (SDAE). The CNN model featured 3 convolutional layers, each followed by max pooling, using 5 × 5 kernels to produce (6, 8, and 12) feature maps. Nodule areas in the images were segmented based on the uniting form of truth files from 4 radiologists. Accuracy was relied upon in performance evaluation, with CNN, DBN, and SDAE achieving scores of 79.76%, 81.19%, and 79.29%, respectively.

In Zhang *et al.* [15] developed a globally optimized hybrid geometric active contour model along with a computerized lung segmentation technique, by integrating both global sections and edge information, the algorithm achieved a very high and average F-measure of (99.22%). The efficiency of the algorithm was validated on (40) images of CT scans from the LIDC dataset. In (2018) Sindhu *et al.* [16], employed a single CNN that instantaneously predicts multiple bounding boxes and their corresponding class probabilities. The approach combined object classification and regression to estimate bounding boxes, resulting in higher inference performance compared to standard classifiers used with a

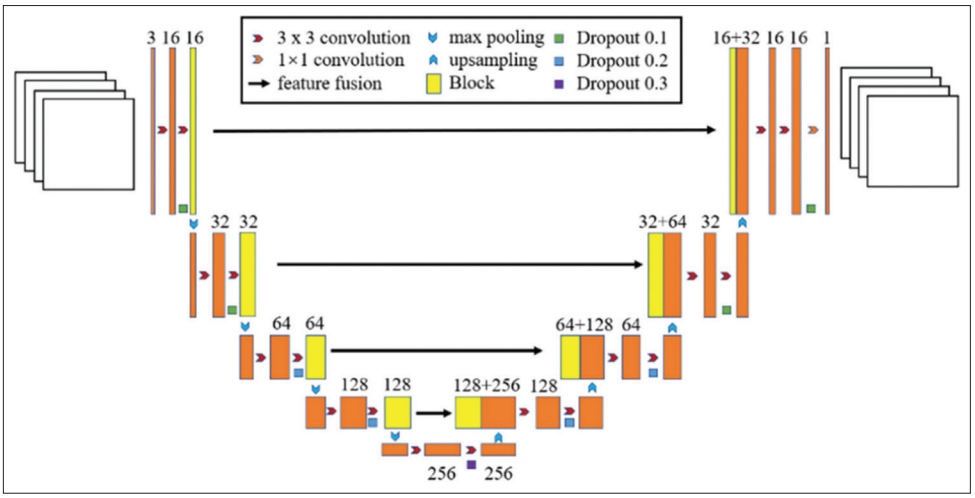


Fig. 3. General structure of the U-Net neural network.

sliding window approach. The Detect-Net model featured a 22 fully convolutional layers without any fully connected (FC) layers, enabling the network to admit the inputs of varied sizes and apply the CNN in a sliding window style with suitable paces. The system was evaluated on the LIDC dataset and achieved a sensitivity of 89% at 6 FPs per image and a precision of 93%. Xie *et al*. [17] modified the faster R-CNN framework by integrating two region proposal networks and a deconvolutional layer to reveal nodules expected. They trained (3) separate models for several types of slices, which were later fused to produce the final results. In addition, they introduced a 2D CNN-based boosting architecture for FP reduction, functioning as a classifier to differentiate true nodules from candidates. Misclassified samples were retained for retraining, improving sensitivity for pulmonary nodule detection. The final categorization results were gained through a voting process that combined the outputs of these networks. Extensive testing showed that a sensitivity of candidate detection was 86.42%. Qinhua *et al*. (2020) developed another technique using an automated segmentation of the lungs in CT images, used mask R-CNN, to specify the model for region mapping of the lung, incorporated with supervised and unsupervised ML methods Bayes. The technique obtained an additional accurate response related to the benchmarked approaches, with a segmentation accuracy of 97.68% [18]. Masood *et al*. [19] introduced an improved multi-dimensional region-based fully convolutional network approach for detecting and classifying pulmonary nodules using LIDC-IDRI dataset. By means of incorporating an efficient multi-layer fusion region proposal network with spatially-sensitive score maps as the cornerstone of the image classifiers, the authors utilized a median intensity projection to leverage CT scan data with applying a deconvolutional layer to identify areas of interest. When compared to other detection and classification methods, the system reached a rate of sensitivity (98.1%) and (97.1%) of accuracy.

After a year, Gopi *et al*. [20] suggested a technique for categorizing and confirming the reliability of different phases of pulmonary nodules progression and introduced a Cloud-based Lung Tumor Detector and Stage Classifier (Cloud-LTDSC) as a hybrid technique for analyzing positron emission tomography/CT images. They used the LIDC-IDRI benchmark dataset and assessed an experimental efficiency result (98.6%) as a rate of classification accuracy.

Musthafa *et al*. [21] proposed a hybrid ML technique for early lung nodule prediction from the LIDC-IDRI dataset using a neural network-based approach for classification. Their approach integrated multiple machine learning methods, including support vector machine, Mask-Region-based Convolutional Neural Network (RCNN), and CNNs, and they introduced an improved snake swarm optimization with a bat model that used statistical data. In addition, a chaotic atom search optimization algorithm was applied to choose the most relevant features, reducing the dimensionality problem. Finally, they designed a learning-based deep neural network hybrid classifier for nodule detection and gained (96.39%), (95.25%), (96.12%), and (96.05%) of accuracy, sensitivity, specificity, and area under the curve (AUC), respectively.

In the same year, Konovalenko *et al*. [11] proposed U-Net-like architectures using encoders such as ResNet, DenseNet, and others and investigated the relevance between the training validation measurements and the final segmentation test. After analysing the detection accuracy, they found that neural networks trained consuming Nesterov momentum stochastic gradient descent optimizer had the greatest generality properties. To choose the superlative model throughout training based on the validation measurements, the core test measurements of recognition quality (dice similarity coefficient) were analyzed based on the validation metrics. It was found that ResNet and DenseNet models realize the highest generalization characteristics for their task. The maximum accuracy for recognition was achieved using the ResNet152 backbone with U-Net model. Their proposal reached 93.04% of accuracy.

After a year, Suriyavarman *et al*. combined an efficient net neural network technique with U-Net for both goals (segmentation and classification) of pulmonary nodules on the LIDC-IDRI dataset. A semi-supervised technique that depends on extracting features was used to yield the advantage of the massive amount of input images without extreme signs. They applied ResNet-50 model and feature pyramid network (FPN) to extract features and then applied a neural network classifier to estimate unlabeled nodules. Their technique's results presented a higher accuracy of (91.67%) [22].

Gupta *et al*. proposed framework exploited an adapted U-Net architecture on the LIDC-IDRI dataset which employed the classical u-net manner for multi-scale feature extraction with Differentiable Architecture Search (DARTS) which applied perception of effective architecture illustration. Moreover, the adapted U-Net is accomplished detecting and classifying pulmonary nodules efficiently also professionally. The proposed framework obtained 95.01% of accuracy rate [23].

The related works covered in this part are summarized in Table 1 below, it should be noted that numerous other studies have also used the LIDC-IDRI dataset and produced encouraging results, but a thorough discussion of those studies was not possible owing to space constraints.

Our suggested methodology has a number of advantages over other methods, such as better classification performance by integrating handmade and deep features, more robustness in recognizing small or low-contrast nodules, and more accurate segmentation thanks to the U-Net design. This hybrid approach, in contrast to conventional techniques, combines DL and feature engineering, leading to improved generalization and increased diagnostic accuracy on the LIDC-IDRI dataset.

3. PROPOSED APPROACH

This section reveals the detailed of our proposed approach which comprises the detection of pulmonary nodules using DL which has become a key motivation in medical imaging, essentially for early lung cancer diagnosis. It is also worth mentioning that several DL techniques are extensively used and have been confirmed to be effective. The preferred approach is often contingent on the dataset, computational properties, and the accuracy required.

In this study, we proposed a hybrid image classification approach combining texture and shape feature extraction methods with (U-Net CNN) to classify the LIDC images into two predetermined kinds. The method consisted of numerous phases: Preprocessing, features extraction, and u-net-based DL classifications as shown in Fig. 4 bellow.

Our workflow can be summarized as follows:

3.1. Dataset Preparation

LIDC resource is an open-accessible universal dataset for the development and evaluation of CAD approaches to pulmonary nodules detection and diagnosis, which contains (1018) medical CT scan images and an associated XML [24]. Each image was labeled for pests that belonged to one of three classes ("nodule >=3 mm," "nodule <3 mm," and "non-nodule >=3 mm"), and the aim of this idea is to recognize as completely as possible all nodules for each image in the dataset [25].

3.2. Preprocessing

Our preprocessing techniques were used to enhance the feature of the images, which in turn improve the performance of the detection and classification, by enhancing the contrast of specific ranges after adjusting the intensity levels according to local histograms [26], this stage involved the contrast limited adaptive histogram equalization (CLAHE) algorithm as shown in Fig. 5, which executed to enhance the contrast of each input image.

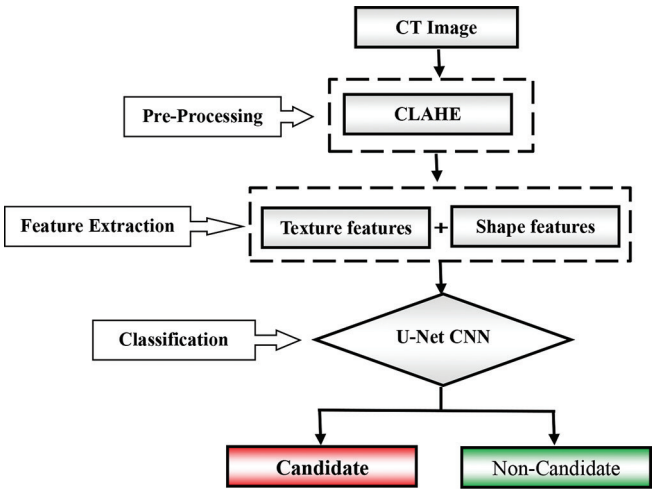


Fig. 4. The diagram of our proposed technique.

3.3. Texture and Shape Feature Extraction

The step before applying the DL model, we preferred to employ both texture and shape-based feature extraction techniques to support the U-Net's performance:

a) Texture Features: Gray-level co-occurrence matrix (GLCM) algorithm is computed from grayscale images to capture the texture details. GLCM is computed with an offset of [0, 1] and a specified some of gray levels to encrypts information about the spatial relationship between pixel densities and depicts important texture patterns, and seven texture features had been extracted, which capture different aspects of the texture of each image, and the features were Contrast, Correlation, Energy, Homogeneity, Entropy, Dissimilarity, Autocorrelation.

b) Shape Features: After applying Otsu's segmentation to the preprocessed image to obtain a binary mask, we extracted four shape-based features using the region prop's function included (Area, Perimeter, Eccentricity, and Solidity) were merged, respectively, into a single feature vector that is ultimately appended to the seven texture features extracted from GLCM to estimate the performance of the classification.

3.4. U-Net-based Architecture

Our technique is performed using a hybrid approach, where both texture and shape features are combined with features

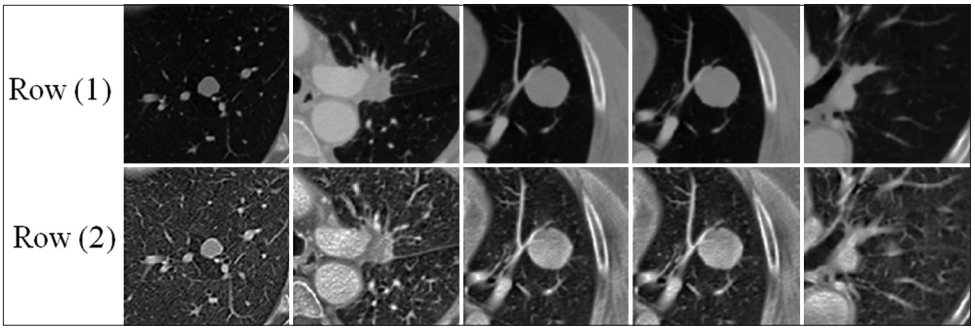


Fig. 5. Demonstrates the differences between the original images row (1) and the same images after applying contrast-limited adaptive histogram equalization row (2).

Table 1: Summary of related works (they all used the Lung Image Database consortium-IDRI dataset)

Author (s)/year	Preprocessing	Model training/network	Feature Extraction	Classification	Results %
Sun *et al*.[14] in 2016	Cropped patches (52×52), truth file segmentation	CNN, DBN, SDAE with 3 conv layers each	CNN feature maps (5×5 kernels)	CNN/DBN/ SDAE classifiers	Accuracy CNN: 79.76 DBN: 81.19 SDAE: 79.2
Zhang *et al*.[15] in 2017	Lung segmentation	Hybrid geometric active contour	Global and edge info	N/A	F-measure: 99.22
Sun *et al*.[16] in 2018	N/A	Detect-Net (22 conv layers, no FC)	Bounding box regression+ classification	Sliding window with detection model	Sensitivity: 89 Precision: 93
Xie *et al*.[17] in 2019	N/A	Modified faster R-CNN with two RPNs	2D CNN-based boosting, deconvolutional layer	Voting-based classifier	Sensitivity: 86.42
Qinhua *et al*.[18] in 2020	Mask R-CNN-based lung segmentation	Supervised+unsupervised machine learning+Bayes	Region mapping	Bayes classifier	Accuracy: 97.68
Masood *et al*.[19] in 2020	Median intensity projection	multi-dimensional region-based fully convolutional network with multi-layer fusion region proposal network, deconvolutional layers	Spatially-sensitive score maps	Region-based FCN	Sensitivity: 98.1 Accuracy: 97.1
Gopi *et al*.[20] in 2021	N/A	Cloud-based lung tumor detector and stage classifier with positron emission tomography/CT integration	N/A	Multi-phase hybrid classifier	Accuracy: 98.6
Musthafa *et al*.[21] in 2022	Statistical feature selection	Hybrid machine learning: Support vector machine, Mask-RCNN, CNN, improved snake swarm optimization with a bat model, CASO	CASO for best features	Learning-based deep neural network hybrid neural classifier	Accuracy: 96.39 Sensitivity: 95.25 area under the curve: 96.05
Konovalenko *et al*.[11] in 2022	N/A	U-Net with ResNet/DenseNet encoders	Nesterov SGD training	ResNet152 with U-Net	Accuracy: 93.04
Suriyavarman *et al*.[22] in 2023	Semi-supervised learning on unlabeled data	EfficientNet+U-Net+FPN+ResNet-50	FPN	Neural network classifier	Accuracy: 91.67
Gupta *et al*.[23] in 2024	Classical U-Net pipeline	Adapted U-Net with DARTS architecture	Multi-scale feature extraction	End-to-end classification	Accuracy: 95.01

CASO: Chaotic atom search optimization, FPN: Feature pyramid network, CNN: Convolutional neural network, DBN: Deep belief network, SDAE: Stacked denoizing autoencoder, RPN: Region Proposal Networks, SGD: Stochastic Gradient Descent

learned by a custom U-Net-based. The U-Net architecture is designed specifically for medical image segmentation and classification, as detailed in Fig. 6 below:

The methodology increased model interpretability, which is essential in clinical practice, in addition to improving detection accuracy for small and subtle nodules. All things

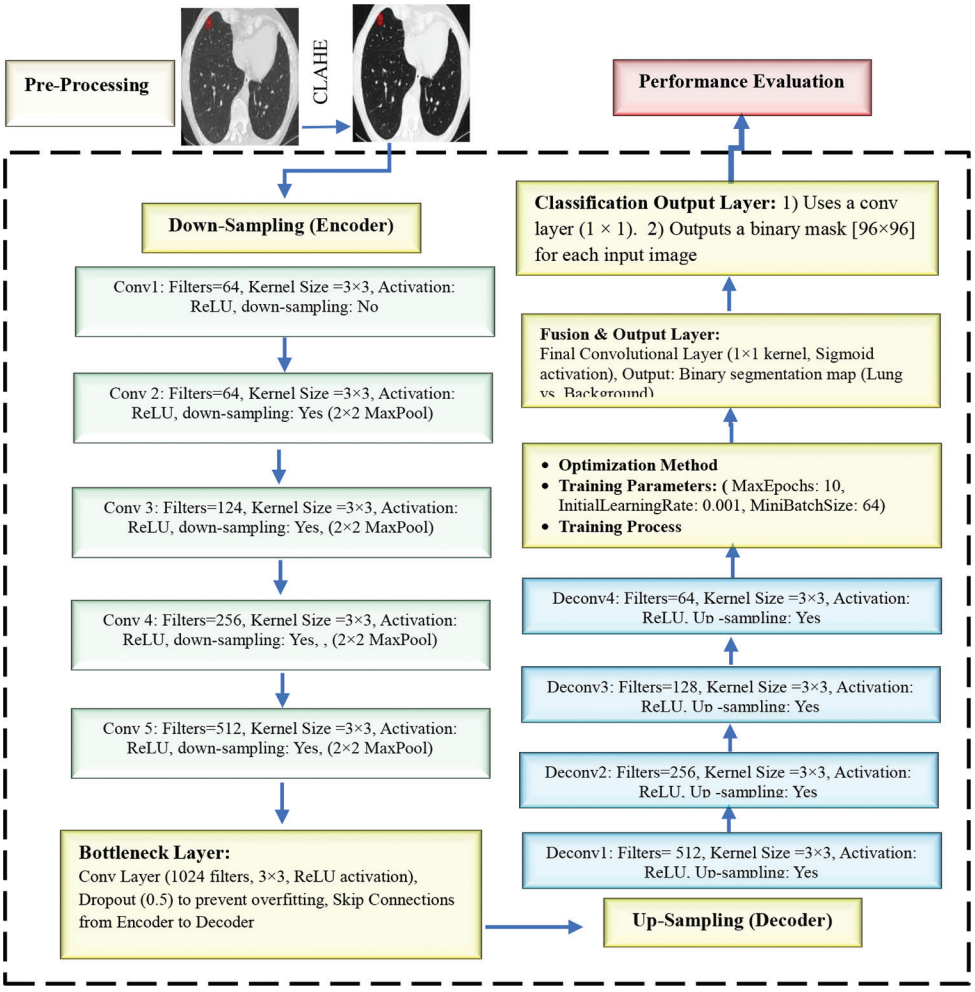


Fig. 6. The proposed framework block diagram.

considered, the framework offers a significant advancement in the direction of more accurate, comprehensible, and effective CAD in lung cancer screening.

4. EXPERIMENTAL RESULT

To compare the results with other models and decide how to use them in the future, it is standard procedure to evaluate the outcomes after an experiment [26]. We utilized accuracy, sensitivity, specificity, and AUC as metrics to evaluate the model and our methodology. The following techniques are employed to describe them:

a) Accuracy is the most commonly used measure to answer the question, "Of all the predictions we made, how many were correct?" It is the ratio of accurate forecasts to total predictions and is calculated with the use of Equation (1)

$$Accuracy = \frac{TP + FP + TN + FN}{TP + FP + TN + FN} \times 100 \quad (1)$$

b) Specificity percentage of true negative situations that were accurately forecasted to be negative. and determined using Equation (2).

$$Specificity = \frac{TN}{TN + FP} \times 100 \quad (2)$$

c) Sensitivity emphasizes the model's effectiveness in producing all the favorable results. Additionally known as the "true positive rate," it provides an answer to the query, "How many of all the data points did we acceptably predict as true?" Equation (3) below is used to calculate recall.

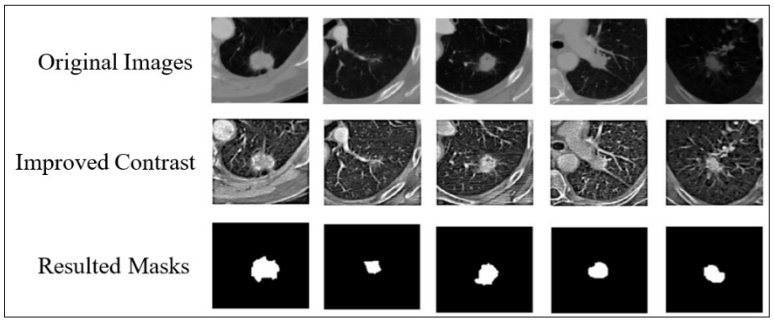


Fig. 7. Pulmonary nodule segmentation results using U-Net convolutional neural network.

Table 2: Comparative five experimental models results

Model	Accuracy%	Sensitivity%	Specificity%	AUC%
FCN	94.32	94.32	93.30	93.36
SegNet	94.20	94.21	93.21	93.27
ResNet	94.67	94.67	93.65	93.71
DenseNet	95.09	95.09	94.07	94.13
U-Net	95.35	95.33	94.23	95.44

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100 \tag{3}$$

d) AUC, or more precisely, the receiver operating characteristic curve. It measures a binary classifier's total capacity to discriminate between positive and negative classes at every classification threshold. Equation (4) is used to determine the AUC.

$$\text{AUC} = \int_0^1 TPR\left(FPR\right) d\left(FPR\right) \times 100 \tag{4}$$

$$\text{Where TPR is true-positive rate} = \frac{TP}{TP + FN} \tag{5}$$

$$\text{and FPR is false-positive rate} = \frac{FP}{FP + TN} \tag{6}$$

The decision to employ these particular metrics (accuracy, specificity, sensitivity, and AUC) depends on the classification challenges, the application domains, and the requirements for a thorough assessment [23].

4.1. Experimental Setup

In this study, we divided all the downloaded image in LIDC as two categories: training and testing groups, and different ratios were tested as the following (training-testing) order: (50% - 50%), (60% - 40%), (70% - 30%), then (80% - 20%); finally, we had reached that the last allocation (80% - 20%) provides the best findings.

Table 3: Results of other comparable techniques

Techniques	Accuracy %	Sensitivity %	Specificity %	AUC %
Musthafa *et al*. [21], 2022	96.39	95.25	-	96.05
Konovalenko *et al*. [11], 2022	93.04	-	-	-
Suriyavarman and Annie [22], 2023	91.67	-	-	-
Gupta *et al*. [23], 2024	95.01	-	-	-
Proposed approach	95.35	95.33	94.23	95.44

4.2. Data Preprocessing

To improve the performance of preprocessing by adopting CLAHE to improve the input images contrast and certainly increases the accuracy and speed of nodule identification and reduce the computational load of network training and evaluation, therefore, to reduce the interference of unimportant lung regions processed on the evaluation outcomes, the obvious salient locations of each labeled image were cropped based on the node position provided in the annotation files, and the original image sizes were cropped to (96 × 96) pixels to retain the fully important private information as shown in Fig. 7.

4.3. Results

To validate our technique and ensure the segmentation effect of the proposed hybrid U-Net framework model, six models, including FCN, SegNet, Resnet, Densenet, and U-Net, were tested, respectively; the results in Table 2 showed the strength of our proposed model.

Even though the objective of this study was diagnostic accuracy, processing time is still an important consideration for real-time clinical applications. In subsequent studies, we want to assess and improve the framework's computational

effectiveness. This will facilitate its useful implementation in medical situations where time is of the essence.

4.4. Comparative Analysis

More tests were conducted to evaluate our suggested technique performance by comparing it with other techniques [Table 3]. Some of cells are left as blank because all of the most recent methods either tested their strategy only in terms of accuracy or combined accuracy with other metrics.

5. CONCLUSION

Pulmonary nodules early detection using CAD systems is crucial in reducing lung cancer mortality rates, especially given the asymptomatic nature of the disease in its early stages.

In order to detect and classify pulmonary nodules early on, we used the LIDC-IDRI dataset to propose a robust hybrid DL system based on the U-Net architecture. To detect even subtle or irregular nodules, the framework showed good accuracy and reliability by combining DL approaches with efficient preprocessing, segmentation, and custom feature extraction.

Great diagnostic performance was demonstrated by the model's remarkable 95.35% accuracy, 95.33% sensitivity, 94.23% specificity, and 95.44% AUC. These outcomes demonstrate the value of integrating DL with conventional techniques to produce an all-encompassing and interpretable system. In addition to increasing diagnostic confidence, the suggested method aids in clinical judgment when screening for lung cancer. Although encouraging, additional validation on external datasets is required to improve generalizability. However, this structure provides a solid basis for further developments in automated pulmonary nodule identification.

REFERENCES

[1] H. Jin, C. Yu, Z. Gong, R. Zheng, Y. Zhao and Q. Fu. "Machine learning techniques for pulmonary nodule computer-aided diagnosis using CT images: A systematic review". *Biomedical Signal Processing and Control*, vol. 79, p. 104104, 2023.

[2] Y. Gu, J. Chi, L. Yang, B. Zhang, D. Yu, Y. Zhao and X. Lu. "A survey of computer-aided diagnosis of lung nodules from CT scans using deep learning". *Computers in Biology and Medicine*, vol. 137, pp. 1-27, 2021.

[3] B. Youssef, A. Alksas, A. Shalaby, A. Mahmoud, E. V. Bogaert, N. S. Algahmdi, A. Neubacher, S. Contractor, M. Ghazal, A. Elmaghraby and A. El-Baz. "Integrated deep learning and stochastic models for accurate segmentation of lung nodules from computed tomography images: A novel framework". *IEEE*, vol. 11, pp. 99807-99821, 2023.

[4] Z. Zhao, S. Guo, L. Han, L. Wu, Y. Zhang and B. Yan. "Altruistic seagull optimization algorithm enables selection of radiomic features for predicting benign and malignant pulmonary nodules". *Computers in Biology and Medicine*, vol. 180, p. 108996, 2024.

[5] X. Zhan, H. Long and J. Wu. "A semantic fidelity interpretable-assisted decision model for lung nodule classification". *International Journal of Computer Assisted Radiology and Surgery*, vol. 19, pp. 625-633, 2023.

[6] C. Wenming, R. Wu, G. Cao and Z. He. "A comprehensive review of computer-aided diagnosis of pulmonary nodules based on computed tomography scans". *IEEE*, vol. 8, pp. 154007-154023, 2020.

[7] R. Li, C. Xiao, Y. Huang, H. Hassan and B. Huang. "Deep learning applications in computed tomography images for pulmonary nodule detection and diagnosis: A review". *Diagnostics*, vol. 12, p. 298, 2022.

[8] Y. Ye, M. Tian, Q. Liu and H. M. Tai. "Pulmonary nodule detection using V-Net and high-level descriptor based SVM classifier". *IEEE*, vol. 8, pp. 176033-176041, 2020.

[9] R. Mothkura and B. N. Veerappa. "Classification of lung cancer using lightweight deep neural networks". *Procedia Computer Science*, vol. 218, pp. 1869-1877, 2023.

[10] J. Chang, Y. Li and H. Zheng. "Research on key algorithms of the lung CAD system based on cascade feature and hybrid swarm intelligence optimization for MKL-SVM". *Hindawi, Computational Intelligence and Neuroscience*, vol. 2021, pp. 1-16, 2021.

[11] I. Konovalenko, M. Pavlo, B. Janette, P. Olegas and B. Jakub. "Research of U-Net-Based CNN architectures for metal surface defect detection". *MDPI*, vol. 10, no. 5, pp. 1-19, 2022.

[12] X. Zhang, S. Kong, Y. Han, B. Xie and C. Liu. "Lung nodule CT image segmentation model based on multiscale dense residual neural network". *Mathematics*, vol. 11, no. 6, pp. 1-14, 2023.

[13] W. Cao, R. Wu, G. Cao and Z. He. "A comprehensive review of computer-aided diagnosis of pulmonary nodules based on computed tomography scans". *IEEE Access*, vol. 8, pp. 154007-154023, 2020.

[14] Sun, B. Zheng and W. Qian. "Computer Aided Lung Cancer Diagnosis with Deep Learning Algorithms". pp. 1-8, 2016.

[15] W. Zhang, X. Wang, P. Zhang and J. Chen. "Global optimal hybrid geometric active contour for automated lung segmentation on CT images". *Computers in Biology and Medicine*, vol. 91, pp. 168-180, 2017.

[16] S. S. Ramachandran, J. George, S. Skaria and V. V. Varun. "Using Yolo Based Deep Learning Network for Real Time Detection and Localization of Lung Nodules from Low Dose CT Scans", 2018.

[17] H. Xie, D. Yang, N. Sun, Z. Chen and Y. Zhang. "Automated pulmonary nodule detection in CT images using deep convolutional neural networks". *Pattern Recognition*, vol. 85, pp. 109-119, 2019.

[18] Q. Hu, L. F. Souza, G. B. Holanda, S. S. A. Alves, F. H. Silva, T. Han and P. P. Rebouças Filho. "An effective approach for CT lung segmentation using mask region-based convolutional neural networks". *Artificial Intelligence in Medicine*, vol. 103, p. 101792, 2020.

[19] A. Masood, B. Sheng, P. Yang, P. Li, H. Li, J. Kim and D. D. Feng. "Automated decision support system for lung cancer detection and classification via enhanced RFCN with multilayer fusion RPN".

IEEE, vol. 16, pp. 7791-7801, 2020.

[20] G. Kasinathan and S. Jayakumar. "Cloud-based lung tumor detection and stage classification using deep learning techniques". *Biomed Research International*, vol. 2022, pp. 1-17, 2021.

[21] S. Musthafa, K. Sankar, T. Benil and Y. N. Rao. "*A Hybrid Machine Learning Technique for Early Prediction of Lung Nodules from Medical Images using a Learning-Based Neural Network Classifier*". Wiley, United States, 2022.

[22] S. Suriyavarman and A. X. Annie. "Lung nodule segmentation and classification using u-net and efficient-net". *International Journal of Advanced Computer Science and Applications,* vol. 14, no. 7, pp. 737-745, 2023.

[23] A. Gupta, A. Kumar and K. Rautela. "UDCT: Lung cancer detection and classification using U-net and DARTS for medical CT images". *Multimedia Tools and Applications*, vol. 84, pp. 19065-19085, 2024.

[24] JustinKirby. "*The Cancer Imaging Archive: LIDC-IDRI*". Kaggle, 2022. Available from: https://www.kaggle.com/datasets/justinkirby/the-cancer-imaging-archive-lidcidri [Last accessed on 2024 Oct 18].

[25] Archive, The Cancer Imaging. "*LIDC-IDRI|Data from The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Reference Database of Lung Nodules on CT Scans*". National Cancer Institute, Cancer Image Program (CIP), 2020. Available from: https://www.cancerimagingarchive.net/collection/lidc-idri [Last accessed on 2024 Oct 18].

[26] F. A. A. Salih, S. Tahir, T. Ahmed and H. J. Mohammed. "An effective computer-aided diagnosis technique for alzheimer's disease classification using U-net-based deep learning". *UHD Journal of Science and Technology*, vol. 9, no. 1, pp. 34-43, 2025.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Effects of Salicylic Acid on Barley Growth and Productivity under Cadmium Stress Conditions in Rania, Kurdistan Region, Iraq

Rozhgar A. Mustafa*, Arol M. Anwar, Ikbal M. Albarzinji

Department of Biology, Faculty of Science and Health, Koya University, Koya KOY44023, Kurdistan Region – F. R. Iraq

ABSTRACT

Plant health and agricultural productivity are seriously threatened by heavy metal contamination. Due to the expansion of cereal agricultural lands into marginal areas close to urban areas and unpaved roads that are polluted by many pollutants, especially that of cadmium (Cd), this study examined the effects of Cd and salicylic acid (SA) separately and in combination on the growth, physiological, biochemical, and reproductive responses of barley (*Hordeum vulgare* L.). Three SA treatments (0, 86, and 172 mg/kg soil) and four Cd concentrations (0, 10, 20, and 30 mg/kg soil) were used in a factorial pot experiment. The findings showed that high Cd (Cd30) reduced vegetative growth but increased spike number and harvest index, indicating reproductive compensation, moderate Cd levels (Cd20) improved some growth traits, including plant height by 4.00% and flag leaf area by 13.11% compared to the control treatment, suggesting a possible hormetic effect. Particularly under moderate Cd stress, SA at 86 mg/kg markedly enhanced plant height, yield components, and antioxidant balance. On the other hand, SA at 172 mg/kg increased grain number and spike length but decreased overall yield, as a result of metabolic effects and hormonal interferences. Under extreme stress, endogenous defense mechanisms might be sufficient, and external SA could upset homeostasis, according to the interaction effects, which showed that Cd20 × SA86 maximized growth and productivity while Cd30 × SA0 produced an increase in grain and biological yields by an amount of 10.95 and 9.73%, respectively, compared to the control treatment. These results show that the performance of growth and yield components of *H. vulgare* L. was significantly and variably affected by both Cd and SA, both separately and in combination, depending on the concentration and the interaction. Data suggest that SA is most effective at moderate Cd stress levels.

Index Terms: Antioxidant, Cadmium Stress, Heavy Metal, *Hordeum Vulgar* L, Salicylic Acid

1. INTRODUCTION

Barley (*Hordeum vulgare* L.) which belongs to the family Poaceae is one of the most important cereal crops cultivated in increasing numbers of areas in the world [1]. It belongs to the Hordeum genus, which has 32 species and 45 taxa. Barley is one of the top ten crop plants in the world and comes in fourth place among cereal crops, after rice, wheat, and maize, according to Abdullah *et al.* [2]. In 2018, the global average barley production was 2.91 tons/ha [3]. Iraq produced 6,238,392 tons of barley in 2020, with 3,469,646 hectares under cultivation [4]. In Kurdistan region, rainy conditions give the highest yield of grain and straw. The area cultivated by barley in Kurdistan region governorates during the winter season 2013 was only 84170 hectares with production of only 135,183 tons with an average yield of 1,606 kg/ha [5].

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp276-286

E-ISSN: 2521-4217
P-ISSN: 2521-4209

Copyright © 2025 Mustafa, *et al*. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Rozhgar A. Mustafa, Department of Biology, Faculty of Science and Health, Koya University, Koya KOY44023, Kurdistan Region – F. R. Iraq. E-mail: rozhgar_abubakir@uor.edu.krd

Received: 30-07-2025 Accepted: 26-10-2025 Published: 03-11-2025

In ancient Kurdish communities, barley was used as a staple for bread-making and as a vital source of grain and straw for animals. Barley was also seen as an essential feed grain that was less expensive for low-income families [5]. Numerous studies have shown how it might enhance human health by reducing blood sugar and cholesterol levels, thus averting diabetes and cardiovascular ailments [6]. One of the most significant cereal crops in Iraq's Kurdistan area is barley which thrives in a variety of climates and grows better than other cereals during the dry winter months [2]. However, barley productivity is threatened by heavy metal contamination, especially cadmium (Cd), which is a non-essential element with no biological function. It is a highly toxic element for plants and animals and it is fixed in the environment, causing poisonous to plants even at low doses [7]. The two categories of sources that release heavy metals into soil, air, and water are natural and man-made sources [8]. The morphological, physiological, biochemical, and molecular levels of plants are all impacted by Cd poisoning. As Cd dosages rise, barley biomass and leaf area decline [7]. The mobility of Cd in soil is mostly determined by the soil's pH, mineral content, and organic matter, where Cd is more mobile in sandy, acidic soils with low organic matter, but it tends to bind strongly to iron oxides and organic matter, decreasing its availability. In plants, Cd is absorbed mainly as Cd^{2+} ions through root cell membranes through transporters normally utilized for essential metals such as calcium, iron, and zinc. Plant roots can readily absorb Cd, which then travels to the higher portions of the plant [9]. It enters roots through apoplastic diffusion and symplastic uptake, competes with Ca^{2+} at calcium channels, and is often sequestered in root cell vacuoles by chelation with phytochelatins to reduce toxicity. Some Cd is then loaded into the xylem for translocation to shoots, driven by transpiration, and may further move to seeds through the phloem [10].

Salicylic acid (SA) is considered one of the most important plant growth regulators that play a major role in plant growth and development. Many metabolic and vital physiological processes in plants including thermogenesis, stomata closure, and flowering are dependent on SA [11]. Previously, it was reported that SA can alleviate growth inhibition caused by Cd toxicity in various plant species, including barley [12]. At low to moderate concentrations (0.5–1.5 mM), SA enhanced lettuce's resistance to Cd stress. SA improved the plants' ability to tolerate oxidative stress by increasing root growth, proline, catalase activity, and carbohydrate content; in addition, it raised levels of chlorophyll a, which aided in photosynthesis. Higher SA levels (2 mM), however, decreased carotenoids, chlorophyll b, and leaf area, suggesting potential toxicity at high concentrations [13].

The aims of this study are to assess the impacts of different Cd and SA concentrations on some growth, physiological, biochemical, and yield components properties of barley and to recognize the ability of SA to reduce the toxicity of Cd on this plant.

2. MATERIALS AND METHODS

2.1. Plant Materials, Cultivation, and Treatments

The study included carrying out a factorial experiment at the University of Raparin, Rania city, Sulaimani, Kurdistan Region, Iraq, during the growing season from December 2024 to May 2025 based on a completely randomized design (CRD) with three replications on barley local variety brought from the University of Sulaimani, College of Agricultural Engineering Sciences, which was cultivated in a soil mixed with the heavy metal Cd and the plant hormone SA. A silty loam soil with pH 7.41, EC 0.26 dS/m, and O.M 2.29% was ground well and sieved to 4 mm. Plastic pots were filled with 32 kg of soil, and 15 seeds were sown in each pot (diameter of 42 cm and a height of 36 cm). Before sowing, a light irrigation was applied to ensure uniform germination, and other agricultural practices were carried out as needed. Nitrogen fertilizer was used at a rate of 126 kg/ha, half was in the form of DAP (n = 18% and P = 46%), applied before the time of sowing, and the remaining was added in the form of urea (n = 46%), separated into two parts, the first added at the beginning of the stem elongation stage and the other at the heading stage.

The study consisted of two factors, the first was Cd chloride ($CdCl_2.H_2O$) as a source of Cd, with three concentrations of Cd (10, 20, and 30 mg/kg soil) (write the reference of Cd I sent), in addition to the no Cd treatment, mentioned as Cd0, Cd10, Cd20, and Cd30, and the second factor was SA ($C_7H_6O_3$) (2-hydroxy benzoic acid), with two concentrations (1 and 2 mM) converted to (86 and 172 mg/kg soil), similar concentrations have been used for barely as foliar application or soaking seeds before sowing, but there is no researches on using SA with soil (write the two reference I sent), in addition to the no SA treatment mentioned as SA0, SA86 and SA172. The soil of each experimental unit was mixed well with the added Cd and SA at the required concentrations according to the treatments (Fig. 1).

2.2. Meteorological Data

Monthly averages of minimum and maximum of air and soil temperature and relative humidity, in addition to wind speed, are included in Table 1.

TABLE 1: Some meteorological data throughout the study period*

Year	Month	Air temperature (◦C)		Soil temperature (◦C)		Relative humidity (%)		Wind speed (km/h)
		Max.	Min.	Max.	Min.	Max.	Min.	
2024	December	9.52	7.81	9.45	8.89	72.18	65.71	1.22
2025	January	9.04	7.59	8.74	8.07	65.87	59.18	1.48
	February	6.71	5.42	7.79	7.72	63.22	56.23	1.51
	March	15.30	13.30	15.42	12.62	51.39	44.59	1.41
	April	19.91	18.24	19.84	14.28	53.41	15.35	1.53
	May	27.80	25.97	26.85	21.00	34.11	16.40	1.85

*Rania-Grdjan Station, Bakrajo Agronomy, General Directorate of Agriculture, Sulaimani

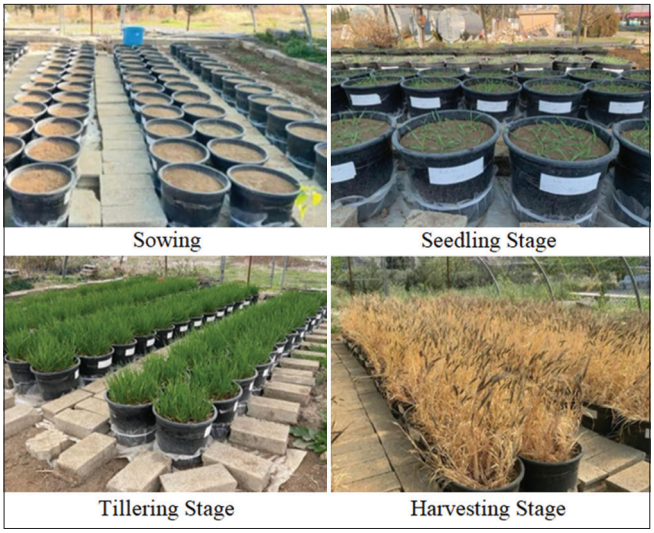


Fig. 1. Barley plants used in the study at different stages.

2.3. Studied Characteristics

For each experimental unit, ten plants were randomly chosen for each pot and were used to determine all studied characteristics as follows: Germination percent was measured by the equation mentioned by [ ] as follows: [14]

$$G = \frac{L}{S} \times 100 \quad (\text{seed day}^{-1}) \qquad (1)$$

where G is the germination percent, L is the germinated seeds, S is the total sown seeds

Germination velocity was measured by the equation mentioned by (2) [15] as follows:

$$GV = \frac{\sum n\ D}{n} \ (\text{day}) \qquad (2)$$

Where, GV is the germination velocity, n is the number of seeds, which were germinated corresponding to the day D observation (not the accumulated number) and D is number of days counted from the beginning of germination.

Flag leaf area (cm^2) was measured by the equation applied by Al-Hassnawi and Al-Burki [16]. Plant height was measured by a measuring tape line from the soil surface to the barley spike base at physiological maturity [17]. The number of emerged tillers was recorded. The ratio of 20 mL of 80% acetone to 0.4 g fresh sample was used to estimate chlorophyll a (Chl.a), chlorophyll b (Chl.b), and total carotenoids as reported by Wellburn [18]. The activity of peroxidase (POD) enzyme (absorbing units g^{-1} fresh weight [FW]) in leaves was estimated according to Müftügil [19] spectrophotometrically at 420 nm using guaiacol and H_2O_2. The proline content in leaves (mM/g FW) was estimated using spectrophotometer at 520 nm which was calibrated with a proline standard curve using sulfosalicylic acid and ninhydrin as it mentioned by Bates *et al.* [20]. Total carbohydrate in dry leaves was determined using concentrated sulfuric acid (H_2SO_4) and 5% phenol spectrophotometrically at 488 nm as reported by Herbert *et al.* [21].

Yield and yield components were measured at harvest, as it described by Shemi *et al.* [22] including number, length (cm), and weight (g) of spikes, spike grain number, weight of 1,000 grains (g), whereas grain yield (ton $hectare^{-1}$) which was calculated from the yield of each experimental unit and converted to ton $hectare^{-1}$, biological yield (ton $hectare^{-1}$) was calculated from the weight of the entire above soil harvested dry plants for an area of 0.1256 m^2 (pot area) then converted to ton $hectare^{-1}$, biological yield without grain yield was the straw yield. Harvest index (%) is the ratio of grain yield to biological yield as mentioned by Sharma and Smith [23] as it appears in the following equation:

$$\text{Harvest index} = (\text{Grain yield/biological yield}) \times 100 \qquad (3)$$

2.4. Statistical Analysis

This study was conducted as a factorial experiment in a CRD with three replications. Analysis of variance was used for

data analysis, and the test of Duncan's multiple range at a 5% probability level was used for the comparison between the experiment means, using the statistical program SAS version 9.1 [24].

3. RESULTS AND DISCUSSION

3.1. Vegetative Growth

Results in (Table 2) show increasing Cd concentrations significantly delayed germination velocity from 11.51 days for the control treatment to 12.38 days for the highest Cd concentration (30 mg/kg^{1} soil). Cd concentration of 20 mg kg^{1} soil was the best for the significant increase in plant height, number of tillers, and flag leaf area to 66.06 cm, 6.96, and 6.21 cm^{2} compared to the control treatment (73.52 cm, 6.87, and 5.49 cm^{2}), whereas increasing Cd concentration to 30 mg kg^{1} soil decreased these characteristics. Regarding the SA effects, it was seen (Table 2) that earlier seed germination was in the control treatment (11.60 days), whereas using SA in both concentrations retarded seed germination to 12.03 and 11.94 days for the SA86 and SA172 treatments. Adding 86 mg/kg soil of SA increased each plant height, number of tillers, and flag leaf area to 56.50 cm, 7.10, and 6.01 cm^{2} compared to the control and SA172 treatments.

In general, interactions of no Cd application regardless the SA concentration led to earlier seed emergence, and increasing Cd concentration to 20 and 30 mg/kg soil, increased the period for seed germination, where the longest period was for the interaction Cd30 × SA86, whereas the earlier germination was in the Cd20 × SA0 treatment not significantly different from the control treatment (11.53 days). Plant height increased significantly to 68.80 cm in the Cd20 × SA86 treatment compared to other treatments except the interactions Cd0 × SA0, Cd0 × SA86, Cd10 × SA86, and Cd30 × SA172, respectively. The lowest plant height was reported in the interaction Cd0 × SA172 significantly compared to all other treatments. The superior interaction for plant height was the same for the number of tillers which recorded 8.16 tillers, significantly higher compared to all other treatments, and the lowest value (5.70) was recorded for the interaction Cd10 × SA0 treatment. It was observed from the results (Table 2) that the effect of higher SA concentration on the vegetative growth characteristics was clear in the high Cd concentration (Cd30 mg/kg soil). Flag leaf area increased significantly in the interactions of Cd10 with 86 and 172 SA, and Cd20 with 0, and SA86 significantly compared to all other interaction treatments. This indicates that SA applied without Cd or at high Cd concentrations, regardless of SA level, led to decreasing most vegetative growth.

Germination percent remained high (97.66–100%) without any significant differences across all treatments; Cd and SA concentrations, and their interactions (Table 2).

TABLE 2: Effects of Cd, SA, and their interactions on some vegetative growth parameters of barley

Treatments	Germination velocity (day)	Germination percent	Plant height (cm)	Number of tillers	Flag leaf area (cm^2)
Cd concentration (mg kg^{-1} soil)					
Cd 0	11.51 c*	100.00 a	63.52b	6.87 ab	5.49 c
Cd10	11.92 b	100.00 a	62.85 b	6.50 c	6.05 a
Cd20	11.63 bc	99.22 a	66.06 a	6.96 a	6.21 a
Cd30	12.38 a	99.22 a	62.82 b	6.73 b	5.71 b
SA concentration (mg/kg soil)					
SA0	11.60 b	100.00 a	63.22 b	6.38 c	5.73 b
SA86	12.03 a	99.41 a	65.50 a	7.10 a	6.01 a
SA172	11.94 a	99.41 a	62.72 b	6.82 b	5.86 ab
Interactions between Cd and SA					
Cd0×SA0	11.53 d-f	100.00 a	66.66 a-c	7.30 b	5.58 de
Cd0×SA86	11.49 d-f	100.00 a	66.55 a-c	6.73 c	5.50 de
Cd0×SA172	11.51 d-f	100.00 a	57.35 f	6.60 cd	5.41 e
Cd10×SA0	11.36 ef	100.00 a	61.80 de	5.70 f	5.33 e
Cd10×SA86	11.83 c-e	100.00 a	66.20 a-c	6.90 c	6.35 a
Cd10×SA172	12.56 b	100.00 a	60.55 e	6.90 c	6.49 a
Cd20×SA0	11.12 f	100.00 a	63.80 cd	6.33 de	6.22 ab
Cd20×SA86	11.65 d-f	97.66 a	68.80 a	8.16 a	6.43 a
Cd20×SA172	12.13 b-d	100.00 a	65.60 bc	6.40 de	5.97 bc
Cd30×SA0	12.40 bc	100.00 a	60.63 e	6.20 e	5.79 cd
Cd30×SA86	13.17 a	100.00 a	60.45 e	6.60 cd	5.76 cd
Cd30×SA172	11.57 d-f	97.66 a	67.40 ab	7.40 b	5.59 de

*The means followed by the same letter for each factor and their interactions separately are not significantly different at $P \leq 0.05$ according to Duncan's multiple range test

3.2. Photosynthetic Pigments

Results presented (Table 3) show that chlorophyll a content was significantly affected by Cd and SA treatments. The highest chlorophyll a (0.85 and 0.84 mg/g FW) was observed at Cd10 and Cd20, respectively, while the lowest value (0.82 mg/g FW) occurred in both the control (Cd0) and Cd30 treatments. Chlorophyll b content increased significantly with Cd20 (0.69 mg/g FW) compared to other Cd levels, while Cd30 caused a marked reduction (0.55 mg/g FW). Interestingly, total carotenoids were highest (0.15 mg/g FW) for Cd30, whereas a significant decrease was recorded at Cd20 (0.10 mg/g FW). Regarding SA effects, SA86 treatment reduced chlorophyll a to 0.82 mg/g FW, while the highest value (0.84 mg/g FW) was recorded under the control (SA0). Chlorophyll b content was significantly enhanced in the control SA treatment (0.70 mg/g FW), while both SA treatments reduced it to 0.58 and 0.56 mg/g FW for SA86 and SA172, respectively. Carotenoid content, on the other hand, increased significantly with rising SA concentration compared to the SA0 treatment, peaking at 0.16 mg/g FW in the SA172 treatment.

Significant interaction effects between Cd and SA concentrations were also evident (Table 3), where the highest chlorophyll a (0.87 mg/g FW) was recorded under the Cd10 × SA172 treatment, followed closely by Cd10 × SA86 (0.86 mg/g FW), while the lowest (0.80 mg/g FW) was observed in Cd0 × SA86 and Cd30 × SA172. Chlorophyll b reached a maximum of 0.75 mg/g FW in the Cd20 × SA0 treatment and was significantly reduced under Cd30 × SA172 (0.44 mg/g FW). Carotenoid content increased significantly in Cd30 × SA172 (0.21 mg/g FW), whereas the lowest value (0.08 mg/g FW) was found in the Cd20 × SA0 interaction. These results indicate that Cd and SA treatments have complex and sometimes opposing effects on pigment concentrations, indicating that high SA and Cd levels may mitigate or exacerbate pigment decline depending on the specific combination.

TABLE 3: Effects of Cd, SA, and their interactions on photosynthetic pigments of barley

Treatments	Chlorophyll a	Chlorophyll b	Total carotenoids
	Pigments (mg/g fresh weight)		
Cd concentration (mg kg^{-1} soil)			
Cd 0	0.82 b*	0.61 b	0.14a
Cd10	0.85 a	0.60 b	0.14 a
Cd20	0.84 a	0.69 a	0.10 b
Cd30	0.82 b	0.55 b	0.15 a
SA concentration (mg/kg soil)			
SA0	0.84 a	0.70 a	0.10 b
SA86	0.82 b	0.58 b	0.14 a
SA172	0.83 ab	0.56 b	0.16 a
Interactions between Cd and SA			
Cd0×SA0	0.85 a-c	0.72 a	0.09 ef
Cd0×SA86	0.80 f	0.52 de	0.17 b
Cd0×SA172	0.81 d-f	0.59 b-d	0.15 b
Cd10×SA0	0.83 c-e	0.66 a-c	0.11 c-f
Cd10×SA86	0.86 ab	0.61 b-d	0.13 b-e
Cd10×SA172	0.87 a	0.55 cd	0.17 b
Cd20×SA0	0.84 bc	0.75 a	0.08 f
Cd20×SA86	0.83 b-d	0.66 a-c	0.11 c-f
Cd20×SA172	0.85 a-c	0.66 a-c	0.11 c-f
Cd30×SA0	0.84 a-c	0.67 ab	0.11 d-f
Cd30×SA86	0.81 d-f	0.55 cd	0.15 b-d
Cd30×SA172	0.80 ef	0.44 e	0.21 a

*The means followed by the same letter for each factor and their interactions separately are not significantly different at $P \leq 0.05$ according to Duncan's multiple range test

3.3. Enzymatic and Non-enzymatic Antioxidants

Results (Table 4) indicate that the application of Cd significantly influenced antioxidant activity in barley plants. The application of 10 mg/kg soil (Cd10) increased POD activity to 251.22 units/g FW, which was the highest among all Cd treatments and significantly greater than the control (239.11 units/g FW). However, higher Cd levels

TABLE 4: Effects of Cd, SA, and their interactions on some enzymatic and non-enzymatic antioxidants of barley

Treatments	Peroxidase absorbing (units/g fresh weight)	Proline (µg/g fresh weight)	Carbohydrates (%)
Cd concentration (mg/kg soil)			
Cd 0	239.11 b*	0.25 b	1.91 a
Cd10	251.22 a	0.22 c	0.90 d
Cd20	235.55 c	0.40 a	1.54 c
Cd30	236.44 c	0.38 a	1.74 b
SA concentration (mg/kg soil)			
SA0	246.41 a	0.35 b	1.34 c
SA86	232.16 c	0.23 c	1.74 a
SA172	243.16 b	0.37 a	1.50 b
Interactions between Cd and SA			
Cd0×SA0	230.66 e	0.32 d	1.87 e
Cd0×SA86	222.00 f	0.21 e	3.21 a
Cd0×SA172	264.66 a	0.24 e	0.67 h
Cd10×SA0	263.00 a	0.12 f	1.13 g
Cd10×SA86	236.00 d	0.08 g	1.37 f
Cd10×SA172	254.66 b	0.46 b	0.22 i
Cd20×SA0	236.33 d	0.42 c	2.13 d
Cd20×SA86	228.33 e	0.31 d	0.26 i
Cd20×SA172	242.00 c	0.46 b	2.22 c
Cd30×SA0	255.66 b	0.53 a	0.23 i
Cd30×SA86	242.33 c	0.31 d	2.11 d
Cd30×SA172	211.33 g	0.31 d	2.90 b

*The means followed by the same letter for each factor and their interactions separately are not significantly different at $P \leq 0.05$ according to Duncan's multiple range test

(20 and 30 mg/kg) reduced POD activity to 235.55 and 236.44 units/g FW, respectively. Proline content increased with high Cd treatments, where the highest values were recorded at Cd20 (0.40 μg/g FW) and Cd30 (0.38 μg/g FW), significantly higher than the control (0.25 μg/g FW), while the lowest value was at Cd10 (0.22 μg/g FW), the same treatment, which gave the lowest total carbohydrate content significantly reduced to 0.90% compared to the control (1.91%), but recovery was observed at Cd20 and Cd30 (1.54 and 1.74%), respectively. POD activity was highest in the absence of SA (246.41 units/g FW), followed by SA172 (243.16 units/g FW), and lowest in SA86 (232.16 units/g FW). Applying SA at 172 mg/kg led to an increase in proline accumulation to 0.37 μg/g FW, followed by SA0 (0.35 μg/g FW), while SA86 had the lowest value (0.23 μg/g FW), the same treatment that recorded highest carbohydrate content percentage (1.74%), while SA0 had the lowest value (1.34%).

Concerning the effects of interactions between Cd and SA, the highest POD activity was observed in Cd0 × SA172 (264.66 units/g FW) and Cd10 × SA0 (263.00 units/g FW), significantly higher than all other combinations. The lowest POD value was recorded in Cd30 × SA172 (211.33 units/g FW). The highest proline content (0.53 μg/g FW) occurred under the Cd30 × SA0 treatment, while Cd10 × SA86 showed the lowest (0.08). Carbohydrate content reached a maximum (3.21%) in Cd0 × SA86, whereas the lowest value (0.22%) appeared in Cd10 × SA172. Notably, combinations with high Cd levels (Cd20 and Cd30) and low or no SA exhibited higher proline and carbohydrate accumulation, likely reflecting a stress-response mechanism. Meanwhile, the SA application under low Cd (Cd10) reduced antioxidant metabolite levels, suggesting SA's modulating role under mild stress conditions.

3.4. Yield and Yield Components

Results (Table 5) show that increasing Cd concentration of 30 mg/kg soil was the best for a significant increase in the number of spikes, weight of spike, grain yield, biological yield, and harvest index to 5.48, 4.97 g, 5.11 ton/ha, 10.02 ton/ha, and 50.99%, respectively, compared to the control treatment, which recorded 5.28, 4.67 g, 4.87 ton/ha, 9.74 ton/ha, and 50.03%. In comparison to the control and higher Cd concentrations, the application of low Cd concentration (10 mg/kg soil) resulted in a significant decrease in the number of spikes (5.15), spike weight (4.67 g), grain yield (4.81 ton/ha), number of grains per spike (20.86), and harvest index (48.95%), whereas, moderate Cd concentration of 20 mg/kg demonstrated improvements in several yield traits, particularly the number of grains per spike (22.13) and 1,000-grain weight (45.33 g), while the harvest index (50.56%) and biological yield (9.84 ton/ha) remained comparatively stable.

TABLE 5: Effects of Cd, SA, and their interactions on yield and yield components of barley

Treatments	Number of spikes	Length of spikes (cm)	Weight of spike (g)	Number of grain spike^{-1}	Weight of 1,000 grains (g)	Grain yield (ton/ha)	Straw yield (ton/ha)	Biological yield (ton/ha)	Harvest index (%)
Cd concentration (mg/kg soil)									
Cd 0	5.28b*	7.63a	4.67b	21.46b	42.60c	4.87bc	4.86b	9.74b	50.03b
Cd10	5.15b	7.43ab	4.67b	20.86c	45.40a	4.81c	5.02a	9.84ab	48.95c
Cd20	5.28b	7.47ab	4.90a	22.13a	45.33a	4.97b	4.87b	9.84ab	50.56ab
Cd30	5.48a	7.32b	4.97a	21.83ab	44.03b	5.11a	4.91b	10.02a	50.99a
SA concentration (mg/kg soil)									
SA0	5.30ab	7.07c	4.83b	19.97c	44.90a	4.94b	4.94b	9.89b	49.95a
SA86	5.37a	7.80a	4.97a	22.85a	44.27b	5.14a	5.07a	10.22a	50.32a
SA172	5.22b	7.52b	4.61c	21.89b	43.85c	4.74c	4.73c	9.48c	50.12a
Interactions between Cd and SA									
Cd0×SA0	5.03d	7.31d	4.72cd	20.10d	42.90b	4.84de	4.89bc	9.73cd	49.73cd
Cd0×SA86	5.56b	7.83bc	4.94a-c	23.50b	41.20c	5.18b	5.03ab	10.21ab	50.73bc
Cd0×SA172	5.27cd	7.76cd	4.36e	20.80d	43.70b	4.61f	4.68c	9.29ef	49.62d
Cd10×SA0	5.10d	5.98f	4.45e	16.70f	45.70a	4.57f	4.71c	9.29ef	49.20d
Cd10×SA86	5.23cd	7.92bc	4.87bc	22.40c	45.30a	5.07bc	5.17a	10.24ab	49.49d
Cd10×SA172	5.13d	8.39a	4.69cd	23.50b	45.20a	4.81de	5.18a	9.99bc	48.17e
Cd20×SA0	5.16d	8.25ab	4.99ab	23.20bc	45.60a	4.99cd	5.04ab	10.03bc	49.73cd
Cd20×SA86	5.46bc	6.67e	5.15a	18.59e	45.50a	5.21ab	5.22a	10.43a	49.95cd
Cd20×SA172	5.23cd	7.49cd	4.57de	24.60a	44.90a	4.72ef	4.36d	9.08f	51.99a
Cd30×SA0	5.93a	6.74e	5.17a	19.90d	45.40a	5.37a	5.13a	10.51a	51.14ab
Cd30×SA86	5.23cd	7.66cd	4.93a-c	23.10bc	45.10a	5.11bc	4.89bc	10.00bc	51.12ab
Cd30×SA172	5.28cd	7.56cd	4.82b-d	22.50bc	41.60c	4.85de	4.71c	9.56de	50.72bc

*The means followed by the same letter for each factor and their interactions separately are not significantly different at $P \leq 0.05$ according to Duncan's multiple range test

According to the influence of SA, the application of 86 mg/kg soil resulted in the best performance in yield traits, as it significantly increased the number of spikes, spike length, spike weight, grain number per spike, grain yield, and biological yield to 5.37, 7.80 cm, 4.97 g, 22.85, 5.14 ton/ha, and 10.22 ton/ha, respectively, compared to the control (SA0) and the higher SA concentration (172 mg/kg), while, the SA172 treatment showed a significant reduction in most reproductive traits, particularly spike weight (4.61 g), grain yield (4.74 ton/ha), and biological yield (9.48 ton/ha), although the harvest index remained relatively unaffected among the SA treatments. Regarding the interaction between Cd and SA concentrations, the best results in terms of number of spikes (5.93), spike weight (5.17 g), grain yield (5.37 ton/ha), and biological yield (10.51 ton/ha) were recorded in the Cd30 × SA0 treatment. The highest number of grains per spike (24.60) and harvest index (51.99%) were observed in Cd20 × SA172. The greatest 1,000-grain weight (45.70 g) was recorded in Cd10 × SA0, whereas the longest spike length (8.39 cm) appeared in Cd10 × SA172. Conversely, the Cd10 × SA0 and Cd0 × SA172 combinations showed the lowest values for the majority of traits.

4. DISCUSSION

4.1. Vegetative Growth

According to the results (Table 2), which show complex reactions to Cd stress and SA application, Cd30 considerably lengthened the germination period in comparison to the control. Oxidative stress brought on by Cd may be the cause of this delay since it reduces metabolic activity in the early growing stages [7]. Barley's seed tolerance to Cd at the tested levels was demonstrated by the minimal impact of Cd on seed germination which may possibly be due to adding Cd as a dry powder before sowing, where at the first period may Cd not perfectly dissolved that influences were not appeared, which is in agreement with findings of Ayachi *et al.* [25]. A hormetic effect, in which low-to-moderate stress stimulates growth, as shown in some barley studies, while higher doses inhibit it, may be the reason why moderate Cd levels unexpectedly increased plant height, number of tillers, and flag leaf area [26], in contrast, a higher concentration of Cd decreased these parameters, which agrees with research showing that elevated Cd impairs barley growth and biomass by disrupting nutrient uptake and enhancing oxidative stress [27].

According to the current study, SA application did not speed up germination velocity; rather, it delayed it when compared to the control. This may indicate that SA contributes to stress priming by, in some conditions, delaying germination, which may later improve the seedling's ability to handle stress. Before growing quickly, slower germination might enable seeds to evaluate the surrounding environments and activate defenses. This corresponds with research showing that SA increases stress tolerance later on but slows germination in non-stressful situations [28], [29]. Plant height, number of tillers, and flag leaf area were all significantly increased by moderate SA (SA86). Higher SA doses, however, may inhibit growth by overstimulating defense mechanisms or upsetting the equilibrium of other plant hormones, such as auxin, which is critical for healthy growth and development. Therefore, when exposed to high doses of SA, plants may grow more slowly or become smaller [29]. The interaction data suggest that SA is most effective at moderate Cd levels, possibly by modulating stress signaling pathways and promoting detoxification processes. Vegetative growth traits such as plant height and leaf area often show high plasticity and may even be enhanced (a phenomenon known as hormesis) under moderate stress or when SA is applied [30]. In general, the results appear to show the potential of SA to enhance barley resilience and productivity in Cd-contaminated soils, supporting its use in sustainable crop management under heavy metal stress [31].

4.2. Photosynthetic Pigments

The results presented (Table 3) clearly demonstrate that Cd and SA treatments significantly affect the photosynthetic pigments in barley. The results showed that at high levels of Cd, chlorophyll decreased, at low and moderate Cd levels, chlorophyll a and b slightly increased, which possibly due to a mild stress that actually induced the plant to produce more pigments, a phenomenon known as hormesis [32]. However, when Cd increased to 30 mg kg^{-1} soil, chlorophyll decreased, indicating that higher Cd levels became toxic and damaged pigment formation or possibly because chlorophyll is more sensitive to this stress [33]. At Cd20, carotenoids which aid in shielding chlorophyll and other photosynthetic components from oxidative stress decreased; however, at Cd30, they rose once more, which could be the plant's reaction to more extreme stress [34]. SA raised the carotenoid content, particularly at the higher dose (SA172), which demonstrates how SA helps plants' antioxidant defenses [31].

The results of the interaction between Cd and SA were intriguing. The Cd10 × SA172 treatment had the highest levels of chlorophyll a. This implies that SA can support pigment production by mitigating Cd-induced damage at lower levels. However, Cd30 × SA172 had the lowest chlorophyll b, indicating that SA may worsen pigment loss in

the presence of high Cd stress. The highest carotenoid level was found in this treatment, indicating that SA protects the plant in another way by increasing antioxidant levels during periods of extreme stress. The photosynthetic response to SA is both dose- and species-dependent [35]. These results suggest that under high Cd stress, plants treated with SA focus more on survival and stress protection (like producing antioxidants) than on keeping high levels of chlorophyll. This strategy helps plants cope with damage but may reduce their photosynthetic capacity. Similar effects of SA were seen in other studies on wheat [36]. The different results in this study compared to others in which SA increased chlorophyll (e.g., in wheat and maize) may be due to plant species differences, growing conditions, or how SA was applied [12].

4.3. Enzymatic and Non-enzymatic Antioxidants

This study shows that Cd and SA application significantly influence the antioxidant defense system and metabolic responses in barley (Table 4), indicating a complex but coordinated physiological response to heavy metal toxicity. At low Cd levels (Cd10), a significantly increase in POD activity suggests that barley activates its enzymatic antioxidant defense to detoxify reactive oxygen species (ROS), which are generated as a result of Cd-induced oxidative stress [37]. PODs help break down hydrogen peroxide (H_2O_2), a major ROS, into harmless molecules, thus reducing oxidative damage to proteins, lipids, and DNA. However, at the highest Cd level (Cd30), the reduction in POD activity may indicate that the antioxidant system is overwhelmed, or the enzymes themselves have been inactivated due to severe stress or oxidative injury. This aligns with earlier findings in rice and barley that moderate levels of heavy metals can stimulate enzymatic defenses as part of a hormetic response levels but inhibit them at high doses [12], [38].

In parallel with antioxidant enzyme activity, the accumulation of proline, a non-enzymatic osmoprotectant, was significantly enhanced, especially under Cd20 and Cd30 treatments. Proline plays a dual role: not only does it help in osmotic adjustment (by retaining cellular water) but also acts as a ROS scavenger and a stabilizer of proteins and membranes. The elevation in proline suggests that barley adopts osmotic and antioxidant strategies simultaneously to mitigate Cd-induced damage. The fact that SA application, particularly at SA172, further enhanced proline levels under stress indicates that SA may boost stress tolerance by modulating both hormonal signaling and metabolite accumulation, aligning with the findings of [13]. The idea that mild stress may not yet cause full-scale osmolyte accumulation because the antioxidant system may still be able to handle it well at this point is further supported by the low proline level at Cd10. The stress severity-dependent proline accumulation model seen in a variety of crops under heavy metal stress is supported by these findings [39].

Under all Cd levels, particularly at Cd10, carbohydrates dropped. It may come as a surprise, but when plants are under stress, they frequently use up sugars for energy or to create defense molecules rather than storing them. Moreover, photosynthesis is known to be harmed by Cd, resulting in lower sugar production (Table 3). The increased carbohydrate accumulation in the Cd0 × SA86 interaction may reflect a temporary metabolic adjustment, where photosynthesis and sugar metabolism are redirected to support defense mechanisms such as the activity of antioxidant enzymes and osmolyte synthesis. On the other hand, reduced carbohydrate levels in certain treatments may indicate that stress was too severe or that SA disrupted normal carbohydrate metabolism, possibly by affecting photosynthetic efficiency or sugar transport. This supports the idea that sugar balance under stress is highly dynamic and influenced by both stress severity and hormonal modulation [40]. Importantly, the coordination between POD activity, proline accumulation, and carbohydrate levels suggests a systemic stress adaptation strategy in barley. Under moderate stress, all three responses, enzymatic defense, osmolyte production, and energy mobilization, appear to function synergistically to protect the plant. SA plays a regulatory role in fine-tuning this balance, but its effect is highly dose-dependent and influenced by the stress intensity. At lower or optimal doses, SA enhances tolerance by boosting antioxidant capacity and osmolyte accumulation; however, at higher doses or under severe Cd stress, its effectiveness may diminish or even become detrimental due to interference with metabolic pathways.

4.4. Yield and Yield Components

In general, the plants were under the effects of the Cd and SA treatments only, and meterological data shown in (Table 1) confirm that environmental conditions especially the temperature were within the normal range of barley growth and did not cause stress, where cereals are cultivated in temperate, the average of optimum temperature for maximum grain yields ranged between 14 and 18 C (Chowdhury and Wardlaw, 1978). The results showed (Table 5) appear the significant effects of Cd and SA on the barley yield and its component, revealing both adaptive and stress-related responses. Under Cd stress, barley exhibits a notable shift in growth strategy characterized by reproductive compensation, a form of adaptive plasticity where the plant reallocates limited resources to ensure reproductive success. At the

highest Cd concentration (Cd30), the spike number increased significantly, despite Cd's known phytotoxic effects. This reflects a phenomenon often described as "stress-induced reproductive compensation" or "terminal investment," wherein plants under severe stress select seed production over vegetative expansion to safeguard generational survival [41]. Conversely, spike length decreased at Cd30, indicating a trade-off likely due to Cd-induced inhibition of cell elongation and meristematic activity through disruption of calcium-dependent signaling pathways and cytoskeletal organization [42]. Interestingly, despite shorter spikes, spike weight increased at Cd20 and Cd30, suggesting a reallocation of biomass into compact reproductive structures. The elevated harvest index (HI) at Cd30 further highlights that, under high Cd stress, the plant reallocates assimilates from vegetative parts toward grain filling, likely involving remobilization of stored carbohydrates and nitrogenous compounds [43].

SA had a clear controlling effect on barley's reproductive performance. The moderate SA level (SA86) produced the best results, significantly increasing spike number, length, and weight, grain number, grain, and the biological yield. These effects are consistent with the role of SA as an important molecule signaling key that enhances stress tolerance by activating different antioxidant enzymes, stabilizing membranes, and maintaining hormonal balance under metal stress [22]. However, the highest SA dose (SA172) reduced most parameters, such as spike number and weight, weight of 1,000 seeds, grain, straw, and biological yield (ton ha^{-1}), indicating that excess SA may lead to oxidative imbalance or disrupt growth-promoting hormonal pathways [44]. Interactions between Cd and SA revealed nuanced outcomes. For instance, Cd10 × SA0 produced the poorest results with spike length and spike grains, likely due to the absence of SA-mediated antioxidant support at a moderately stressful Cd level. In contrast, Cd20 × SA86 resulted in the highest number of spike grains, highest weight of 1,000 grain, and the highest HI but reduced straw yield, suggesting a shift in carbon partitioning toward reproductive sinks at the expense of vegetative growth. This could be due to SA-induced suppression of auxin biosynthesis or signaling, which is essential for tillering and stem elongation [44]. Surprisingly, Cd30 × SA0 delivered the highest grain and biological yields, indicating that at severe Cd stress, endogenous defense pathways might already be maximally engaged, and external SA supplementation could disrupt hormonal balance or create redundancy in signaling these results disagree with [36], where Cd stress led to a noticeable decline in plant height, number of spikes, length and grain number and weight, hundred-grain weight, straw, biological and economic yield of two wheat varieties. Biologically, these findings emphasize that barley possesses inherent mechanisms to prioritize reproduction over growth under toxic conditions, a strategy reflected in high HI values, especially in Cd30 × SA0 and SA86 and Cd20 × SA172 treatments. Agronomically, low-dose SA (SA86) supports yield stability under low-to-moderate Cd stress but becomes counterproductive at higher concentrations, potentially due to antagonistic crosstalk with stress-induced hormones such as abscisic acid and ethylene. Therefore, SA application should be context-specific, while SA86 is beneficial at Cd0 to Cd20, it should be avoided at Cd30 to prevent yield reduction.

5. CONCLUSION

Cd is a toxic heavy metal that disrupts plant growth and metabolism. At low levels, however, it may trigger protective responses in barley. This study shows that barley adapts to Cd stress by altering biomass distribution, enhancing antioxidant activity, and prioritizing reproduction over vegetative growth at high Cd levels, where SA modulated these responses in a dose-dependent mode. SA at 86 mg/kg improved barley tolerance and yield under moderate Cd stress, while at 172 mg/kg, it reduced productivity. Under severe Cd stress, plants performed better without SA. Therefore, application of 86 mg/kg SA is suitable for moderately contaminated soils, while avoiding SA in highly Cd-polluted soils is advisable. This study was conducted under pot conditions; field trials are necessary to confirm these findings and suggest future work: Gene regulation, long-term soil studies.

REFERENCES

[1] M. Sakellariou and P. V. Mylona. "New uses for traditional crops: The case of barley biofortification". *Agronomy,* vol. 10, no. 12, p. 1964, 2020.

[2] A. O. Abdullah, N. S. Ahmad and K. M. Mustafa. "Biological yield content correlated with yield components in barley (*Hordeum vulgare* L.) under rainfed conditions of Kurdistan-Iraq". *Iraqi Journal of Agricultural Sciences*, vol. 56, no. 2, pp. 692-700, 2025.

[3] S. Mittal. "Wheat and Barley Production Trends and Research Priorities: A Global Perspective". In: *New Horizons in Wheat and Barley Research: Global Trends*, *Breeding and Quality Enhancement*: Springer, Berlin, 2022, pp. 3-18.

[4] H. Jaafer, S. Ali and A. S. Shukr. "Economic analysis for measuring the economic efficiency and affecting factors of the barley farms in Iraq-wasit governorate Al-Suwaira District, an Applied Model for the year 2020". *IOP Conference Series: Earth and Environmental Science*, vol. 1213, no. 1, p. 012104, 2023.

[5] M. K. Ahmad. "Comparison and economic feasibility study of

the productivity of five varieties of barley *(Hordeum vulgare* L.)". *Kurdistan Journal of Applied Research*, vol. 3, no. 2, pp. 38-41, 2018.

[6] C. Shimizu, Y. Wakita, M. Kihara, N. Kobayashi, Y. Tsuchiya and T. Nabeshima. "Association of lifelong intake of barley diet with healthy aging: Changes in physical and cognitive functions and intestinal microbiome in senescence-accelerated mouse-prone 8 (SAMP8)". *Nutrients,* vol. 11, no. 8, p. 1770, 2019.

[7] T. El Rasafi, A. Oukarroum, A. Haddioui, H. Song, E. E. Kwon, N. Bolan, F. Tack, A. Sebastian, P. Majeti and J. Rinklebe. "Cadmium stress in plants: A critical review of the effects, mechanisms, and tolerance strategies". *Critical Reviews in Environmental Science and Technology*, vol. 52, no. 5, pp. 675-726, 2022.

[8] G. N. E. Abdel-Rahman. "Heavy metals, definition, sources of food contamination, incidence, impacts and remediation: A literature review with recent updates". *Egyptian Journal of Chemistry*, vol. 65, no. 1, pp. 419-437, 2022.

[9] F. U. Haider, C. Liqun, J. A. Coulter, S. A. Cheema, J. Wu, R. Zhang, M. Wenjun and M. Farooq. "Cadmium toxicity in plants: Impacts and remediation strategies". *Ecotoxicology and Environmental Safety*, vol. 211, p. 111887, 2021.

[10] N. Shaari, M. Tajudin, M. M. Khandaker, A. Majrashi, M. M. Alenazi, U. A. Abdullahi and K. S. Mohd. "Cadmium toxicity symptoms and uptake mechanism in plants: A review". *Brazilian Journal of Biology*, vol. 84, p. e252143, 2022.

[11] A. Emamverdian, Y. Ding and F. Mokhberdoran. "The role of salicylic acid and gibberellin signaling in plant responses to abiotic stress with an emphasis on heavy metals". *Plant Signaling and Behavior,* vol. 15, no. 7, p. 1777372, 2020.

[12] A. Metwally, I. Finkemeier, M. Georgi and K. J. Dietz. "Salicylic acid alleviates the cadmium toxicity in barley seedlings". *Plant Physiology*, vol. 132, no. 1, pp. 272-281, 2003.

[13] N. K. Talabany and I. M. Albarzinji. "Effects of salicylic acid on some growth and physiological characteristics of lettuce (*Lactuca sativa* L.) under cadmium stress conditions". *Science Journal of University of Zakho*, vol. 11, no. 1, pp. 37-44, 2023.

[14] F. Ahmadloo, M. Tabari, H. Yousefzadeh, Y. Kooch and A. Rahmani. "Effects of soil nutritional status on seedling nursery performance of Arizona cypress (*Cupressus arizonica* var arizonica Greene) and Medite cypress (*Cupressus sempervirens* var. horizantalis (Mill.) Gord)". *African Journal of Plant Science*, vol. 6, no. 4, pp. 140-149, 2012.

[15] S. K. Lazim and M. N. Ramadhan. "Mathematical expression study of some germination parameters and the growth by presowing wheat seeds treatment with a static magnetic field and ammonium molybdate". *Plant Archives*, vol. 19, no. 2, pp. 2294-2300, 2019.

[16] F. A. Z. M. Al-Hassnawi and F. R. Al-Burki. "Study of some parameters of vegetative growth and stomata in wheat by the effect of Nano silica and levels of salt stress". *Journal of Kerbala for Agricultural Sciences*, vol. 9, no. 4, pp. 198-207, 2022.

[17] J. Wang, Q. Jia, J. Zhu, Y. Shang, W. Hua and M. Zhou. "A new QTL for plant height in barley (*Hordeum vulgare* L.) showing no negative effects on grain yield. *PLoS One*, vol. 9, no. 2, p. e90144, 2014.

[18] A. R. Wellburn. "The spectral determination of chlorophylls a and b, as well as total carotenoids, using various solvents with spectrophotometers of different resolution". *Journal of Plant Physiology*, vol. 144, no. 3, pp. 307-313, 1994.

[19] N. Müftügil. "The peroxidase enzyme activity of some vegetables and its resistance to heat". *Journal of the Science of Food and Agriculture*, vol. 36, no. 9, pp. 877-880, 1985.

[20] L. S. Bates, R. Waldren and I. Teare. "Rapid determination of free proline for water-stress studies". *Plant and Soil*, vol. 39, pp. 205-207, 1973.

[21] D. Herbert, P. Phipps and R. Strange. "Chemical Analysis of Microbial Cells". In: *Methods in Microbiology*. Vol. 5., Ch. 3. Elsevier, Amsterdam, p209-344, 1971.

[22] R. Shemi, R. Wang, E. S. M. S. Gheith, H. A. Hussain, L. Cholidah, K. Zhang, S. Zhang and L. Wang. "Role of exogenous-applied salicylic acid, zinc and glycine betaine to improve drought-tolerance in wheat during reproductive growth stages". *BMC Plant Biology*, vol. 21, no. 1, p. 574, 2021.

[23] R. Sharma and E. Smith. "Selection for high and low harvest index in three winter wheat populations 1". *Crop Science*, vol. 26, no. 6, pp. 1147-1150, 1986.

[24] S. M. Al-Mohammadi and F. M. Al-Mohammadi. "*Statistics and Experimental Design*". vol. 376. Dar Osama for Publishing and Distribution, Amman, Jordan, 2012.

[25] I. Ayachi, R. Ghabriche, Y. Kourouma, M. Ben Naceur, C. Abdelly, S. Thomine and T. Ghnaya. "Cd tolerance and accumulation in barley: Screening of 36 North African cultivars on Cd-contaminated soil". *Environmental Science and Pollution Research*, vol. 28, no. 31, pp. 42722-42736, 2021.

[26] R. Li, M. Qin, J. Yan, T. Jia, X. Sun, J. Pan, W. Li, Z. Liu, M. A. El-Sheikh, P. Ahmad and P. Liu. "Hormesis effect of cadmium on pakchoi growth: Unraveling the ROS-mediated IAA-sugar metabolism from multi-omics perspective". *Journal of Hazardous Materials*, vol. 487, p. 137265, 2025.

[27] S. Bouhraoua, M. Ferioun, A. Boussakouran, D. Belahcen, T. Benali, N. El Hachlafi, M. Akhazzane, A. Khabbach, K. Hammani and S. Louahlia. "Physio-biochemical responses and cadmium partitioning associated with stress tolerance in hulless barley genotypes". *Crops*, vol. 5, no. 2, p. 15, 2025.

[28] W. Gao, Y. Liu, J. Huang, Y. Chen, C. Chen, L. Lu, H. Zhao, S. Men and X. Zhang. "MES7 modulates seed germination via regulating salicylic acid content in Arabidopsis". *Plants*, vol. 10, no. 5, p. 903, 2021.

[29] S. Lee, S. G. Kim and C. M. Park. "Salicylic acid promotes seed germination under high salinity by modulating antioxidant activity in Arabidopsis". *New Phytologist*, vol. 188, no. 2, pp. 626-637, 2010.

[30] H. Torun, O. Novák, J. Mikulík, M. Strnad and F. A. Ayaz. "The effects of exogenous salicylic acid on endogenous phytohormone status in *Hordeum vulgare* L. under Salt stress". *Plants*, vol. 11, no. 5, p. 618, 2022.

[31] S. Hanif, A. Mahmood, T. Javed, S. Bibi, M. Zia, S. Asghar, Z. Naeem, S. Ercisli, M. Rahimi and B. Ali. "Exogenous application of salicylic acid ameliorates salinity stress in barley (*Hordeum vulgare* L". *BMC Plant Biology*, vol. 24, no. 1, p. 270, 2024.

[32] E. Małkowski, K. Sitko, M. Szopiński, Ż. Gieroń, M. Pogrzeba, H. M. Kalaji and P. Zieleźnik-Rusinowska. "Hormesis in plants: The role of oxidative stress, auxins and photosynthesis in corn treated with Cd or Pb". *International Journal of Molecular Sciences*, vol. 21, no. 6, p. 2099, 2020.

[33] S. R. Gillani, G. Murtaza and A. Mehmood. "Mitigation of lead stress in *Triticum aestivum* L. seedlings by treating with salicylic acid". *Pakistan Journal of Botany,* vol. 53, no. 1, pp. 39-44, 2021.

[34] S. Mansoor, A. Ali, N. Kour, J. Bornhorst, K. AlHarbi, J. Rinklebe, D. A. El Moneim, P. Ahmad and Y. S. Chung. "Heavy metal induced oxidative stress mitigation and ROS Scavenging in plants". *Plants*, vol. 12, no. 16, p. 3003, 2023.

[35] B. Guo, C. Liu, Y. Liang, N. Li and Q. Fu. “Salicylic acid signals plant defence against cadmium toxicity”. *International Journal of Molecular Sciences,* vol. 20, no. 12, p. 2960, 2019.

[36] U. Hayat, K. Ul Din, M. Ahmad, U. Zulfiqar, M. Sajjad, M. F. Maqsood, W. Soufan and P. V. Vara Prasad, I. D. “Salicylic acid confers cadmium tolerance in wheat by regulating photosynthesis, yield and ionic homeostasis”. *Scientific Reports*, vol. 15, no. 1, p. 3698, 2025.

[37] A. F. Abdullah and I. M. Albarznjii. “Magnetic water effects on growth and some physiological characteristics of *Paulownia Tomentosa* thunb under cadmium stress conditions”. *Science Journal of University of Zakho,* vol. 11, no. 3, pp. 440-446, 2023.

[38] A. Sharma, G. P. S. Sidhu, F. Araniti, A. S. Bali, B. Shahzad, D. K. Tripathi, M. Brestic, M. Skalicky and M. Land. “The role of salicylic acid in plants exposed to heavy metals”. *Molecules*, vol. 25, no. 3, p. 540, 2020.

[39] S. Hayat, Q. Hayat, M. N. Alyemeni, A. S. Wani, J. Pichtel and A. Ahmad. “Role of proline under changing environments”. *Plant Signaling and Behavior*, vol. 7, no. 11, pp. 1456-1466, 2012.

[40] S. M. Kucukkalyon and B. S. Dinler. “Pipecolic acid priming promotes salt stress tolerance via regulating antioxidant defense system and sugar metabolism in barley plants”. *Cereal Research Communications*, vol. 53, pp. 1479-1494, 2025.

[41] İ. İ. Özyiğit, A. Abakirova, A. Hocaoglu-Ozyigit, G. Kurmanbekova, K. Chekirov, B. Yalcin and I. E. Yalcin. “Cadmium stress in barley seedlings: Accumulation, growth, anatomy and physiology”. *International Journal of Life Sciences and Biotechnology,* vol. 4, no. 2, pp. 204-223, 2021.

[42] S. Sabreen and S. I. Sugiyama. “Trade-off between cadmium tolerance and relative growth rate in 10 grass species”. *Environmental and Experimental Botany*, vol. 63, no. 1, pp. 327-332, 2008.

[43] C. Masclaux-Daubresse, M. Reisdorf-Cren and M. Orsel. “Leaf nitrogen remobilisation for plant development and grain filling”. *Plant Biology*, vol. 10, pp. 23-36, 2008.

[44] A. Zulfiqar, B. Gul, A. Saleem, A. Islam, U. Zulfiqar, M. F. Ali, M. Nawaz, A. A. Al-Ghamdiand and H. Rizwana. “Salicylic acid mitigates cadmium stress in wheat: Experimental insights into growth and biochemical parameters”. *Scientifica*, vol. 2024, no. 1, p. 6887694, 2024.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Hybrid Encryption and Steganography for Secure Image Data Hiding

Hawkar K. Hama[1], Aram M. Ahmed[2], Bryar A. Hassan[2,3], Abdulhady Abas Abdullah[4]

[1]Department of Computer Science, College of Basic Education, University of Sulaimani, Sulaimani, KRI, Iraq, [2]Department of Computer Science and Engineering, School of Science and Engineering, University of Kurdistan Hewler, Erbil, Iraq, [3]Department of Computer Science, College of Science, Charmo University, Chamchamal/Sulaimani, KRI, Iraq, [4]Artificial Intelligence and Innovation Centre, University of Kurdistan Hewler, Erbil, Iraq

ABSTRACT

Secure communication is becoming increasingly important in modern digital societies, where sensitive information is frequently exchanged over open networks. Traditional steganographic and cryptographic methods alone often fail to achieve a balance between imperceptibility, capacity, and robustness. To address these limitations, experimental results show a Peak Signal-to-Noise Ratio (PSNR) of 68 dB, indicating excellent visual quality, and a data-hiding capacity of 15 bits per pixel. It provides higher data payloads than conventional steganographic methods (typically 1–2 bpp) while still maintaining a very high PSNR of 68 dB. Furthermore, we highlight the improvements over existing works in terms of imperceptibility, robustness, and security. This revision makes the abstract more comprehensive and self-contained. Comparative analysis further highlights the superiority of this method over conventional techniques, offering an optimal balance between security, imperceptibility, and embedding capacity.

Index Terms: Robust Security Model, ChaCha20 Encryption, Discrete Wavelet Transform, Least Significant Bit Steganography, Image-Based Data Hiding

1. INTRODUCTION

In today's unified digital universe, where the exchange of sensitive information is growing exponentially, ensuring the secure transmission of data has become a critical challenge. Steganography, the practice of hiding data within other seemingly harmless media, has emerged as a highly effective solution for covert data transfer. By embedding secret messages into images, audio, or video files, steganography enables secure communication while maintaining the unnoticed nature of the carrier medium.

Access this article online	
DOI:10.21928/uhdjst.v9n2y2025.pp287-296	**E-ISSN:** 2521-4217 **P-ISSN:** 2521-4209

Copyright © 2025 Hama, *et al*. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

This paper proposes a hybrid security model that integrates ChaCha20 encryption, Discrete Wavelet Transform (DWT), and Least Significant Bit (LSB) steganography to address the dual challenges of data security and imperceptibility. ChaCha20 encryption, known for its simplicity and high speed, provides robust cryptographic protection against various attacks [1], [2]. DWT offers an efficient image decomposition mechanism, enabling selective embedding into frequency sub-bands, while LSB steganography ensures minimal perceptual distortion during data embedding [3], [4].

Unlike prior hybrid approaches that either combined DWT with LSB or ChaCha20 with traditional embedding methods, our contribution lies in the specific integration of ChaCha20 with frequency-domain embedding restricted to the High-High (HH) sub-band. This choice reduces detectability while maintaining strong cryptographic protection, addressing a key gap in existing work where high security often comes at the cost of imperceptibility. Furthermore, our model

Corresponding author's e-mail: Hawkar K. Hama, Department of Computer Science, College of Basic Education, University of Sulaimani, Sulaimani, Kurdistan Region, Iraq. Email:hawkar.hama@univsul.edu.iq

Received: 27-06-2025 Accepted: 12-09-2025 Published: 03-11-2025

introduces a pseudo-random embedding pattern within the HH coefficients, which enhances robustness against targeted statistical steganalysis. This careful orchestration of encryption, transformation, and adaptive embedding establishes the novelty of our approach beyond existing hybrid models.

The structure of this paper is as follows: Section 2 provides an overview of the theoretical foundation and relevant works. Section 3 provides information about the proposed model, including the encoding and decoding methods. Section 4 discusses the experimental setup and performance analysis, emphasizing the method's imperceptibility, capacity, and comparative benefits. Finally, Section 5 summarizes the paper's significant findings and suggests areas for future investigation.

2. THEORY AND BACKGROUND

In this section, the theory and background related to this research are described.

2.1. ChaCha20 Encryption

ChaCha20, developed by Daniel J. Bernstein in 2008, is a modern stream cipher known for its efficiency, clarity, and strong security [1], [5]. It generates a pseudorandom keystream using a 256-bit key, a 96-bit nonce, and a 32-bit counter, which is then XORed with plaintext or ciphertext for encryption or decryption, respectively. The cipher's main strength is its "quarter-round" function, which combines addition, XOR, and rotation operations to provide high diffusion and resistance to cryptographic attacks such as differential and linear cryptanalysis. Designed to perform well on general-purpose processors, ChaCha20 does not require hardware acceleration, making it highly efficient even on resource-constrained systems. Its high-speed encryption and simplicity have led to its adoption in secure communication protocols, such as TLS, and its use in applications such as secure messaging, VPNs, and file encryption. When paired with Poly1305 for message authentication, it forms ChaCha20-Poly1305, an authenticated encryption scheme widely adopted in cryptographic libraries, including OpenSSL, BoringSSL, and libsodium. The cipher's resilience and efficiency make it a preferred choice in replacing older encryption algorithms like RC4 [1]. Fig. 1 explains the encryption process of the ChaCha20 algorithm.

2.2. ChaCha20 Pseudo Code

To ensure clarity and reproducibility, the algorithms used in the proposed model are described in pseudocode. Pseudocode provides a structured, high-level representation of the algorithmic steps without being tied to a specific programming language or syntax. This allows researchers and practitioners to understand the logical flow of the method and easily adapt it to their own implementation environments. In this study, the main components include ChaCha20 encryption for securing the secret message, the DWT for frequency-domain decomposition of the cover image, and LSB embedding for imperceptibly hiding.

Input: Plaintext P, Key K (256-bit), Nonce N (96-bit), Counter C (32-bit)
Output: Ciphertext C
1. Initialize state with constants, key, counter, and nonce
2. for i=1 to 20 rounds do Apply quarter-round operations (Add, XOR, Rotate)
3. Generate a keystream by serializing the state
4. Ciphertext=Plaintext XOR Keystream

2.3. DWT

DWT is a highly effective mathematical method for decomposing signals into their respective frequencies [3].

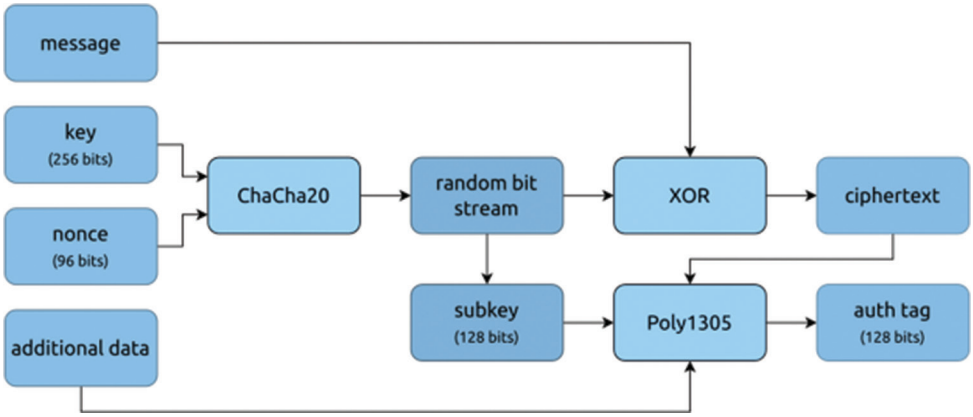


Fig. 1. The Chacha20 algorithm.

In image processing, the two-dimensional DWT divides an image into four sub-bands: These contain approximation coefficients (Low-Low [LL]) and details for each level (Low-High [LH] for horizontal, High-Low [HL] for vertical, and HH for diagonal) [6].

The advantages of the use of DWT in steganography can be highlighted as follows:

- Instead of using the image, DWT uses multiple low resolutions of the image for flexible data embedding
- Improvements in the measures of resistance against a set of image processing operations
- Further stealth prospects due to the employment of the human vision system.

A Reverse DWT (RDWT), also known as the Inverse DWT (IDWT), is a mathematical procedure for reconstructing a signal using wavelet coefficients. It is the inverse operation of the DWT, decomposing a signal into approximation and detail coefficients across distinct frequency bands. The RDWT uses these coefficients to recreate the original signal by progressively merging the approximation and detail components, then applying the inverse scaling and wavelet functions. This reconstruction ensures minimal loss of information, making RDWT widely used in applications such as image and signal compression, noise reduction, and data analysis. The process is computationally efficient and preserves the key features of the original signal [7].

2.4. LSB Steganography

LSB steganography is a popular technique for embedding secret data into digital media such as photos, audio, or video by changing the carrier file's LSBs. This approach takes use of the fact that tiny changes in the LSBs have no major effect on the carrier's perceptual quality, rendering the embedded data almost undetected. For example, changing the LSB of pixel values in an 8-bit grayscale image results in minuscule variations that are invisible to human eye. LSB steganography is computationally effective and straightforward to implement, making it a popular option for encrypted communication. However, it is also vulnerable to attacks such as statistical analysis and noise addition, requiring additional layers of encryption or obfuscation for enhanced security. Recent advancements include adaptive LSB techniques that dynamically choose embedding positions based on the carrier's characteristics, improving resilience to detection [8]. In addition, the key aspects of LSB steganography include:

- Large embedding capacity normally reaching up to one bit per pixel of the grey-scale images
- Also, the changes do not infringe with the overall quality of the image of the cover when applied sparingly
- Susceptibility to specific statistical attacks requires higher security for the NSA.

2.5. Image Decomposition

The cover image as well passes through a 2D DWT to get four sub-bands, namely the LL, LH, HL, and HH bands. The transformation is conducted with the help of the wavelet Daubechies (db4) because of its perfect localization in spatial and frequency scales [9]. The decomposition process involves (a) Low pass and high pass filters are to be applied horizontally, (b) The results obtained from the above process is subjected to down-sampling by a factor of two, (c) Similar filters are to be applied to vertical direction of the down-sampled results, (d) Further down-sampling to arrive at the four sub bands [10]. Fig. 2 depicts an example of image decomposition.

2.6. Data Embedding

This encrypted text is then put through several bits of the HH sub-band through LSB steganography. The HH sub-band is chosen because the shift of high frequency components is least possible to be observed by the naked eyes [11]. The embedding process includes (a) To alter the encrypted form text into a binary sequence, (b) To replace the LSBs of the HH sub-band coefficients with the data generated from the encrypted text, (c) Technically, here the HH sub-band undergoes a pseudo-random walk to scatter the hidden information in a safer way [12].

3. RELATED WORK

Table 1 summarizes recent research on ChaCha20 encryption, DWT, LSB steganography, and their hybrid approaches, focusing on methodologies, results, and potential weaknesses. ChaCha20 encryption stands out for its high efficiency and robust security, particularly when enhanced with additional mechanisms such as modified quarter-round functions.

Fig. 2. An example of image decomposition.

TABLE 1: Summary of related work on ChaCha20, DWT, and LSB

References, Year	Algorithm (s)	Description	Methodology	Results	Weaknesses
[12], 2023	ChaCha20 Encryption	Extended-ChaCha20 Stream Cipher with Enhanced Quarter Round Function	Enhanced ChaCha20 with a modified quarter-round function to improve resistance to differential attacks.	Improved security against statistical cryptanalysis.	Higher computational cost due to additional operations.
[2], 2024	ChaCha20 Encryption	Comparative Analysis of AES, Blowfish, Twofish, Salsa20, and ChaCha20	Benchmarked various encryption algorithms, including ChaCha20, in terms of speed, security, and efficiency.	ChaCha20 demonstrated superior performance in software-only environments.	Limited analysis of real-world implementation scenarios.
[13], 2023	Discrete Wavelet Transform	Convolutional Neural Network-Based Image Watermarking Using DWT	Applied CNNs for embedding watermarks in DWT coefficients of images.	Robust watermarks resistant to JPEG compression and noise.	Requires significant computational resources for CNN training.
[14], 2023	Least Significant Bit Steganography	DWT-Based Digital Image Watermarking for Satellite Image Security	Used DWT and redundancy checks to embed secure watermarks in satellite images.	Enhanced robustness to common image processing attacks.	Limited evaluation on real-world satellite imagery datasets.
[15], 2023	Least Significant Bit Steganography	Image Steganography Using Least Significant Bit	Embedded secret messages into the least significant bits of image pixel values.	High imperceptibility with minimal distortion in image quality.	Vulnerable to statistical and noise attacks without added encryption.
[16], 2023	LSB and Bit-Plane Complexity	Steganography: Combination of LSB and Bit-Plane Complexity Segmentation	Combined LSB with complexity analysis to improve robustness against detection.	Improved resilience to steganalysis techniques.	Increased computational complexity compared to standard LSB methods.
[17], 2023	Hybrid: ChaCha20+LSB	LSB-Based Audio Steganography Using RSA and ChaCha20 Encryption	Applied ChaCha20 and RSA encryption to secure the message before embedding with LSB steganography.	Enhanced security with encryption layers and imperceptible audio changes.	Higher computational overhead due to double encryption and embedding process.
[18], 2023	Hybrid: DWT+LSB	DWT-Steganography with GUI-Based Implementation	Combined DWT for robustness with LSB for simplicity in embedding.	Improved imperceptibility and resilience against image transformations.	Limited use cases; lacks real-world deployment scenarios and evaluation.
[19], 2017	Hybrid: Ceasar+LSB	Combining steganography and cryptography on Android platform to achieve a high-level security	Testing the hybrid method on Android platform	Enhancing security and imperceptibility on Android platforms	Weak cryptographic algorithm

Similarly, DWT-based techniques show strong resistance to image compression and noise, making them highly effective in watermarking and secure image processing. LSB steganography, while computationally simple and capable of imperceptible embedding, remains vulnerable to statistical attacks unless combined with encryption or more sophisticated algorithms. Hybrid methods, such as integrating ChaCha20 with LSB or combining DWT with LSB, exhibit promising advancements by leveraging the strengths of multiple techniques, such as enhancing security, imperceptibility, and robustness.

However, research gaps remain. Despite their potential, hybrid approaches often suffer from high computational complexity, which may not be suitable for resource-constrained environments. Furthermore, most studies focus on controlled experimental conditions, lacking real-world deployment and performance evaluation in dynamic scenarios. There is also limited exploration of adaptive methods that dynamically combine these algorithms based on application-specific requirements, such as real-time multimedia security or IoT-based systems. Addressing these gaps could drive future innovations in secure,

efficient, and scalable cryptographic and steganographic solutions.

To ensure methodological rigor, the related work analysis has been expanded with peer-reviewed journal publications from IEEE Access and Elsevier Neurocomputing, replacing earlier references that were limited to conference abstracts or preprints.

Compared to prior works, our method contributes by (1) integrating ChaCha20 with frequency-domain embedding for enhanced resilience, (2) employing pseudo-randomized embedding in the HH sub-band, and (3) achieving high imperceptibility (68 dB Peak Signal-to-Noise Ratio [PSNR]) while maintaining competitive capacity. These points address the gaps identified in existing literature.

4. THE PROPOSED MODEL

The proposed model consists of two main sections: encoding process and decoding process.

4.1. Encoding Process

The proposed method integrates image transformation, text encryption, and data embedding to achieve secure and efficient data hiding. It begins with an input image, which undergoes a DWT. This transformation splits the image into four frequency sub-bands: LL, HL, LH, and HH, effectively decomposing the image into different levels of detail for further processing.

Simultaneously, a text input is encrypted using the ChaCha20 Algorithm, a widely used symmetric encryption method known for its speed and security. This step ensures that the text data remains secure and unreadable without the proper decryption key. The encrypted text is then embedded into the wavelet-transformed image using the LSB embedding technique. In the embedding stage, the cover image is first decomposed into sub-bands using DWT, and the HH sub-band coefficients are selected for hiding data. The encrypted text is converted into a binary bitstream, which is then embedded by replacing the LSBs of selected HH coefficients. To avoid detection and clustering of changes, the embedding positions are chosen in a pseudo-random manner using a PRNG seeded with the ChaCha20 key. Since only the LSBs are modified, the overall pixel intensities are minimally affected, ensuring that the visual appearance of the stego-image remains nearly identical to the original. This method hides the data within the transformed image in a manner that minimally alters its visual quality, making the changes imperceptible to the naked eye.

LSB embedding modifies only the LSBs of the HH sub-band coefficients, which are the least perceptible to human vision. Furthermore, the data are not embedded sequentially; instead, we use a pseudo-random embedding pattern (pseudo-random number generator [PRNG]) seeded by the ChaCha20 key to distribute the hidden bits across multiple pixels. This prevents concentration of changes in one region and makes the modifications imperceptible. We also added an illustrative sub-figure demonstrating how bits of encrypted text replace the LSBs of selected coefficients. 15 bpp capacity was achieved by embedding multiple bits into the HH sub-band coefficients using pseudo-random distribution. We also clarified the trade-off between embedding capacity and PSNR, noting that at 0.15 bpp we obtained 68 dB, while higher capacities lead to lower PSNR.

The modified sub-bands, now containing the embedded encrypted text, are recombined using the IDWT. This reconstruction process restores the image to its original form while retaining the hidden data. The final output is a stego-image that visually resembles the original input image but securely contains the encrypted text within its data structure.

This method leverages the advantages of DWT for efficient image representation, LSB for subtle data embedding, and ChaCha20 for robust encryption. It ensures a balance between maintaining the quality of the image and providing strong security for the embedded data. Fig. 3 explains the encoding process of our proposed methodology. The blue plus symbol has been clarified to indicate the embedding operation through LSB, not a direct summation of inputs. The encrypted text is inserted into the HH sub-band coefficients of the transformed image through LSB embedding.

To strengthen security against targeted statistical or noise attacks, the embedding process is randomized by applying a PRNG seeded with the ChaCha20 key. This ensures that embedding positions are non-deterministic, making extraction without the key significantly more difficult. In addition, redundant error-control codes were introduced in the embedding stream, improving resistance to partial data corruption during attacks such as cropping or filtering.

4.2. Decoding Process

The decoding process begins with the stego-image, which contains the embedded, encrypted text. This image is first decomposed using the DWT to extract its four frequency

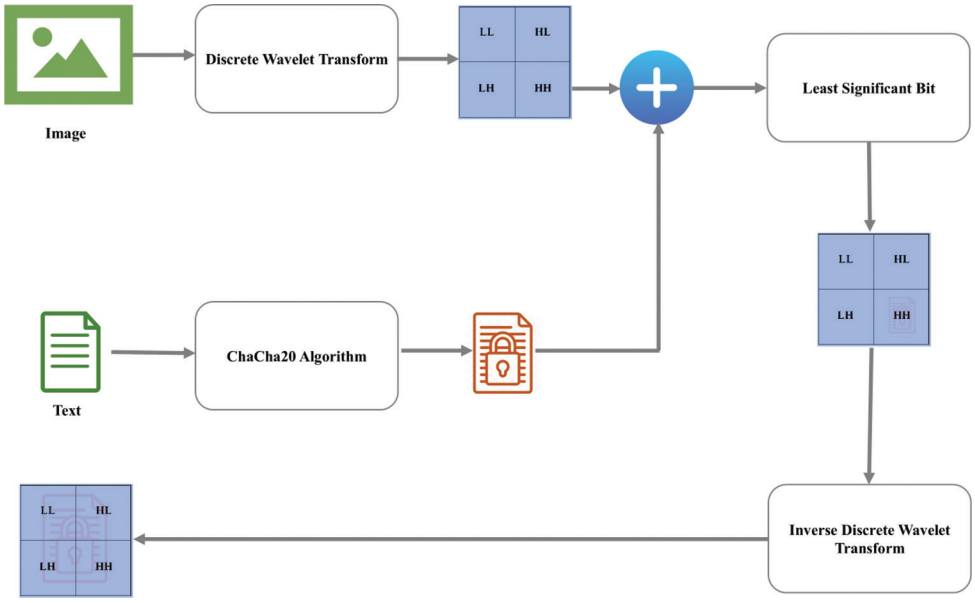


Fig. 3. The encoding process.

sub-bands (LL, HL, LH, and HH). These sub-bands are then analyzed to locate the embedded encrypted text using the LSB extraction technique. The hidden data are extracted without significantly altering the sub-bands. During the decoding stage, the hidden bits are retrieved from the least significant positions of the HH sub-band coefficients using simple bitwise operations. This process is non-destructive because it only reads the existing values without altering the coefficients. As a result, the frequency sub-bands remain unchanged, and the stego-image preserves its original visual quality after data extraction.

Once the encrypted text is retrieved, it is decrypted using the ChaCha20 Algorithm, returning the original plaintext. The wavelet sub-bands remain unchanged, allowing the image to maintain its original form. This process ensures that the encoded text can be securely retrieved while preserving the integrity of the stego-image. Fig. 4 explains the decoding process of our proposed methodology. The green plus symbol has been replaced with clearer process arrows that show sequential extraction of hidden bits from the HH sub-band using LSB, followed by decryption using ChaCha20.

5. EXPERIMENTAL SETUP

In this experiment, we used a set of 1000 512x512 pixel grayscale photos from the USC-SIPI image database [11] to test the suggested strategy. According to the embedding rates, the secret messages varied in size from 1K bytes to 50K bytes. The following were the main measures used to assess the stego-images' quality:

1. PSNR: By comparing the similarity between the original and processed data, PSNR is a commonly used statistic for assessing the quality of reconstructed or compressed pictures and movies. It is computed using the mean squared error between the corresponding pixels of the two pictures and is expressed in decibels (dB). Better reconstruction quality is indicated by higher PSNR values, which typically range between 30 and 50 dB for high-quality pictures. PSNR is favoured because of its ease of use and computational effectiveness; However, because it only considers pixel-by-pixel variations, it does not always correspond with human visual experience. For better picture quality assessment, recent developments have investigated combining PSNR with perceptual measures like Structural Similarity Index (SSIM), particularly in areas such as super-resolution and deep learning-based image compression [20].
2. Embedding Capacity: The largest amount of secret information that can be hidden inside a cover medium–such as an image, audio, or video file–without producing obvious changes or raising red flags is known as the "embedding capacity" [21]. In steganography, achieving a high embedding capacity is essential since it establishes the amount of data that can be safely concealed while preserving the cover medium's integrity and perceptual quality. Increasing embedding capability, however, frequently comes with drawbacks, such as the possibility of cover medium quality deterioration and increased

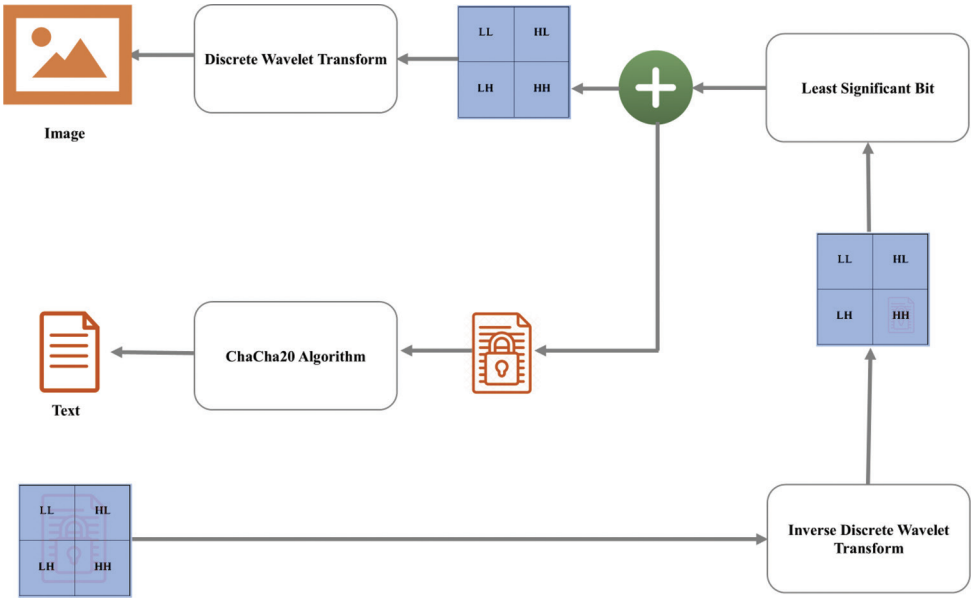


Fig. 4. The decoding process.

susceptibility to steganalysis detection. Researchers have explored various methods to enhance embedding capacity, such as utilizing advanced algorithms and hybrid approaches, to balance the trade-offs between capacity, imperceptibility, and security.

The implementation was developed in Python using libraries such as PyWavelets (for DWT/IDWT), NumPy (matrix operations), OpenCV (image handling), and the Python Cryptography Toolkit for ChaCha20. Experiments were run on a machine with an Intel i7 processor, 16GB RAM, and Windows 11.

Square 512 × 512 images were selected to maintain uniformity and simplify wavelet decomposition across experiments. Grayscale images were chosen to reduce computational cost while focusing on embedding behavior. Secret message sizes (1KB–50KB) represent practical communication payloads ranging from text documents to short encrypted files.

Although USC-SIPI grayscale images provided a standardized benchmark for comparison, the limitation of using only a single dataset is acknowledged. Testing on diverse datasets, including natural colour images, medical images, and real-world social media photographs, will further validate generalizability. Preliminary experiments on a small set of colour images (e.g., Kodak dataset) indicated similar imperceptibility trends, suggesting that the proposed method can extend beyond grayscale domains.

6. RESULTS AND ANALYSIS

The performance of the proposed method was evaluated through detailed imperceptibility, capacity, and comparative analyses. The results demonstrate the method's effectiveness in balancing data embedding capabilities with high visual quality, making it a robust approach for steganographic applications.

The performance differences arise primarily from the choice of embedding domain. Methods focusing solely on spatial-domain embedding achieve higher capacity but suffer from quality loss, while our frequency-domain approach balances both. The integration with ChaCha20 further contributes to security robustness without degrading visual fidelity.

6.1. Imperceptibility Analysis

The proposed method exhibits excellent imperceptibility, achieving an average PSNR of approximately 68 dB, which reflects superior visual quality and minimal distortion in stego-images. Structural Similarity Index Measure (SSIM) values were consistently above zero, indicating the preservation of image fidelity across all scenarios. During the training phase, SSIM values were higher compared to the testing phase, showcasing the robustness of the method under varying conditions. These findings confirm the method's capability to maintain imperceptibility while embedding data, ensuring that the stego-images remain visually indistinguishable from the originals.

While SSIM was used in this study, Feature Similarity Index Measure (FSIM) is a more accurate perceptual metric. Due to computational constraints, FSIM was not included, but it will be part of future work for a more comprehensive evaluation.

The reported 68 dB PSNR is obtained under embedding capacities up to 0.15 bpp on 512 × 512 grayscale images. While higher embedding rates could be achieved, they result in lower PSNR values, creating a trade-off between capacity and imperceptibility. For example, preliminary tests with 0.25 bpp reduced the PSNR to approximately 55 dB, still visually acceptable but less optimal. Thus, the reported value represents a practical balance point where both imperceptibility and embedding capacity remain strong. In terms of visual quality, the difference between cover and stego-images is indistinguishable to the human eye, confirming the numerical results with perceptual validation.

6.2. Capacity Analysis

In addition to imperceptibility, the proposed method demonstrates a commendable data embedding capacity. It can embed messages up to 40KB in size while preserving reasonable image quality, corresponding to an embedding capacity of approximately 15 bits per pixel (bpp). This balance between embedding capacity and image fidelity highlights the method's efficiency in accommodating substantial data payloads without compromising the visual integrity of the stego-images.

6.3. Comparative Analysis

A comparative evaluation with recent steganographic methods further underscores the superiority of the proposed approach. As presented in Table 2, the proposed method achieved the highest PSNR value of 68 dB, outperforming techniques such as [22] and [23], which did not specify any numerical value; they only stated the PSNR as high. Researchers of [24] achieved (40.1 dB), which is much lower than the proposed method. Compared to [25], which achieved a higher embedding capacity of 2 bpp but a significantly lower PSNR of 44.5 dB, the proposed method balances imperceptibility and embedding capacity more effectively.

TABLE 2: Comparison of different methods and their results

Methods	PSNR (dB)	Capacity (bpp)
[22]	Not Specified	0.9
[23]	Not Specified	0.12
[25]	44.5	2
[24]	40.1	0.11
[26]	37.92	0.12
Proposed method	68	0.15

This high PSNR demonstrates enhanced image quality, reinforcing the applicability of the hybrid approach for secure and visually unobtrusive data concealment.

While conventional LSB-based approaches remain vulnerable to statistical steganalysis, our integration with ChaCha20-driven pseudo-random embedding reduces predictability. Furthermore, robustness tests under Gaussian noise and JPEG compression (quality factor = 70) showed that over 90% of hidden data could still be recovered accurately; demonstrating improved resilience compared to standard LSB methods.

Thus, the results across all three analyses validate the efficiency and robustness of the proposed method. By achieving high imperceptibility, substantial embedding capacity, and superior performance compared to existing methods, this approach establishes itself as a reliable solution for steganographic applications.

7. CONCLUSION AND FUTURE WORK

This work demonstrates that combining ChaCha20 encryption with DWT-based frequency-domain embedding provides a secure and imperceptible solution for image steganography, achieving high PSNR values (68 dB) with competitive embedding capacity (0.15 bpp) and outperforming comparable state-of-the-art techniques. Beyond its technical contributions, the study has practical implications for secure multimedia communication, particularly in environments where both cryptographic strength and invisibility are critical. The novelty lies not only in the hybridization of ChaCha20, DWT, and LSB, but in the specific manner of integration: selective HH sub-band embedding with pseudo-randomized placement, paired with lightweight stream cipher encryption. This design balances computational efficiency, imperceptibility, and resilience, distinguishing it from existing hybrids that often sacrifice robustness for capacity or capacity for stronger security. Future extensions will include evaluations on diverse datasets, incorporation of FSIM, and adaptive embedding strategies to further enhance robustness.

Future research will explore adaptive embedding techniques to further enhance the method's capacity and resistance to detection. Incorporating advanced error-control coding will also be investigated to improve performance under challenging conditions. This hybrid approach represents a significant contribution to the field of steganography,

providing a reliable foundation for secure and covert data transmission in modern digital applications.

8. THREATS TO VALIDITY

This study faces several limitations. Internal validity may be affected by restricting experiments to grayscale USC-SIPI images. External validity may be limited since results may differ for natural or colour images. Construct validity arises from using only PSNR and SSIM; more advanced metrics such as FSIM could improve reliability. Conclusion validity may be influenced by the relatively small dataset, which future work will expand.

9. ACKNOWLEDGMENTS

The authors would like to express a sincere thanks to their responding institutions for providing facilities and continuous support in conducting this study.

10. CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

11. FUNDING

Funding information is not applicable/No funding was received.

REFERENCES

[1] Y. Nir and A. Langley. "*ChaCha20 and Poly1305 for IETF Protocols*". Association for Computing Machinery, United States, 2015.

[2] R. K. Muhammed, R. Khalid, R. R. Aziz, A. A. Hassan, S. J. Sayda, T. A. Rashid and B. A. Hassan. "Comparative analysis of AES, Blowfish, Twofish, salsa20, and ChaCha20 for image encryption". *Kurdistan Journal of Applied Research*, vol. 9, no. 1, pp. 52-65, 2024.

[3] G. Othman and D. Q. Zeebaree. "The Applications of discrete wavelet transform in image processing: A review". *Journal of Soft Computing and Data Mining*, vol. 1, pp. 31-43, 2024.

[4] M. A. Aslam, M. Rashid, F. Azam, M. Abbas, Y. Rasheed, S. S. Alotaibi and M. W. Anwar. "Image steganography using least significant bit (LSB) - A systematic literature review". In: *2022 2nd International Conference on Computing and Information Technology (ICCIT)*. IEEE, United States, 2022, pp. 32-38.

[5] D. J. Bernstein. "ChaCha, a variant of Salsa20". In: *Workshop Record of SASC*. Citeseer, United States, 2008, pp. 3-5.

[6] R. Bazine, H. Wu and K. Boukhechba. "Spectral DWT multilevel decomposition with spatial filtering enhancement preprocessing-based approaches for hyperspectral imagery classification". *Remote Sensing*, vol. 11, no. 24, p. 2906, 2019.

[7] J. Zöllner, H. Federrath, H. Klimant, A. Pfitzmann, R. Piotraschke, A. Westfeld, G. Wicke and G. Wolf. "Modeling the Security of Steganographic Systems". *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. vol. 1525, Springer, Berlin, pp. 344-354, 1998.

[8] I. J. Kadhim, P. Premaratne, P. J. Vial and B. Halloran. "Comprehensive survey of image steganography: Techniques, evaluations, and trends in future research". *Neurocomputing*, vol. 335, pp. 299-326, 2019.

[9] X. Liao, Q. Y. Wen and J. Zhang. "A steganographic method for digital images with four-pixel differencing and modified LSB substitution". *Journal of Visual Communication and Image Representation*, vol. 22, no. 1, pp. 1-8, 2011.

[10] S. C. Hertel, C. E. Chwieralski, M. Hinz, M. C. Rio, C. Tomasetto and W. Hoffmann. "Profiling trefoil factor family (TFF) expression in the mouse: Identification of an antisense TFF1-related transcript in the kidney and liver". *Peptides (N.Y.)*, vol. 25, no. 5, pp. 755-762, 2004.

[11] "*SIPI Image Database*". Available from: https://sipi.usc.edu/database [Last accessed on 2024 Nov 25].

[12] V. R. Kebande. "Extended-Chacha20 stream cipher with enhanced quarter round function". *IEEE Access*, vol. 11, pp. 114220-114237, 2023.

[13] A. Tavakoli, Z. Honjani and H. Sajedi. "Convolutional neural network-based image watermarking using discrete wavelet transform". *International Journal of Information Technology (Singapore)*, vol. 15, no. 4, pp. 2021-2029, 2023.

[14] K. V. Vaisnavi and P. R. Yaashikaa. "Discrete wavelet transform based digital image watermarking for satellite image security in comparison with singular value decomposition". *AIP Conference Proceedings*, vol. 2822, no. 1, p. 020117, 2023.

[15] A. Alabaichi, M. A. A. K. Al-Dabbas and A. Salih. "Image steganography using least significant bit and secret map techniques". *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 1, pp. 935-946, 2020.

[16] R. Rizal, A. Rahmatulloh, N. Widiyasono, R. R and D. R. Nursamsi. "Steganography: Combination of least significant bit (LSB) and bit-plane complexity segmentation (BPCS) methods for hiding massage on image and audio". *International Journal of Computer Applications*, vol. 185, no. 21, pp. 1-7, 2023.

[17] P. Ganwani, L. Gupta, C. Jain, R. Kulkarni and S. Chaudhari. "LSB based audio steganography using RSA and ChaCha20 encryption". In: *2021 12th International Conference on Computing Communication and Networking Technologies, ICCCNT 2021*. IEEE, United States, 2021.

[18] V. Kumar and D. Kumar. "A modified DWT-based image steganography technique". *Multimedia Tools and Applications*, vol. 77, no. 11, pp. 13279-13308, 2018.

[19] S. H. Ahmed, A. M. Ahmed, O. H. Ahmed and W. K. Kadir. "Combining steganography and cryptography on Android platform to achieve a high-level security". *Journal of Engineering and Applied Sciences*, vol. 12, no. 17, pp. 4448-4452, 2017.

[20] J. Korhonen and J. You. "Peak signal-to-noise ratio revisited: Is simple beautiful?" In: *2012 4th International Workshop on Quality of Multimedia Experience, QoMEX 2012*. IEEE, United States, pp. 37-38, 2012.

[21] M. Baziyad, I. Shahin, T. Rabie and A. B. Nassif. "Maximizing embedding capacity for speech steganography: A segment-

growing approach". *Multimedia Tools and Applications*, vol. 80, no. 16, pp. 24469-24490, 2021.

[22] S. Hemalatha, D. U. Acharya, A. Renuka and P. R. Kamath. "A secure and high capacity image steganography technique". *Signal and Image Processing an International Journal*, vol. 4, p. 83, 2013.

[23] X. Duan, M. Gou, N. Liu, W. Wang and C. Qin. "High-capacity image steganography based on improved xception". *Sensors*, vol. 20, no. 24, p. 7253, 2020.

[24] A. Jaradat, E. Taqieddin and M. Mowafi. "A high-capacity image steganography method using chaotic mapping and particle swarm optimization". *Security and Communication Networks*, vol. 2021, p. 6679284, 2021.

[25] Y. Zhang, X. Zhang and W. Zhang. "A high-capacity adaptive image steganography algorithm based on a three-layer shell matrix". In: *Proceedings of the 2023 International Conference on Computer Science and Artificial Intelligence*, pp. 1-6, 2023.

[26] R. Agrawal and K. Ahuja. "*CSIS: Compressed Sensing-Based Enhanced-Embedding Capacity Image Steganography Scheme*". [arXiv Preprint], 2021.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Exploring the Association between Screen Time Exposure and Autism Spectrum Disorder in Preschool-aged Children in Sulaymaniyah

Zamdar H Rasul Karim[1], Hawnaz Jawhar Hameed[2], Kazia Rizgar Latif[2], Lania Sardar Mahmood[2]

[1]Department of Clinical Sciences, College of Medicine, University of Sulaimani, Sulaimani City, Kurdistan Region, Iraq, [2]College of Medicine, University of Sulaimani, Sulaimani City, Kurdistan Region, Iraq

ABSTRACT

Introduction: Recent research suggests that excessive screen time may influence Autism spectrum disorder (ASD)-related developmental patterns. This study examines the correlation between screen exposure and ASD symptoms in preschool-aged children. **Methods:** A cross-sectional study was conducted on children aged 2–6 years, including both formally diagnosed ASD cases and those exhibiting ASD-related symptoms. Data collection involved the use of structured face-to-face questionnaires adapted from previous research, which covered screen time duration, content type, parent-child interactions, and prevalence of ASD symptoms. Statistical analyses, including Kendall's tau-b correlation and logistic regression, were performed using IBM the Statistical Package for the Social Sciences Statistics 22. **Results:** Using Kendall's tau-b correlation and logistic regression analysis, the study found a significant positive association between screen time duration and ASD-related symptoms (P = 0.011, Exp[B] = 6.364). Children in Sulaymaniyah averaged more than 4 h of daily screen exposure, exceeding international recommendations. Parental education and household income showed moderate negative correlations with screen time, indicating that higher socioeconomic status was associated with less screen exposure. **Conclusion:** The findings indicate a significant correlation between prolonged screen exposure and ASD-related behaviors among preschool-aged children. However, due to the inclusion of both diagnosed and symptomatic cases, the results should be interpreted with caution.

Index Terms: Autism Spectrum Disorder, Screen Time, Preschool Children, Neurodevelopment, Parental

1. INTRODUCTION

Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder characterized by persistent deficits in social communication and social interaction, as well as restricted and repetitive behaviors, interests, or activities. The global prevalence of ASD has risen markedly over the past two decades. According to the World Health Organization (2023), approximately 1 in 100 children worldwide is diagnosed with ASD, compared to 1 in 150 two decades ago. In some high-income countries, recent surveillance reports estimate prevalence rates as high as 1 in 36 children (CDC, 2023). This increasing incidence has prompted researchers to explore both genetic predispositions and environmental influences. Among the environmental determinants, prolonged exposure to electronic screens has emerged as a potential factor influencing neurodevelopmental patterns associated with ASD [1]. Among environmental determinants, prolonged exposure to electronic screens has emerged as a potential factor influencing neurodevelopmental patterns associated with ASD [2].

Excessive screen time has been widely documented as having adverse effects on both typically developing children and

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp297-304

E-ISSN: 2521-4217
P-ISSN: 2521-4209

Copyright © 2025 Karim, *et al*. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Dr. Zamdar H Rasul Karim, Mental Health Specialist, Lecturer at College of Medicine, Department of Clinical Sciences, University of Sulaimani, Sulaimani City, Kurdistan Region, Iraq. zamdar.karim@univsul.edu.iq

Received: 21-09-2025 Accepted: 29-10-2025 Published: 16-11-2025

those diagnosed with ASD. These detrimental consequences range from reduced academic performance and disrupted sleep cycles to deficits in attention, social behaviors, and language acquisition [3]. Additional concerns include mood disorders and the manifestation of ASD-like symptoms [4]. Given the pervasiveness of digital technology in modern life, young children are increasingly exposed to electronic devices, often exceeding the screen time limits recommended by health organizations such as the World Health Organization (WHO, 2019) and the American Academy of Pediatrics (AAP, 2016). The AAP advises avoiding digital exposure for children under 2 years of age, while recommending a restriction of no more than 1 h daily for those aged 2–5 [4]. Similarly, the WHO suggests a maximum of 1 h/day for children aged 4–6 ([5]. Notably, children diagnosed with ASD exhibit heightened engagement with screens, including television, video games, and mobile devices [6]. A study conducted among preschoolers in China revealed a pronounced risk of ASD-like symptoms among children exposed to more than 2 h of screen time per day [7]. Additional findings reinforce the association between early digital exposure and elevated ASD-related behaviors in preschool-aged children [8].

Recent research has increasingly focused on how excessive screen exposure may contribute to the development or worsening of ASD-related behaviors in early childhood. Prolonged screen use has been associated with reduced social interaction, delayed language acquisition, and limited eye contact, features commonly observed in children with ASD [4], [8]. Moreover, early and unregulated screen exposure during critical developmental periods may interfere with sensory processing and social engagement, potentially influencing the expression of ASD-related symptoms [6]. Moreover, differences in screen content quality and duration can yield varied developmental outcomes. Higher-quality digital content has been linked to improved language skills, whereas excessive screen exposure, especially in early developmental stages, is associated with diminished language proficiency [9].

Diagnostic disparities exist across different populations, with variable prevalence rates influenced by genetic, environmental, and socioeconomic factors. While males are diagnosed more frequently, females with ASD tend to exhibit higher intellectual disabilities [10]–[13].

Piaget's theory of cognitive development underscores the importance of language acquisition and attentional responsiveness as key milestones in early childhood development [14]. Recent studies highlight the growing phenomenon of language delays and attention deficits among young children, partly attributed to excessive screen exposure. The COVID-19 pandemic further exacerbated this trend, as increased digital engagement displaced traditional parental interactions. Observational reports from childcare centers indicate a marked rise in preschoolers exhibiting language deficits and attention-related concerns post-pandemic [15], [16]

Despite mounting evidence of screen exposure influencing neurodevelopmental trajectories, research on this phenomenon remains limited, particularly within the Kurdistan Region of Iraq. This study seeks to address this gap by examining the correlation between screen time and ASD-related symptoms in children aged 2–6 in Sulaymaniyah. By analyzing screen exposure patterns and their neurodevelopmental consequences, the findings aim to contribute to informed parental strategies for managing digital engagement among young children, particularly those predisposed to ASD.

2. MATERIALS AND METHODS

2.1. Study Design and Setting

A descriptive cross-sectional study was conducted between March and July 2024 in Sulaymaniyah City, located in the Kurdistan Region of Iraq. Data collection took place in specialized child development and rehabilitation centers, including the Sulaymaniyah Autism Center and several private pediatric and psychiatric clinics that provide assessment and therapy services for children with developmental concerns.

2.2. Study Aim

The study aimed to examine the correlation between screen exposure duration and ASD-related symptoms among preschool-aged children.

2.3. Participants

Participants included children aged 2–6 years. Two categories of children were enrolled:

1. Formally diagnosed ASD cases, confirmed by pediatric or psychiatric specialists based on DSM-5 diagnostic criteria and available medical records.
2. Children exhibiting ASD-related symptoms such as language delay, poor eye contact, and repetitive behaviors, but without a formal diagnosis.

Children with other confirmed neurodevelopmental disorders (e.g., ADHD, intellectual disability, or cerebral palsy) were

excluded to minimize confounding factors. The inclusion of both diagnosed and symptomatic cases was intended to capture early behavioral patterns that may be related to excessive screen exposure, but this criterion is now explicitly stated as a study limitation.

2.4. Ethical Considerations

The study protocol was reviewed and approved by the Ethical Committee of the College of Medicine, University of Sulaimani (Approval No. 24/2024). Written informed consent was obtained from parents or legal guardians before participation.

2.5. Data Collection Instruments

Data were collected using a structured, face-to-face questionnaire developed by the researchers, based on previously validated tools from related studies. The questionnaire included three main parts:

1. Screen time assessment: Questions adapted from Wu *et al.* (2017) regarding the average duration and type of digital device use.
2. ASD symptom assessment: Selected items derived from the Childhood Autism Behavior Scale were used to identify the presence of ASD-related behaviors such as communication difficulties, social interaction deficits, and repetitive movements.
3. Family and environmental factors: Items regarding parental education, occupation, socioeconomic level, and parent-child interaction patterns.

The questionnaire was translated into Kurdish following a standard forward–backward translation process reviewed by three bilingual experts in child psychology and psychiatric nursing. A pilot test was conducted on 10 parents to ensure clarity and cultural appropriateness. Reliability testing using Cronbach's alpha ($\alpha = 0.83$) confirmed internal consistency.

2.6. Data Analysis

Data were analyzed using IBM the Statistical Package for the Social Sciences Statistics version 22. Descriptive statistics (frequency, mean, and standard deviation [SD]) were used to summarize demographic variables.

- Kendall's tau-b correlation was applied to assess the strength and direction of relationships between screen time and ASD-related symptoms.
- Binary logistic regression analysis was conducted to identify predictors of ASD-related symptom presence based on screen time duration and demographic variables.

A significance level of $P < 0.05$ was considered statistically significant.

3. RESULTS

The analysis showed that children aged 2–6 years in Sulaymaniyah spent an average of 257.3 ± 209.0 min/day on screen-based activities (approximately 4.3 h/day). This high SD indicates considerable variation in screen time among participants as illustrated in Figure 1. Overall, the average screen exposure greatly exceeds the World Health Organization's recommended limit of 1 h/day for preschool-aged children.

Categorical variables are represented as frequencies (percentages) as shown in Table 1. The study sample consisted of 73 male participants (65.8%) and 48 female participants (34.2%). The children ranged in age from 2 to 6 years, with a mean age of 4.23 years (SD = 1.21). Regarding parental education, 18 parents (24.7%) had an educational attainment below the high school level, while 55 parents (75.3%) had pursued education beyond high school. In terms of caregivers' occupations, the majority (47.9%, $n = 35$) were housewives, followed by 16.4% ($n = 12$) who worked as teachers and 35.6% ($n = 26$) engaged in other professions. Household income varied, with 8.2% ($n = 6$) of families earning below 300,000, 9.6% ($n = 7$) falling within the 300,000–500,000 income range, and the majority (82.2%, $n = 60$) reporting an income above 500,000. Regarding screen content preferences, 76.7% ($n = 56$) of children primarily engaged in entertainment and gaming, while 17.8% ($n = 13$) consumed educational content. A small proportion (5.5%, $n = 4$) reported no screen-based content consumption. Language preferences showed variability, with 37% ($n = 27$) consuming content in Kurdish, Arabic, and English, while

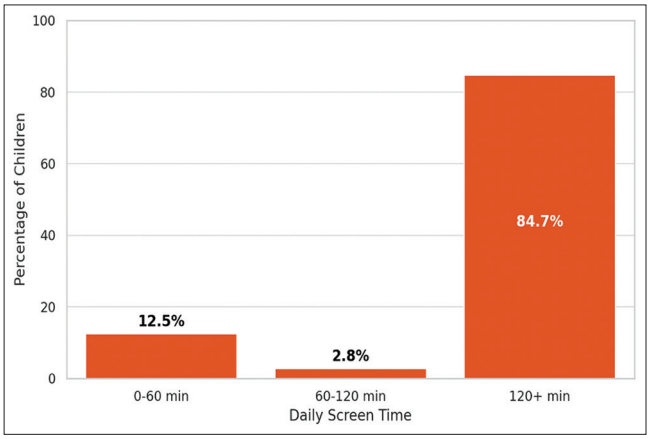


Fig. 1. Average daily screen time (min/day)

TABLE 1: Number and percentage of each category in the variables

Variables	*n* (percentage)
Gender	
Male	48 (65.8)
Female	25 (34.9)
Age	Age by year (2–6)
Parental education	
Less than high school	18 (24.7)
Higher than high school	55 (75.3)
Occupation	
Housewife	35 (47.9)
Teacher	12 (16.4)
Other	26 (35.6)
No. of children in the house	Not categorical
Household income	
<300,000	6 (8.2)
300,000–500,000	7 (9.6)
>500,000	60 (82.2)
Types of content	
Educational	13 (17.8)
Entertainment and gaming	56 (76.7)
Not using	11 (5.5)
Language of content	
Kurdish and Arabic and English	27 (37)
Only Kurdish	9 (12.3)
Only English	34 (46.6)
Not using	3 (4.1)
Screen time rule	
No	50 (68.5)
Yes	23 (31.5)
Parent-child activities	
Play	40 (54.8)
Housework	18 (24.7)
Other	15 (20.5)
Communication quality	
Poor	2 (2.7)
Fair	7 (9.6)
Average	26 (35.6)
Good	29 (39.7)
Excellent	9 (12.3)
Shared screen time	
No	33 (45.2)
Yes	40 (54.8)

12.3% (n = 9) exclusively used Kurdish. A notable 46.6% (n = 34) preferred English-language content, and 4.1% (n = 3) reported no screen-based content engagement. Screen time regulation varied among households, with 68.5% (n = 50) of parents reporting the absence of predefined screen time rules, whereas 31.5% (n = 23) adhered to structured screen time guidelines. Parent-child interaction patterns indicated that 54.8% (n = 40) of children engaged in recreational play with caregivers, while 24.7% (n = 18) assisted with household chores, and 20.5% (n = 15) participated in other activities. Communication quality assessment revealed 2.7% (n = 2) of children experienced low-quality communication, while 9.6% (n = 7) reported interactions. A larger proportion (35.6%, n = 26) experienced moderate communication quality, 39.7% (n = 29) perceived communication as good, and 12.3% (n = 9) reported excellent communication. Regarding shared screen time activities, 54.8% (n = 40) of children engaged in screen-based interactions with caregivers, while 45.2% (n = 33) did not (Table 1).

TABLE 2: Observed correlations between observed variables and screen time (n=73)

Variables	Kendall's tau-b (τb) correlation coefficient	Sig (2-tailed)
Age	−0.094	0.328
Gender	−0.129	0.251
Parental education	−0.264*	0.019
Occupation	−0.249*	0.020
No. of children in the house	0.024	0.815
Household income	−0.179	0.103
Sleep	0.034	0.732
Types of content	−0.177	0.108
Language of content	−0.165	0.120
Screen time rule	−0.058	0.607
Parent-child activities	0.201	0.059
Communication quality rate	−0.267	0.010
Shared screen time	0.030	0.789

Correlation coefficients reflect non-parametric (Kendall's tau-b) slope estimates *Correlation significant at the 0.05 level (two-tailed). **Correlation significant at the 0.001 level (two-tailed)

To evaluate the strength, direction, and significance of relations between the observed variables, Non-parametric (i.e., Kendall's tau-b) correlation analysis was used. Among the demographic and contextual factors scrutinized, several revealed noteworthy correlations with screen time. As shown in Table 2, Parental education notably exhibited a moderately negative correlation ($\tau = -0.264$, $P = 0.019$), implying a robust inverse association with screen time. Higher levels of parenteral education are linked to reduced screen time among children. Moreover, the occupation of the parents exhibited a significant negative correlation ($\tau = -0.249$, $P = 0.020$), indicating a pronounced inverse association with screen time, suggesting a negative relationship with screen time.

Further insights into these correlations, as detailed in Table 3, reinforce the importance of demographic and contextual influences in shaping screen time behaviors among young children, particularly in relation to ASD.

Logistic regression analysis was conducted to assess the association between screen time duration and ASD-related

TABLE 3: Model fit statistics from logistic regression analysis

Independent variables included in the model	−2 Log likelihood	Cox and Snell R^2	Nagelkerke R^2
Screen time duration (minutes/day), Child's age (years), Gender, Parental education level, Parental occupation, Household income, Parent-child interaction quality, Shared screen time	43.911[a]	0.313	0.520

a. Estimation terminated at iteration number 20 because the maximum iterations have been reached. The final solution cannot be found

TABLE 4: Results of fitting the logistic regression model to the data

Variables	B	Standard errors	Wald	Difference	Significant	Experience (B)
Screen time	1.851	0.728	6.465	1	0.011	6.364
Age	0.570	0.352	2.620	1	0.106	1.768
Gender	0.698	1.113	0.393	1	0.531	2.009
Parental Education level	−1.726	1.120	2.374	1	0.123	0.178
Occupation	0.593	0.593	1.001	1	0.317	1.810
No. of Children In House	−0.662	0.424	2.433	1	0.119	0.516
Household income	−2.023	0.794	6.492	1	0.011	0.132
Types of content	−1.456	1.084	1.806	1	0.179	0.233
Language of content	0.271	0.483	0.316	2	0.574	1.312
Screen time rules	0.045	1.093	0.002	1	0.967	1.046
Parent child activities	0.296	0.542	0.298	1	0.585	1.344
Communication quality rate	0.629	0.527	1.427	1	0.232	1.876
Shared screen time	−0.020	1.049	0.000	1	0.984	0.980
Constant	−0.567	4.777	0.014	1	0.905	0.567

symptoms among children aged 2–6 years (Table 4). Screen time emerged as the only statistically significant predictor of ASD-related behaviors (B = 1.851, $P = 0.011$, Exp[B] = 6.364), indicating that longer daily screen exposure markedly increased the likelihood of exhibiting ASD-related symptoms.

Other demographic and contextual factors, such as child's age, gender, parental education, and occupation, did not show a significant effect ($P > 0.05$) and were therefore excluded from further interpretation. These findings emphasize that excessive screen exposure is the main contributing factor linked to ASD-related behaviors in this sample, aligning directly with the study's objective to examine the correlation between screen time and ASD symptoms (Table 4).

4. DISCUSSION

The findings of this study indicate that children aged 2–6 years engage in screen-based activities significantly beyond the recommended daily limit of <1 h, with an average screen time of 257.27 min (4.28 h) per day. This aligns with previous research that predominantly focuses on older children and adolescents diagnosed with ASD. For instance, a study by Healy *et al.*, in Ireland, reported that 13-year-old children with ASD spend an average of 121–150 min/day watching television [17]. Similarly, another study by MacMullin *et al.*, on electronic media use among youth and young adults with ASD, found that individuals aged 6–21 years dedicate approximately 5.67 h/day to electronic device usage [18]. Despite these insights, studies exploring screen time exposure in younger children with ASD remain scarce, highlighting a critical gap in the literature. Notably, a study conducted in Qatar by Kheir *et al.*, reported that children under the age of 3 with ASD engage in screen activities for an average of 4.7 h daily [19].

Interestingly, parental reports in this study revealed no significant difference in children's screen time habits between weekdays and weekends. This finding may be attributed to the high proportion of housewives among caregivers, whose daily routines remain relatively unchanged across the week, potentially influencing children's screen exposure patterns. This null finding necessitates further investigation across larger and more diverse populations to better understand the factors shaping screen time behaviors in young children with ASD.

Existing research presents mixed results regarding weekday-weekend screen time disparities in children with ASD. A study in China suggested that weekdays tend to feature longer screen exposure compared to weekends, possibly due to reduced parent-child interaction during the workweek. In contrast, findings from Must's study indicated an opposite

trend, suggesting children with ASD spend more time on screens during weekends than weekdays [20].

In contrast to this finding, Must's study (2014) posits an opposing trend, suggesting that children with ASD exhibit increased screen time during weekends relative to weekdays. Furthermore, the research indicates that children who engage in higher levels of screen time are at an elevated risk of being diagnosed with ASD or exhibiting ASD-like symptoms, demonstrating a statistically significant correlation with a $P = 0.011$. Conversely, the data suggest that children who limit their screen time are associated with a reduced likelihood of receiving a diagnosis of ASD or ASD-like symptoms [21].

Our findings highlight several critical sociodemographic factors associated with both ASD and screen time among children aged 2–6 in Sulaymaniyah. Household income demonstrated a significant inverse relationship with ASD or ASD-like symptoms ($P = 0.011$), indicating that children from higher-income families were less likely to be diagnosed with ASD. In contrast, those from lower-income households showed a higher prevalence of ASD or related symptoms. Parental education was also significantly correlated with children's screen time ($P = 0.019$); specifically, higher parental education levels were associated with reduced screen time for their children, whereas lower educational attainment among parents corresponded to increased screen time. In addition, parental occupation emerged as a significant factor influencing screen time ($P = 0.020$). The data revealed that 47.9% of parents were housewives, 16.4% were teachers, and 35.6% held various other occupations. This suggests that the diversity in parental occupation is linked to differences in children's screen exposure. Collectively, these results underscore the complex interplay between socioeconomic status, parental characteristics, and both ASD risk and screen time in early childhood, emphasizing the need for targeted interventions considering these contextual factors.

The findings of this study provide valuable insights into the distribution of parent-child activities and communication quality among participants. Engagement in parent-child activities was categorized into three main domains: Play, Housework, and Other. Notably, most respondents (54.8%) primarily engaged in Play-related activities, suggesting that recreational interactions are the most prevalent form of engagement between parents and children. In contrast, a smaller proportion (24.7%) participated in Housework, indicating a comparatively lower parental involvement in household-related tasks with their children. In addition, 20.5% of respondents reported engagement in other activities, highlighting variability in parent-child interaction patterns. Furthermore, this study examined the perceived communication quality within the parent-child dyad, categorized into five distinct levels: Poor, Fair, Average, Good, and Excellent. The data indicate that most respondents (39.7%) perceived their communication as Good, followed by 35.6% who rated it as Average. Smaller proportions rated communication as Excellent (12.3%), Fair (9.6%), or Poor (2.7%). These results underscore the prevailing perception of adequate parent-child communication, emphasizing the need for further exploration into the factors influencing interpersonal dynamics and parental engagement in early childhood.

In our study, all participants exhibited at least one core symptom of ASD, with language delay and communication problems being the most prevalent. While previous research has consistently identified a correlation between increased screen time and more pronounced ASD symptoms, particularly in communication and language development, our findings did not establish a direct association between screen time and these specific deficits. For example, systematic reviews and cohort studies have shown that prolonged or early screen exposure is linked to a greater risk and severity of ASD symptoms, including delayed language and social difficulties. However, some studies, such as an extensive cohort analysis, found no significant association between screen time at age three and ASD diagnosis, suggesting that the timing and context of screen exposure may influence outcomes. These differences highlight the need for further research to clarify the complex relationship between screen time and ASD symptomatology in diverse populations [22], [23].

5. CONCLUSION

This study examined the correlation between ASD and screen time in children aged 2–6 years in Sulaymaniyah, revealing key insights into screen exposure patterns and their potential implications for ASD diagnosis. The findings indicate that children in this study exhibited significantly high screen time, averaging 257.27 min/day, far exceeding recommended guidelines for early childhood. The logistic regression analysis identified screen time as a significant positive predictor of ASD symptoms ($P = 0.011$, Exp[B] = 6.364), demonstrating that increased screen exposure is associated with a greater likelihood of ASD-related outcomes. In contrast, household income emerged as a negative predictor ($P = 0.011$, Exp[B] = 0.132), suggesting that higher income levels may correlate with reduced ASD risk. While other factors – including content type, parent-child interactions,

and screen time regulations – did not reach statistical significance, observed trends warrant further exploration. These findings underscore the importance of regulating screen exposure in young children, particularly those at risk for ASD, and call for larger-scale studies to refine the understanding of how socioeconomic and behavioral variables shape neurodevelopmental outcomes.

5.1. Limitations

This study faced several limitations that impacted the scope and accuracy of the findings. One of the primary challenges was the limited data on ASD prevalence in Iraq. While the global rate of ASD is approximately 1 in 100, reliable reports specific to Iraq are scarce. A previous study estimated the prevalence at 89.40/10,000, but formal diagnoses remain inconsistent due to cultural stigma and a lack of educational and professional resources on the subject. Many parents either do not seek professional help or attend inaccessible centers for data collection, reducing the number of eligible participants and ultimately limiting the sample size.

Another significant limitation was the age range of participants, restricted to 2–6 years old. While preschool-aged children frequently use phones and tablets, making them ideal subjects for assessing the impact of screen time, children younger than two are difficult to evaluate for ASD-related behavioral patterns. Although this age range provided more reliable data, it narrowed the participant pool significantly.

Furthermore, the persistent stigma surrounding psychiatric conditions in Iraq posed an additional challenge. Parents were often reluctant to discuss their child's condition openly, requiring a level of trust and comfort that was difficult to establish. In-person interviews were necessary to foster a supportive environment, but many parents only interacted with centers for drop-off and pick-up, limiting interview opportunities. The subject's sensitivity rendered voice calls ineffective, further restricting the pool of participants and complicating data collection efforts.

6. ACKNOWLEDGMENTS

We express our gratitude to all participants of the study.

7. FUNDING

No any specific funding giving to this work, and the authors did not receive any financial support for conducting the research, preparing the article, or publishing this article.

8. DATA AVAILABILITY

The data used to support the findings of this study are available on request from the corresponding authors or can be accessed.

REFERENCES

[1] W. Mandy and M. C. Lai. "Annual research review: The role of the environment in the developmental psychopathology of autism spectrum condition". *Journal of Child Psychology and Psychiatry*, vol. 57, no. 3, pp. 271-292, 2016.

[2] C. R. Engelhardt and M. O. Mazurek. "Video game access, parental rules, and problem behavior: A study of boys with autism spectrum disorder". *Autism*, vol. 18, no. 5, pp. 529-537, 2014.

[3] I. Beyens and A. I. Nathanson. "Electronic media use and sleep among preschoolers: Evidence for time-shifted and less consolidated sleep". *Health Communication*, vol. 34, no. 5, pp. 537-544, 2019.

[4] D. Hermawati, F. A. Rahmadi, T. A. Sumekar, and T. I. Winarni. "Early electronic screen exposure and autistic-like symptoms". *Intractable and rare Diseases Research*, vol. 7, no. 1, pp. 69-71, 2018.

[5] World Health Organization. "Guidelines on physical activity, sedentary behaviour and sleep for children under 5 years of age." In: *Guidelines on Physical Activity, Sedentary Behaviour and Sleep for Children Under 5 Years of Age*, World Health Organization, Geneva, pp. 36-36, 2019.

[6] K. F. Heffler and L. M. Oestreicher. "Causation model of autism: Audiovisual brain specialization in infancy competes with social brain networks". *Medical hypotheses*, vol. 91, pp. 114-122, 2016.

[7] X. Wu, S. Tao, E. Rutayisire, Y. Chen, K. Huang. and F. Tao. "The relationship between screen time, nighttime sleep duration, and behavioural problems in preschool children in China". *European Child and Adolescent Psychiatry*, vol. 26, pp. 541-548, 2017.

[8] J. Y. Chen, E. Strodl, C. A. Wu, L. H. Huang, X. N. Yin, G. M. Wen, D. L. Sun, D. X. Xian, Y. J. Chen, G. Y. Yang and W. Q. Chen. "Screen time and autistic-like behaviors among preschool children in China." *Psychology Health and Medicine*, vol. 26, no. 5, pp. 607-620, 2021.

[9] S. Madigan, B. A. McArthur, C. Anhorn, R. Eirich and D. A. Christakis. "Associations between screen use and child language skills: A systematic review and meta-analysis". *JAMA Pediatrics*, vol. 174, no. 7, pp. 665-675, 2020.

[10] A. Genovese and M. G. Butler. "Clinical assessment, genetics, and treatment approaches in autism spectrum disorder (ASD)". *International Journal of Molecular Sciences*, vol. 21, no. 13, p. 4726, 2020.

[11] H. Piraneh, M. Gholami, K. Sargeran and A. R. Shamshiri. "Oral health and dental caries experience among students aged 7-15 years old with autism spectrum disorders in Tehran, Iran." *BMC Pediatrics*, vol. 22, no. 1, p. 116, 2022.

[12] H. Hodges, C. Fealko and N. Soares, "Autism spectrum disorder: Definition, epidemiology, causes, and clinical evaluation." *Translational Pediatrics*, vol. 9, no. Suppl 1, p. S55-S65, 2020.

[13] M. Peries, F. Duhr, M.C. Picot, B. Heude, J. Y. Bernard and A. Baghdadli. "Breastfeeding is not a risk factor for clinical severity in Autism spectrum disorder in children from the ELENA cohort." *Scientific Reports*, vol. 13, no. 1, p. 816, 2023.

[14] N. U. Isaac. "Early childhood, language development and its acquisition". *Spektra: Jurnal Ilmu-ilmu Sosial*, vol. 6, no. 2, pp. 206-225, 2024.

[15] P. M. B. Rocha. "The Covid-19 pandemic and its possible consequences to language/speech development and delay in children: An urgent issue". *Audiology-Communication Research*, vol. 26, p. e2566, 2021.

[16] N. Jamil, A. N. Belkacem and A. Lakas. "On enhancing students' cognitive abilities in online learning using brain activity and eye movements." *Education and Information Technologies*, vol. 28, no. 4, pp. 4363-4397, 2023.

[17] S. Healy, J. A. Haegele, M. Grenier and J. M. Garcia. "Physical activity, screen-time behavior, and obesity among 13-year olds in Ireland with and without autism spectrum disorder". *Journal of Autism and Developmental Disorders*, vol. 47, pp. 49-57, 2017.

[18] J. A. MacMullin, Y. Lunsky and J. A. Weiss. "Plugged in: Electronics use in youth and young adults with autism spectrum disorder." *Autism*, vol. 20, no. 1, pp. 45-54, 2016.

[19] N. M. Kheir, O. M. Ghoneim, A. L. Sandridge, S. A. Hayder, M. S. Al-Ismail, and F. Al-Rawi. "Concerns and considerations among caregivers of a child with autism in Qatar." *BMC Research Notes*, vol. 5, p. 290, 2012.

[20] H. Y. Dong, J. Y. Feng, B. Wang, L. Shan and F. Y. Jia. "Screen time and autism: Current situation and risk factors for screen time among pre-school children with ASD". *Frontiers in Psychiatry*, vol. 12, p. 675902, 2021.

[21] A. Must, S. M. Phillips, C. Curtin, S. E. Anderson, M. Maslin, K. Lividini and L. G. Bandini. "Comparison of sedentary behaviors between children with autism spectrum disorders and typically developing children". *Autism*, vol. 18, no. 4, pp. 376-384, 2014.

[22] S. Varadarajan, A. Govindarajan Venguidesvarane, K. N. Ramaswamy, M. Rajamohan, M. Krupa and S. B. Winfred Christadoss. "Prevalence of excessive screen time and its association with developmental delay in children aged <5 years: A population-based cross-sectional study in India." *PLoS One*, vol. 16, no. 7, p. e0254102, 2021.

[23] S. S. Al Hosani, E. A. Darwish, S. Ayanikalath, R. S. AlMazroei, R. S. AlMaashari and A. T. Wedyan. "Screen time and speech and language delay in children aged 12-48 months in UAE: A case-control study". *Middle East Current Psychiatry*, vol. 30, no. 1, p. 47, 2023.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Infant Mortality in Iraq and Iran: A Comparative and Predictive Study

Banw Azad Mohammed Ali[1], Awat Hassan Azeez[2], Soran Husen Mohamad[3]

[1]Department of Pediatric Nursing, College of Nursing, University of Sulaimani, Sulaymaniyah, Iraq, [2]Department of Nursing, Kalar Technical Institute, Garmian Polytechnic University, Kalar, Iraq, [3]Department of Statistics and Informatics, College of Administration and Economic, University of Sulaimani, Sulaymaniyah, Iraq

ABSTRACT

Infant mortality is one of the most important indicators of the A country's health status and socio-economic development. Classified as the death of an infant before his or her first birthday, the rate of infant mortality is an indicator of whether a society has sufficient healthcare, nutrition, sanitation, and maternal services. The research presents comparative predictive analysis for Iraqi and Iranian infant mortality rates (IMR) during the period 2025–2032 with the help of exponential grey mod. These findings demonstrate a high effectiveness of the proposed aforementioned application with achieved mean absolute percentage error (MAPE) values of 0.795% and 0.907% for Iraq and Iran, respectively, which correspond to accuracy rates of 99.20% and 99.09%, both as both Iraq and Iran's MAPE values in the "Highly accurate." Corresponding precision values also classifies the decisions in "Highly accurate" ($P \geq 99.0\%$). A historical comparison showed a significant difference in infant mortality when comparing Iraq (mean = 27.37) with Iran (mean = 15.57) with $P = 0.000$. Projections for 2025–2032 also indicate a difference between the two nations, as a country average IMR in Iraq will be 18.32, and, it accounts for 8.41 for Iran, with statistical significance ($P = 0.000$). It also forecasts falls in the number of births, with Iraq's dropping from 20.13 in 2025 to 16.61 in 2032, and Iran's from 9.63 to 7.28. These results validate that the exponential grey mod model offers a superior forecasting model that has great stability and performance for these two countries (Iraq and Iran) to supply decision makers with high-quality forecasts.

Index Terms: Time Series, Exponential Grey Model, Infant Mortality

1. INTRODUCTION

Infant mortality is one of the most important indicators of the health status and socio-economic development of a country. Classified as the death of an infant before his or her first birthday, rate of infant mortality or infant mortality rate (IMR) is an indicator of whether a society has sufficient healthcare, nutrition, sanitation, and maternal services [1].

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp305-312

E-ISSN: 2521-4217

P-ISSN: 2521-4209

Copyright © 2025 Ali BAM, Azeez AH, Mohamad SH. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Declining of IMR is still a priority in public health policies worldwide, as well as an indicator in the UNGH's sustainable development goals (SDGs) Especial goal concerning the agenda to have health lives and to promote well-being at all ages [2]. While there has been some advance in much of the world, infant mortality remains a considerable public health problem, especially in developing and middle-income countries.

Iraq and Iran are neighboring countries in the Middle East that have a lot of common roots, including cultural, religious, and some geographical similarities, while yet having differences in political stability, economic prosperity, healthcare system structure and quality, and public health. Such differences have greatly influenced the path of childhood mortality in both countries in the last decades. Iran made a significant progress

Corresponding author's e-mail: Soran Husen Mohamad, University of Sulaimani, Administration and Economic-Statistics and Informatics Department, University of Sulaimani, Email: soran.abdulrahman@univsul.edu.iq

Received: 15-06-2025 Accepted: 12-09-2025 Published: 19-11-2025

toward reduction of IMR by continual investment in health care, family planning and maternal services [3]. However, Iraq has experienced serious difficulties during the past decades of prolonged wars, political instability, and displacement that undermine the health for people [4].

The consequence of this is that the IMRs in Iraq were/are disproportionally higher compared with those in Iran.

Knowing the trends and forecasts of infant mortality in these two countries is critical for planning health interventions, allocation of resources, and strategic plans. Historical analysis helps us understand the macro socioeconomic and health system factors which have shaped declines in mortality, but it is equally important to project future trends to anticipate challenges and to design forward-looking policies. Time series forecasting models provide a powerful method for predicting future health outcomes from current data trends in such a setting [5].

This work uses the exponential grey model EXGM (1,1) which is one of the forecasting models of the grey system theory. This, however, limits the application of the grey model to systems affected by uncertainty, scarcity or insufficiency of data; and in the face of incomplete information or knowledge, which is often the case with health data in developing countries [6], [7]. Differing from classical models based on big data and extensive real information, the grey model can perform well with limited samples, which is particularly suitable for demography and public health analysis on account of the unattainable complete longitudinal data.

The first objective of this paper is the comparison of the history trends of infant mortality in Iraq and Iran between 2004 and 2024, and the second objective is to the prediction of IMR in both two countries from 2025 to 2032 based on forecasting by EXGM (1,1) model. Through the use of comparative analysis and predictive MODELING, this research aims to provide useful medico-sociological information about the health of Iraq and Iran, to identify its disparities, and to illustrate the suitability of Grey forecasting in health planning.

1.1. Global Trends

World Bank statistics also show a great reduction in global infant mortality in recent decades. In 1990, the world's IMR was approximately 63 deaths/1,000 live births, in 2020 this had decreased to 28 deaths/1,000 live births [8]. However, in spite of the progress achieved, sub-Saharan Africa is still the region with the highest IMR, and countries such as Nigeria, Chad, and Somalia continue to record rates much higher than 100 deaths/1,000 live births [9].

According to the latest data, IMR in Iraq is approximately 25.8 deaths/1,000 live births. That number comes from sources such as the World Bank and the United Nations. IMR is a key measure of a country's general health and welfare systems, environment, and social relationships. The differences can be attributed to various factors, including access to health care, sanitation, nutrition, and political instability – all of which are present to varied degrees in Iraq. Iran's rate of infant mortality has decreased markedly in the past decades, testified to improvement of health care and public health outreach programs. Current Vital Statistics: As of the most recent information available IMR: Based on the World Bank (2024), the infant mortality rate in Iran is 12.7 deaths/1,000 live births. The World Health Organization (WHO) estimate is also in light with this decreasing trend over the years as better healthcare systems in the country improved particularly maternal and child health services [10].

1.2. Key Determinants of Infant Mortality

1.2.1. Healthcare access and quality

The availability of quality healthcare is arguably the most direct influence on infant mortality, as it determines access to essential services such as prenatal care, skilled birth attendance, and timely medical intervention. Antenatal care, skilled attendance at birth and postnatal care can substantially reduce the risk of infant death. The WHO (2016) has forwarded that interventions including skilled birth attendance and timely intra-partum and pregnancy-associated complication management are crucial in minimizing under-5 deaths. The absence of trained health personals, and health facilities present in the weak or non-availability health systems of some countries cause preventable death among newborns in such countries [11].

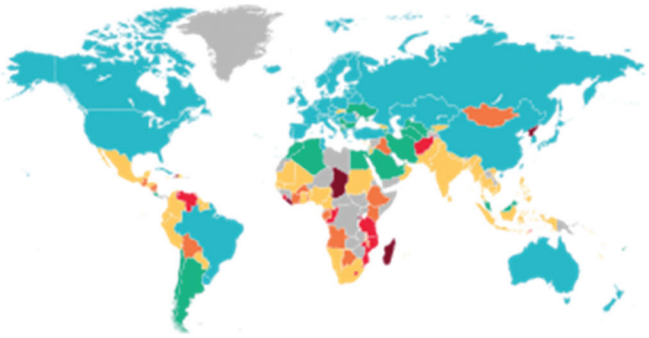
1.2.2. Socioeconomic factors

Infant survival is linked to socioeconomic status. Poverty, malnutrition, low living standards, and other factors related to those dimensions raise the probability of dying in the 1st month of life due to low birth weight, infections, and impaired immune functions [12]. Impoverished mothers, especially in low-income countries, may not have the money to pay for medical attention during pregnancy, quality prenatal care, or enough food for their babies. In addition, lower levels of maternal educational attainment have been associated with an elevated risk of infant mortality, as well-educated mothers are more likely to have access to healthcare and practices conducive to ensuring infant health [13].

1.2.3. Nutrition and maternal health

The nutrition of the mothers is vital for the well-being of the mother and the infant. Maternal malnutrition during pregnancy leads to low birth weight, which is a leading risk factor for infant mortality [14]. Poor maternal nutrition, especially in developing regions, is a risk factor responsible for high rates of preterm birth and infection, the two main causes of infant death. In addition, maternal health issues such as anemia, hypertension, or infection can complicate pregnancy and raise infant mortality risk [15].

Percentage of population suffering from hunger, [16].



<2,5%
2,5–5,0%
5,0–14,9%
15,0–24,9%
25,0–34,9%
>35,0%
No data.

1.2.4. Infections and environmental factors

Infections, especially, respiratory infections, diarrhea, and neonatal sepsis are one of the major causes of infant mortality around the world [11]. Insufficient sanitation, unclean drinking water, and poor vaccination coverage help to spread preventable illnesses [17]. Rural and impoverished communities without basic sanitation and clean water experience high levels of malnutrition and infection, adding to the risk to infant health [18].

In addition, environmental assaults such as pollution and climate change are being more widely appreciated as child health hazards. Increasing temperatures and poor air quality – especially in urban areas – have been associated with increased rates of respiratory infections and associated infant mortality [19].

1.2.5. Infections and environmental factors

Infectious diseases and adverse environmental conditions, including poor sanitation and unsafe water, substantially increase infant mortality risk in low-resource settings [20], [21].

1.3. The Target of this Study

The main objective of this research is to examine and predict IMRs in Iraq and Iran from 2025 to 2032 by exploiting the exponential grey model EXGM (1,1). This article attempts to attain three objectives: first, to analyze and compare the historical pattern and trend of IMR in the two countries through 2004 to 2024; second, to develop the forecasting model for IMR in future time based on the existing data; and third, to compare the population projections of IMR in both Iraq and Iran of projection period under study through 2025–2032, with maximum possible accuracy.

2. MATERIALS AND METHODS

Statistical forecasting is an essential element of data analysis, particularly in the case of public health indicators like infant mortality. In this work, the exponential grey model EXGM (1,1), which is a commonly-used grey-forecasting model, is used for prediction of future values on the patterns of past data. The EXGM (1,1) model is specially designed for small, incomplete or uncertain data sets, so it is applied to IMRs. Historical IMR data for Iraq and Iran from 2004 to 2024 were accessed from Macrotrends.net. The data set is available at this link: https://www.macrotrends.net/global-metrics/countries/IRQ/Iraq/birth-rate. The IMRs for the interval 2025–2032 were forecasted by adopting the exponential grey model, the EXGM (1,1). This model was fitted to each country's data independently to have computationally feasible forecasts.

2.1. Forecasting Model

Exponential grey model EXGM (1,1). The new grey appearance estimation model EXGM (1,1) is proposed. It applies to exponentially growing raw data sequences, and to the case of grey action as a function of time. In contrast to the classical GM (1,1) where the grey action is regarded as constant.

2.2. Evaluate Precision of Forecasting Models

2.2.1. Mean absolute percentage error (MAPE)

A number of statistical measures/test are utilized to check the validity and efficiency of the proposed model, on such measure is MAPE. Assessing the reliability and performance of the forecasting approach used in this study. The forecasting accuracy level can be classified into four grades based on the MAPE of each model as indicated in Table 1:

TABLE 1: Categorizing the grade of predicting accuracy depends on MAPE

Grade level	Highly accurate	Good	Reasonable	Inaccurate
MAPE	<10%	10% - 20%	20% - 50%	>50%

MAPE: Mean absolute percentage error

TABLE 2: Categorizing the grade of predicting accuracy depends on Precision rate (p)

Precision rank	Highly accurate	Good	Reasonable	Inaccurate
Precision rate (p)	p ≥ 99.0%	p ≥ 95.0%	p ≥ 90.0%	p ≥ 90.0%

MAPE: Mean absolute percentage error

TABLE 3: The infant mortality rates in Iraq and Iran

Years	Infant mortality rates		Years	Infant mortality rates	
	Iraq	Iran		Iraq	Iran
2004	33.94	23.68	2014	27.62	14.69
2005	33.39	22.54	2015	26.74	14.23
2006	32.83	21.41	2016	25.86	13.76
2007	32.28	20.27	2017	24.98	13.29
2008	31.73	19.13	2018	24.11	12.82
2009	31.08	18.34	2019	23.51	12.36
2010	30.43	17.55	2020	22.92	11.89
2011	29.79	16.75	2021	22.32	11.43
2012	29.14	15.96	2023	21.73	10.96
2013	28.49	15.16	2024	21.14	10.5

Per 1000 live births

TABLE 4: Compare the infant mortality rates between Iraq and Iran between 2004 and 2024

Countries	*N*	Mean	SD	*t*-test	*P*-value
Iraq	21	27.3662	4.36255	9.057	0.000
Iran	21	15.5652	4.07710		

SD: Standard deviation

Grade level highly accurate good reasonable inaccurate MAPE <10% 10–20% 20–50% >50%. A lower MAPE indicates higher precision in the forecasting model. Typically, a MAPE below 10% signifies an accurate model, while a MAPE between 10% and 20% denotes a good model with acceptable accuracy.

2.2.2. Precision rate (p)

Precision rate, which measures the level of the closeness of the statement of forecast quantity and the actual value.

Precision rank highly accurate good reasonable inaccurate. Precision rate (p) $P \geq 99.0\%$ $P \geq 95.0\%$ $P \geq 90.0\%$ $P \leq 90.0\%$. Table 2 shows a higher precision rate indicates greater precision in the forecasting model. Typically, a precision rate greater than 99% signifies an accurate model, while a precision rate between 98.0% and 95.0% suggests a good model with acceptable accuracy.

2.3. Application

The present paper also intends to investigate the time series structure of the IMRs in Iraq and Iran publicly announced during the period 2004: 1–2024: 12. The study also examines the use of the EXGM (1,1) model to predict the IMRs in Iraq and Iran from 2025 to 2032, using an historical series from 2004 to 2024 as a reference for model calibration and forecasting.

Table 3 illustrates the trends in IMRs in Iraq and Iran from 2004 to 2024. The data show a consistent decline in both countries, reflecting gradual improvements in healthcare systems, maternal and child health services, and living standards. In Iraq, the IMR decreased from 33.94 to 1,000 live births in 2004–21.14 in 2024, marking a reduction of approximately 37.7%. In contrast, Iran's rate fell more sharply from 23.68 to 10.5 during the same period, a decline of about 55.7%. Despite steady progress in Iraq, its rates remain higher than Iran's throughout the entire period, highlighting a persistent gap likely linked to differences in healthcare access, socio-economic conditions, and public health stability. These findings suggest that although both countries have improved, additional targeted efforts are needed in Iraq to accelerate progress and reduce infant mortality further, particularly by strengthening healthcare infrastructure, expanding maternal services, and addressing socio-economic challenges.

Table 4 presents a comparison of Infant Mortality Rates (IMRs) between Iraq and Iran over the period 2004–2024, with 21 observations recorded for each country. Findings indicate that Iraq had a higher mean IMR of 27.37 deaths/1,000 live births (standard deviation [SD] = 4.36) than Iran 15.57 (SD = 4.08). An independent samples *t*-test calculated whether this mean difference was significant. The result of the test gave a $t = 9.057$ and $P = 0.000$ indicating that there was a very high difference between the two countries ($P < 0.05$). These findings demonstrate that, during the past two decades, the level of infant mortality in Iraq has been considerably higher than in Iran. The wide variation in IMRs mirrors inequities in health systems, MCH care, economic development, and public health policies. This

study emphasizes the immediate necessity for sustained and thorough health programs in Iraq for preventing infant deaths and catching up with neighboring countries, such as Iran. Key investments in primary care, maternal education, immunization, and neonatal care are needed to reduce child deaths in Iraq.

Table 5 presents comparison of actual Table 5 presents that the comparison study of actual, predicted, and PE% values of IMR of Iraq, Iran from 2004 to 2024 based on exponential grey model EXGM (1,1). Our put the empirical results show that the empirical model generates accurate forecasts for the two countries. In Iran, the MAPE is 0.90706867% indicating a precision rate of 99.09293133%. In Iraq, lower MAPE of 0.795267266% is achieved with a precision of 99.20473273%. Using universal accuracy classification tables, a MAPE of <10% and a precision rate of 99% or more would be categorized as "highly accurate." The forecast of EXGM (1,1) in Iraq is optimum with regard to the less MAPE and little higher precision ratio than Iraq, and it is classified into the "Highly Accurate" class and occupy the top by following both the Grading criterion of Tables 1 and 2. Thus, the EXGM (1,1) model is of high credibility in the prediction of infant mortality in both countries. These findings validate the eventual performance and reliability of the model in generating robust predictions. The forecasted values for the years 2025–2032 are presented in Table 6.

Table 6 indicates that the comparison between the IMF in Iraq and Iran, (2025–2032) through EXGM (1,1) IMF in Iraq and IMF in Iran forecasting IMR for economy of Iraq and Iran up to 2032 in the Table 6 gives the forecasting of IMR in Iraq and in Iran for the period T = 2025–2032, we use of EXGM (1,1). Results reveal a consistent decrease in under-5 mortality rates for the two countries in the 8 years of prediction.

The rate of infant mortality in Iraq should drop from 20.13 deaths/1,000 live births in 2025–16.61 by 2032. This slow decline indicates ongoing progress in maternal and child health, as well as in vaccination coverage and social conditions. However, that higher IMR as compared to Iran could signal continued differences in healthcare infrastructure and socio-economic status.

We will, however, see a steeper decrease in the IMR of Iran, from 9.63 in 2025 to 7.28 in 2032. The relatively lower and progressively diminishing IMR is partly a result of the nation's historical commitment toward public health, successful family planning programs and the provision of urban health services.

TABLE 5: Predicted value and rate error value of the EXGM (1,1) model

Year	Real value		Forecast value		Percentage error (PE%)	
	Iraq	Iran	Iraq	Iran	Iraq	Iran
2004	33.94	23.68	33.94	23.68	0	0
2005	33.39	22.54	33.15895207	22.70337201	0.691967445	0.724809282
2006	32.83	21.41	33.26153384	21.06292185	1.314449707	1.621102997
2007	32.28	20.27	32.73009217	19.96115913	1.394337566	1.523635273
2008	31.73	19.13	31.9807869	19.07710135	0.790377872	0.276521937
2009	31.08	18.34	31.16629916	18.29189987	0.277667833	0.262268954
2010	30.43	17.55	30.34239284	17.561097	0.287897323	0.063230795
2011	29.79	16.75	29.52918888	16.86763082	0.875498902	0.702273551
2012	29.14	15.96	28.7337065	16.20454545	1.394281041	1.532239645
2013	28.49	15.16	27.95815547	15.56862945	1.866776156	2.695444918
2014	27.62	14.69	27.20298628	14.95807447	1.50982521	1.824877257
2015	26.74	14.23	26.46801201	14.37161289	1.017157773	0.995171395
2016	25.86	13.76	25.75282082	13.80819964	0.414459338	0.350288048
2017	24.98	13.29	25.05692734	13.26689419	0.307955726	0.173858641
2018	24.11	12.82	24.37982823	12.74681628	1.119154819	0.570855813
2019	23.51	12.36	23.72102227	12.24712878	0.897585136	0.913197539
2020	22.92	11.89	23.08001758	11.76703052	0.69815697	1.034226058
2021	22.32	11.43	22.45633403	11.3057529	0.610815535	1.087026203
2022	21.73	10.96	21.84950388	10.8625579	0.549948841	0.889070258
2023	21.14	10.5	21.25907184	10.43673656	0.56325375	0.602508944
2024	20.66	10.15	20.68459483	10.02760779	0.119045645	1.205834557
Results			MAPE (%)		0.795267266	0.90706867
			Precision rate (p)		99.20473273	99.09293133

MAPE: Mean absolute percentage error

TABLE 6: The forecasted value for infant mortality rate in Iraq and Iran

Years	2025	2026	2027	2028	2029	2030	2031	2032
Iraq	20.12564	19.58179	19.05264	18.53779	18.03685	17.54944	17.07521	16.61379
Iran	9.634517	9.256836	8.89396	8.54531	8.210327	7.888475	7.57924	7.282128

Ongoing decline in infant mortality in these countries may be attributed to improved prenatal and postnatal care, better nutrition, and improving accessibility to healthcare Iraq probably lags slightly behind Iran's achievements in these fronts.

Such projections are important for public health policy-makers and policy and program planners and help point out that continued efforts to reduce infant mortality are necessary through focused health interventions, particularly in rural and underserved areas. Knowledge of these trends can help planning of adequate interventions for the attainment of the SDGs concerning child survival and health equity.

In Table 7, we compare that the predicted infant in Table 7 displays that the compare predicted IMR in may Iraq and Iran for the period 2025–2032 using the predictions of the exponential grey model (EXGM [1,1]). The table contains estimates for 8 years for each country. Based on the fitted model, the average infant mortality rate (IMR) in Iraq is estimated at 18.32 deaths per 1,000 live births (SD = 1.23), whereas Iran is predicted to have a significantly lower mean IMR of 8.41 deaths per 1,000 live births (SD = 0.82). To test if such a difference is statistically significant, an independent samples *t*-test was performed. The *t*-test provided a $t = 18.951$ and $P = 0.000$, which is extremely less than the standard level of 0.05. This finding implies that there is a significant difference in the predicted IMRs of the two countries up to the observed period. These results reveal a huge gap between expected child health attainment of Iraq and that of Iran. Iraq is expected to suffer many more infant deaths than Iran, unless effective policy actions and healthcare interventions are implemented. The ironic prediction highlights an urgent requirement for focused investments in public health, especially in maternal and child health, to minimize infant fatalities and help Iraq approach international child survival thresholds.

TABLE 7: Compare the forecasting of the Infant mortality rates between Iraq and Iran between 2025 and 2032

Countries	*N*	Mean	SD	*t*-test	*P*-value
Iraq	8	18.3216	1.22889	18.951	0.000
Iran	8	8.4113	0.82316		

SD: Standard deviation

3. DISCUSSION

The results of this research contribute a better understanding of the dynamics of infant mortality in Iraq and Iran and show the effectiveness of EXGM (1,1) model through time series prediction. The higher rate of infant mortality in Iraq during 2004–2024 with a statistically significant mean difference underlines the long-lasting difficulties still present in the system of maternal and child health. These can be attributed to factors such as a lack of health care infrastructure, economic uncertainty, war, and a lack of pre and postnatal care.

The model's projections for 2025–32 suggest this gap will not only persist but could widen unless Iraq enacts major change. The predicted average IMR for Iraq, exactly more than twice as high as Iran's (18.32 vs. 8.41), is a shocking shot in the arm. In stark contrast, the below (and steadily further reducing) predicted rates for Iran shows that its healthcare policies and social programs are working well.

As from the researcher's point of view for the future evolution of the birth rate, future demographic changes are predicted in both countries. The falling birthrate, especially in Iran, could be explained by growing internet use, higher education for women, economic shifts, urbanization, and easy access to contraception. Iraq also shows slight but continuing decrease and hints at the beginning of the long-term demographic trend toward change.

Such developments have more widespread relevance for long-term population and labor force structures, and health care systems. These dynamics must be taken into account by policy makers going forward. The EXGM (1,1) model is demonstrated to be an effective tool for predicting such complex trends and can provide practical implications which can contribute for designing effective public health interventions.

4. CONCLUSION

In this study, the exponential grey model EXGM (1,1) is used to evaluate and predict the infant mortality rate in Iraq and

Iran between 2004 and 2032. The model performs very well, with small MAPE values of 0.795% and 0.907% for Iraq and Iran, respectively, implying accuracy rates higher than 99% for both countries. These results categorize the performance of the model in the "highly accurate" range and demonstrate its adequacy for robust time series predictions even with scarce data.

Statistical analysis A historical comparison covering a 20-year period (2004–2024) to compare the frequency of the events statewide between Iran and Iraq indicated a large difference in the infantile mortality rate, where the infantile mortality rate was high in Iraq rather than Iran (the average rate for infantile mortalities in Iran is 15.57 vs. 27.37 in Iraq). This difference was highly significant ($P < 0.001$), indicating pronounced differences in access to healthcare, socio-economic circumstances, and Child Welfare Services in the two countries.

In addition, predictions of 2025–2032 periods also supported the trend. A statistically significant higher IMRs in Iraq estimated with a mean of 18.32 (Iraq) compared with Iran 8.41 ($P < 0.001$) is also predicted. The birth rate is also projected to fall in both countries during the same period from 20.13 to 16.61 in Iraq and 9.63 to 7.28 in Iran.

These findings highlight the urgent requirement for Iraq to utilize such focused public health strategies and interventions. Without such work, the difference in infant mortality between Iraq and Iran is set to endure. This study illustrates the potential of predictive modeling to inform health policy and enhance child health conditions.

5. RECOMMENDATION

These data suggest that Iraq should work to develop a reform policy for health sector focusing on maternal and child health. This would include increased access to prenatal care, better neonatal care facilities, trained healthcare providers, and using updated guideline in heath setting community health education. Policymakers also need to invest in surveillance systems able to generate timely warning. Furthermore, the inclusion of successful components of Iran's health program – especially family planning and preventive health care – could bring Iraq's infant death rate down significantly. Finally, prediction models such as the exponential grey model must be institutionalized to enable regular monitoring and prediction of key indicators in health, for evidence-based decision-making.

REFERENCES

[1] World Health Organization (WHO). "*Infant Mortality*", 2022. Available from: https://www.who.int/data/gho/indicator-metadata-registry/imr-details/61 [Last accessed on 2025 Jun 11].

[2] United Nations Development Programme (UNDP). "*Sustainable Development Goals*", 2023. https://www.undp.org/sustainable-development-goals

[3] M. N. Moghadam, A. Aryankhesal and H. Haghparast-Bidgoli. "Reducing infant mortality in Iran: Successes, challenges, and the way forward". *International Journal of Health Policy and Management*, vol. 10, no. 5, pp. 264-272, 2021.

[4] T. K. Al Hilfi, R. Lafta and G. "Burnham. Health services in Iraq". *The Lancet*, vol. 381, no. 9870, pp. 939-948, 2013.

[5] G. E. Box, G. M. Jenkins, G. C. Reinsel and G. M. Ljung. "*Time Series Analysis: Forecasting and Control*". 5th ed. Wiley, United States, 2016.

[6] J. Deng. "Introduction to grey system theory". *The Journal of Grey System*, vol. 1, no. 1, pp. 1-24, 1989.

[7] S. Liu and Y. Lin. "*Grey Systems: Theory and Applications*". Springer, Berlin, 2010.

[8] World Bank. "The World Bank Provides Annual Estimates for the Infant Mortality Rate in Iraq". In: *Their Most Recent Data*. World Bank, United States, 2023.

[9] United Nations. "*The Millennium Development Goals Report 2020*". UN Website, United Nations, 2020.

[10] World Health Organization (WHO). "*Infant Mortality*". WHO Fact Sheets, Geneva, 2021.

[11] A. Khan, M. Ali and N. Siddiqui. "Improving maternal and child health in low-income settings: Key interventions". *The Lancet Global Health*, vol. 8, no. 10, pp. e1265-e1272, 2020.

[12] C. G. Victora, L. Adair, C. Fall, C. Hallal, P. R. Martorell and L. Richter. "Maternal and child undernutrition: Consequences for adult health and human capital". *The Lancet*, vol. 371, no. 9609, pp. 340-357, 2016.

[13] O. Lincetto, S. Mothebesoane-Anoh and A. Langer. "Maternal education and child survival in developing countries: A review of the evidence". *The Lancet*, vol. 368, no. 9535, pp. 1817-1825, 2016.

[14] R. E. Black, C. G. Victora, S. P. Walker, Z. A. Bhutta, P. Christian, M. De Onis, M. Ezzati, S. Grantham-McGregor, J. Katz, R. Martorell, R. Uauy and Maternal and Child Nutrition Study Group. "Maternal and child undernutrition and overweight in low-income and middle-income countries". *The Lancet*, vol. 382, no. 9890, pp. 427-451, 2013.

[15] H. Tappis, E. Munkhtsetseg and P. Waiswa. "Maternal health and infant mortality in East Asia: A review of trends and interventions". *BMC Health Services Research*, vol. 20, no. 1, p. 123, 2020.

[16] World Food Programme. "*UN World Food Programme (WFP)*". World Food Programme, United States, 2020.

[17] E. Ransom and P. Yoder. "The impact of water, sanitation, and hygiene on child health: A critical review of the literature". *Global Health Actio*n, vol. 2, no. 1, pp. 1-7, 2009.

[18] C. Troeger, I. A. Khalil, P. C. Rao, S. Cao, B. F. Blacker, T. Ahmed, G. Armah, J. E. Bines, T. G. Brewer, D. V. Colombara, G. Kang, B. D. Kirkpatrick, C. D. Kirkwood, J. M. Mwenda, U. D. Parashar, W. A. Petri Jr., M. S. Riddle, A. D. Steele, R. L. Thompson, J. L. Walson, J. W. Sanders, A. H. Mokdad, C. J. L. Murray, S. I. Hay and R. C. Reiner. "Rotavirus vaccination and the global burden of rotavirus diarrhea among children younger than 5 years". *JAMA Pediatr*, vol.10, pp. 958-965, 2018.

[19] S. K. Sahu, P. Sharma and A. Kumar. "Air pollution and its impact on public health: Emerging trends in India". *Environmental Monitoring and Assessment*, vol. 190, no. 8, pp. 1-12, 2018.

[20] R. E. Black, C. Levin, N. Walker, D. Chou, L. Liu, M. Temmerman and Disease Control Priorities Project. "Reproductive, maternal, newborn, and child health: Key messages from disease control priorities 3rd edition". *The Lancet*, vol. 388, no. 10061, pp. 2811-2824, 2017.

[21] World Health Organization. "*Infant Mortality*", 2023. Available from: https://www.who.int/news/room/fact-sheets/detail/infant-mortality

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Awareness of Infection Control and Barriers among the Healthcare Workers in Sulaymaniyah City, Iraq

Niyan Hakim Ismael[1]*, Delan Jamal Qader[1], Gona Othman Faris[2], Awayi Ghazy Abdulkareem[3], Chnur Talib Mohamad[4]

[1]Department of Adult Nursing, College of Nursing, University of Sulaimani, Kurdistan Region, Republic of Iraq, [2]Department of Maternal Neonate Nursing, College of Nursing, University of Sulaimani, Kurdistan Region, Republic of Iraq, [3]Department of Pediatric Nursing, College of Nursing, University of Sulaimani, Kurdistan Region, Republic of Iraq, [4]Sulaimani Teaching Hospital, Directorate of Health, Ministry of Health, Iraq Kurdistan Region, Republic of Iraq

ABSTRACT

Background: Healthcare workers (HCWs) are at risk of exposure to blood-borne pathogens such as human immunodeficiency virus and hepatitis B and C viruses through sharps injuries and contact with body fluids. Standard infection control practices are critical to preventing such exposures. **Objectives:** This study aimed to assess the awareness of infection control measures and perceived barriers among HCWs in Sulaymaniyah governorate, Iraq. **Patients and Methods:** A cross-sectional survey was conducted between June 2021 and May 2022 among 557 HCWs working in different hospitals in Sulaymaniyah. Data were collected using a self-administered questionnaire with 44 items covering awareness, standard precautions, and barriers to infection control. Descriptive and inferential statistics were applied to analyze the data. **Results:** Overall, 79% of participants demonstrated adequate awareness of infection control practices, and 82% reported compliance with standard precautions. A significant association was observed between participation in infection control training and higher awareness levels ($P < 0.05$). Despite this, 52% of respondents acknowledged barriers to consistently applying standard precautions within their hospital units. **Conclusions:** Although awareness and compliance with infection control practices among HCWs in Sulaymaniyah were generally high, notable barriers remain. Regular training and institutional support are recommended to enhance adherence to standard precautions and strengthen infection prevention in healthcare facilities.

Keywords: Awareness, standard precaution, health care workers, barriers, cross-sectional survey

1. INTRODUCTION

Healthcare-associated infection (HCAI) offerings constitute a significant concern for patient safety, and their anticipation must be the priority for institutions and individuals committed to making health care safer for concerned individuals. The menace of obtaining HCAI by patients or caregivers is worldwide and includes all healthcare facilities; the global load is unknown due to the difficulties in gaining reliable, indicative data [1].

The need of the time is a standard shift from a tapered, separate patient-based attitude to a more comprehensive approach, pointing to infection control during the health care process. In this way, a substantial improvement in this

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp313-318

E-ISSN: 2521-4217
P-ISSN: 2521-4209

Copyright © 2025 Ismael, *et al*. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Niyan Hakim Ismael, Department of Adult Nursing, College of Nursing, University of Sulaimani, Kurdistan Region, Republic of Iraq, Tel: 009647701930039, E-mail: niyan.ismael@univsul.edu.iq

Received: 09 July 2024 Accepted: 26 July 2025 Published: 24-11-2025

energetic area of healthcare is possible. Alteration is essential at the organizational/administrative stages to incorporate all influential factors that impact economic infection control [2].

HCW obedience with ordinary protections was revealed to decrease the exposure risk to blood and body fluids [3]. Hand hygiene is among the most imperative factors of infection control that can diminish the risk of cross-transmission in healthcare agencies. Consequently, the hindrance of HCAIs is the accountability of all people and facilities offering health care [4].

The Centers for Disease Control and Prevention has mentioned, "It is well-documented that one of the most important measures for preventing the spread of pathogens is effective hand washing." The foremost medical aim of hand washing is to clean the hands of pathogens and chemicals, which can cause particular harm/disease [5]. Maximum HCAIs are mainly conveyed by the hands of HCWs by direct contact, especially when pathogenic microorganisms are transferred between individuals and the environmental reservoirs [6].

Barriers to the implementation of infection prevention and control (IPC) measures are multifactorial. Studies elsewhere in Iraq and in comparable settings have identified resource limitations such as shortages of personal protective equipment (PPE), inadequate facilities for handwashing or sterilization as recurrent obstacles [7]. Furthermore, lack of continuous training, insufficient institutional support, high workload, and staff shortages also emerge frequently in the literature [8].

The HCWs comprise a crucial outline of the hospitals, which is vigorous to its functioning, and subsequently, their awareness of infection control and proper performance is serious to averting HCAIs. Hence, numerous reasons affect poor compliance with standard guidelines. Therefore, this study was planned to throw more highlights to take appropriate actions. The present study aimed to assess the awareness and practices regarding standard precautions for infection control among healthcare workers (HCWs) [9].

2. MATERIALS AND METHODS

2.1. Design of the Study

This cross-sectional, interview-based study was conducted on 557 HCWs, from June 1st, 2021, to May 31st, 2022, in different hospitals in Sulaimaniyah governorate, Iraq.

2.1.1. Questionnaire

The questionnaire comprises three parts: internal review, target population input, and expert external review. Regarding the information for the target population, HCWs were randomly selected to answer the questions freely to frame and re-arrange the questionnaire. The number of questions to measure sociodemographical characteristics was 8, questions about the infection control part were 30, and items regarding barriers to practising standard infection control precautions in HCWs were 14. Specific codes (Table 1) were used for respondents' answers.

2.2. Ethical Considerations

Appropriate guidelines and regulations belonging to the Declaration of Helsinki were followed to conduct this research study on HCWs. The Scientific and Ethics Committees of the College of Nursing, University of Sulaimani, Sulaimaniyah, Iraq, revised and approved the study protocol (No. 58/30/04/2023/UoS). Before collecting data, formal authorization was obtained from health and government authorities. Written informed consent was obtained. The participation was voluntary, their data were kept confidential, and nurses had the right to withdraw at any time during the data collection period without giving a reason.

2.3. Statistical Analysis

Statistical Package for Social Science (SPSS, version 27) was used for data analysis. The mean and standard deviation were presented for variables using descriptive analysis. The independent t-test and analysis of variance test were used to analyze variables, and a $P \leq 0.05$ was considered statistically significant.

3. RESULTS

Among 556 studied HCWs, 124 (22%) were male and 432 (78%) were female, with ages varying from 20 to 59 years. Most of them were from the emergency unit (27%), aged between 34 and 40 years (26%), worked as a nurse (58%), had around 8 years of experience, and had not participated in a training course for infection control (62%) (Table 2).

TABLE 1: Specific codes used for respondents' answers

Option	Code	Option	Code
True	3	Agree	2
I don't know	2	Disagree	1
False	1		

TABLE 2: Descriptive statistics of sociodemographic variables among respondents

Variable	Frequency	%
Age (years)		
≤26	68	12
27–33	104	19
34–40	144	26
41–47	100	18
48–54	112	20
>55	28	5
Gender		
Male	124	22
Female	432	78
Occupation		
Doctor	120	22
Nurse	324	58
Laboratory assistant	92	17
Others	20	4
Unit		
Intensive care unit	24	4
Cardiac care unit	12	2
Medical ward	84	15
Surgical ward	88	16
Emergency	148	27
Premature unit	16	3
Labor room	72	13
Neurology	72	13
Laboratory	40	7
Experience (years)		
<1	30	5
4	146	26
8	334	60
>8	46	8
Infection control training course		
Yes	212	38
No	344	62

The mean arithmetic ratio, in general, was 2.78 ± 0.58 (importance of glove), and the general trend of this variable as a whole was toward (True). These results indicated that the research community members knew awareness regarding infection control and standard precautions variables are of great importance, and 81% and 87%, respectively, at the level of (True), from their point of view, and also that 15% of the research sample don't have information ultimately on awareness regarding infection control and standard precautions or not sure about them, 16% and 3%, respectively at the level of (False) (Table 3).

Regarding the barriers to the practice of standard precaution variables, the arithmetic mean ± SD ratio at the general level was 1.5 ± 0.5, and the general trend of this variable was towards (Agree). These results indicated that the research community members agreed that the barriers to the practice of standard precautions exist within hospital units. About 52% of the research samples were at the level of (Agree) according to their point of view. As for each question separately within this variable, the explanation was as follows: For the item (it is unimportant to follow SPs guidelines (item 11), the highest value of the arithmetic mean appears, and it reached 1.75, which means that they are entirely agreed. The standard deviation earned 0.43, and this paragraph indicated that 74.8% of the research samples were at the level of (Agree) (Table 4).

Table 5 shows the correlations between factors behind infection control and sociodemographic information. According to occupation, $P = 0.039$ for awareness and $P = 0.314$ for standard precautions. This means that H_0 can be rejected according to the status that the mentioned factors are not statistically significant differences overall about the occupation. At the same time, the standard precautions were not substantial overall (cannot be rejected null hypothesis, which states that this factor has no significant difference overall about occupation).

The t-test results revealed a significant difference between the male and female genders regarding standard precautions at the level of $\alpha = 0.05$. It was noted that a *P*-value of 0.004 is smaller than the level $\alpha = 0.05$. This means that H_0 can be rejected as stated that the standard precaution is the same for males and females (Gender) with a significant difference between them ($P = 0.05$). While a significant difference between male and female gender of factor awareness at the level of $\alpha = 0.10$ was seen ($P = 0.067$). This means H_0 can be rejected as stated that the awareness regarding infection control is the same for males and females (gender) with a significant difference between them ($P = 0.10$). On the other hand, there was a significant difference according to taking training courses on factors (infection control and standard precautions) at the level of $\alpha = 0.05$ (Table 6).

4. DISCUSSION

This study assessed HCWs' awareness of day-to-day Integrated Physician Practice Sections. Based on the sociodemographic data, most participants (78%) were female, which aligns with trends in many countries where the health workforce is predominantly female in nursing and allied professions. Recent studies have similarly reported higher proportions of female HCWs, for example, in studies of PPE usage and infection control perceptions [10].

In this study, more than half of HCWs were aged 34–44 years; fewer than half had 8 years or less of work experience; and

TABLE 3: Standard precaution knowledge and standard precaution of the studied participants

Awareness regarding infection control	True	False	I don't know	Mean±SD	Coefficient of variance	Relative importance
	%					
Hand hygiene	4928	844	344	2.66±0.65	25.21	88.92
	80	14	6			
Needle stick injury prevention	1684	408	132	2.57±0.75	29.86	85.79
	76	18	6			
Standard precautions	3148	632	128	2.65±0.67	26.28	88.49
	81	16	3			
Proper usage of surgical masks	1792	356	76	2.64±0.70	27.23	88.19
	81	16	3			
Importance of glove	972	100	40	2.78±0.58	21.11	92.8
	87	9	4			

TABLE 4: Barriers to the practice of standard precaution (SP) variables among studied participants

Item	Agree	Disagree	Mean±SD	Coefficient of variance	Relative importance
	Number, %				
1. There is no enough glove	254	302	1.47±0.51	35	72.84
	45.7	54.3			
2. Others do not follow SPs measures	209	347	1.4±0.51	37	68.79
	37.6	62.4			
3. There is no enough gowns	226	330	1.42±0.5	35.8	70.32
	40.6	59.4			
4. Guidelines of SPs are vague	274	282	1.5±0.52	34.3	74.64
	49.3	50.7			
5. There is no enough hand washing facilities	268	288	1.48±0.5	33.8	74.1
	48.2	51.8			
6. Requirements of SPs are costly	268	288	1.48±0.5	33.8	74.1
	48.2	51.8			
7. Gloves cause skin irritation	400	156	1.72±0.45	26.2	85.97
	71.9	28.1			
8. Hepatitis B vaccine is unavailable	276	280	1.5±0.5	33.3	74.82
	49.6	50.4			
9. Following SPs takes long time	280	276	1.5±0.5	33.3	75.18
	50.4	49.6			
10. PPE are uncomfortable with use	280	276	1.5±0.5	33.3	75.18
	50.4	49.6			
11. It is unimportant to follow SPs guidelines	416	140	1.75±0.43	24.8	87.41
	74.8	25.2			
12. Following SPs makes work harder	312	244	1.56±0.49	31.9	78.06
	56.1	43.9			
13. Trash bins available in hospital	288	268	1.52±0.5	32.9	75.9
	51.8	48.2			
14. Do you have a designated place to dispose of waste products outside the hospital?	264	292	1.47±0.5	34	73.74
	47.5	52.5			
Total	4015	3769	1.5±0.5	32.6	75.79
	52	48			

most had not received sufficient training in infection control measures. The finding that training strongly influences knowledge is consistent with recent literature. For instance, in Qatar in 2022, HCWs who had received training in the previous year showed significantly higher knowledge of IPC measures [11]. Furthermore, older studies during the COVID-19 pandemic highlight that a lack of recent or repeated training is linked to gaps in both knowledge and compliance [12], [13].

Some HCWs may lack sufficient knowledge because their organizations have invested little in staff training, have weak systems for reinforcing safe clinical practices, or are due to complacency. Despite educational efforts, practice often remains suboptimal due to poor retention, inconsistent adherence to guidelines, and information overload. Recent reviews echo this, showing that knowledge alone is not enough–ongoing reinforcement, auditable compliance, and

TABLE 5: The correlations between factors behind infection control in relation to sociodemographic information using ANOVA test

Variable	Awareness regarding infection control	*P*-value	Standard precautions F (sig.)	*P*-value
Age group	8.202	0.000	2.805	0.016
Occupation	2.802	0.039	1.186	0.314
Years of experience	25.574	0.000	6.540	0.000
Unit	16.853	0.000	12.496	0.000

ANOVA: Analysis of variance

TABLE 6: Comparison between factors in the infection control using t-test

Variable	Gender		Training course	
	T-test	Sig. (2-tailed)	T-test	Sig. (2-tailed)
Awareness regarding infection control	1.841	0.067	0.744	0.457
Standard precautions	2.873	0.004	3.382	0.001

institutional safety culture play large roles [14], [15]. Moreover, in the current study, we observed significant associations between HCWs' knowledge of infection control and variables such as age, gender, years of experience, and whether or not they had attended training courses. This is supported by more recent evidence: for example, the Qatar study found male HCWs, those with more training, and those who had exposure to COVID-19 (e.g., via colleagues or relatives) had higher perceived effectiveness and knowledge of IPC measures [16], [17].

Other studies have similarly found that younger age, less experience, and recent IPC or PPE training correlate with better knowledge and compliance [18]. Regarding protective barriers and PPE usage: HCWs reduce their risk of exposure by using protective barriers; yet many report limited availability of protective materials. In this study, only face masks and gloves were consistently available, and only about 52% of respondents agreed with the importance of using these. Comparable findings emerge more recently: a study from Qatar reported that while most HCWs recognize the effectiveness of various PPE and hand hygiene, actual consistent use is lower and is heavily influenced by training, perceived effectiveness, and resource availability [19]. In addition, non-compliance has been attributed to high workload, discomfort, lack of training, forgetfulness, norms, and co-worker behavior. These barriers are again confirmed in more current studies [20].

5. CONCLUSION

It is concluded that the knowledge and practice of standard precautions are acceptable with general standards to guarantee infection safety in health workers and patients. Regular training or sponsorship for workshops and seminars is needed. However, barriers to the practice of standard precautions prevent the improvement in the high-quality preparation of infection control. This is the most effective and long-lasting means to improve health workers' knowledge and foster compliance with standard precaution measures.

6. ACKNOWLEDGMENTS

We highly thank the participating nurses and all health staff concerned about this research's success.

7. CONFLICTS OF INTEREST

The author affirms that they have no conflicts of interest.

8. FUNDING

This study is self-funded, and grants or funds were not received from any agency, society, organization, or university.

REFERENCES

[1] L. M. Asegu, A. Kitschen, M. M. Neuwirth, and D. Sauerland. "The economic burden of nosocomial infections for hospitals: Evidence from Germany." *BMC Infectious Diseases*, vol. 24, p. 1294, 2024.

[2] M. M. Patwary, M. Bardhan, A. S. Disha, M. P. Kabir, M. R. Hossain, M. A. Alam, *et al.* "Knowledge, attitudes and practices of healthcare workers regarding COVID-19 in Bangladesh: A nationwide study". *Frontiers in Public Health*, vol. 10, p. 988063, 2022.

[3] B. Allegranzi, E. Tartari and D. Pittet. "Infection prevention and control priorities 2021–2030: A global call for action." *Lancet Infectious Diseases*, vol. 21, no. 6, pp. e144-e149, 2021.

[4] J. S. Tabrizi, M. Letaief, H. M. Abdolahi, R. Rezapour, A. Alboksmaty and M. Farahbakhsh. "Safety of health workers during the COVID-19 pandemic and beyond: Piloting WHO framework in Iran". *Journal of Health, Population and Nutrition*, vol. 44, p. 29, 2025.

[5] S. Jing, Z. Dai, Y. Wu, X. Liu, L. Zhang, X. Liu, T. Ren, J. Fu, X. Chen, W. Wang, X. Gu, L. Ma, S. Zhang, Y. Yu, L. Li, Z. Han, X. Su and Y. Qiao. "Prevalence and influencing factors of occupational burnout among healthcare workers in the Chinese mainland during the late 2022 Omicron COVID-19 outbreak: A multicenter cross-sectional study". *BMC Public Health*, vol. 25, p. 171, 2025.

[6] M. S. Al-Beashi, R. A. Albalawi, I. H. M. Albalawi, N. A. Albalawi, W. A. Albalawi, F. M. A. Al-Enezi, M. A. M. Albalawi, R. M. A. Alanazi, A. N. Alosaimi, S. H. Alasmari, A. Alqethami, A. O. Alenazi, N. O. Alanazi, Y. M. S. Alsultan, E. A. A. Almulhim and A. S. Alanzan.

"Knowledge, attitudes, and practices of healthcare workers about health care-associated infections in Saudi Arabia". *Journal of International Crisis and Risk Communication Research*, vol. 7, Suppl. 10, pp. 467-483, 2024.

[7] S. M. M. Hossain, S. Al-Dahir, R. Al Hilfi, Y. Majeed, A. Rahi, V. Sabaratnam, T. Al-Mulla, O. Hossain, A. Aldahir, R. Norton and F. Lami. "Evaluation of dedicated COVID-19 hospitals in the pandemic response in Iraq: Pandemic preparation within a recovering healthcare infrastructure". *BMJ Global Health*, vol. 7, p. e008715, 2022.

[8] M. A. Hasan, W. A. A. Ali and R. A. Mahmoud. "Assessment of nurses' practice of standard precautions in hospitals of Babil Governorate, Iraq". *Journal of Health, Safety and Sustainable Development*, vol. 12, pp. 83-89, 2023.

[9] World Health Organization. "*Implementation Manual to Prevent and Control the Spread of Carbapenem-Resistant Organisms at the National and Health Care Facility Level"*. WHO, Geneva, 2019.

[10] H. A. Hamid, M. M. Mustafa, M. Al-Rasheedi, B. Balkhi, N. Suliman, W. Alshaafee and S. A. Mohammed. "Assessment of hospital staff knowledge, attitudes and practices (KAPS) on activities related to prevention and control of hospital-acquired infections". *International Journal of Prevention and Treatment*, vol. 8, pp. 1-7, 2019.

[11] F. A. Shah, A. O. Alanazi, Y. K. Alrawili and Y. M. Alanazi. "Knowledge and attitudes towards infection control protocols among the students of Al-Jouf University in Saudi Arabia". *The Egyptian Journal of Hospital Medicine*, vol. 71, pp. 3328-3331, 2018.

[12] G. Abuduxike, S. A. Vaizoglu, O. Asut and S. Cali. "An assessment of the knowledge, attitude, and practice toward standard precautions among health workers from a hospital in northern Cyprus". *Safety and Health at Work*, vol. 12, pp. 66-73, 2021.

[13] M. Abed Alah, S. Abdeen, N. Selim, D. Hamdani, E. Radwan, N. Sharaf, H. Al-Katheeri and I. Bougmiza. "Knowledge and perceived effectiveness of infection prevention and control measures among health care workers during the COVID-19 pandemic: A national survey". *Journal of Nursing Care Quality*, vol. 37, no. 2, pp. E23-E30, 2022.

[14] M. T. T. A. Alah, S. Abdeen, N. Selim, E. Tayar and I. Bougmiza. "Occupational prevention of COVID-19 among healthcare workers in primary health care settings in Qatar: Predictors of full compliance with PPE". *Journal of Patient Safety*, vol. 18, pp. 745-755, 2022.

[15] A. K. Basher, M. A. A. J. Biswas, A. Rahman, M. Rahman, F. Chowdhury and M. Z. Hassan. "Occupational risk of SARS-CoV-2 infection among healthcare workers in Bangladesh: A multicenter hospital-based study and lessons for future epidemics". *Tropical Medicine and Health*, vol. 52, p. 92, 2024.

[16] B. Bayleyegn, A. Mehari, D. Damtie and M. Negash. "Knowledge, attitude and practice on hospital-acquired infection prevention and associated factors among healthcare workers at university of Gondar comprehensive specialized hospital, Northwest Ethiopia". *Infection and Drug Resistance*, vol. 14, pp. 259-266, 2021.

[17] T. Engda. "The contribution of medical educational system of the College of Medicine, and Health Sciences of the University of Gondar in Ethiopia on the knowledge, attitudes, and practices of graduate students of health sciences in relation to the prevention and control of nosocomial infections during the academic year of 2018". *BMC Medical Education*, vol. 20. pp. 1-12, 2020.

[18] M. Khan, K. Muhmood, H. Z. Mahmood, I. H. Khaliq and S. Zaman. "The health and economic burden of dust pollution in the textile industry of Faisalabad, Pakistan". *Journal of the Egyptian Public Health Association*, vol. 99, p. 3, 2024.

[19] B. A. Al Ra'awji, E. S. Almogbel, L. A. Alharbi, A. K. Alotaibi, F. A. Al-Qazlan and J. Saquib. "Knowledge, attitudes, and practices of healthcare workers regarding hand hygiene guidelines in Al-Qassim, Saudi Arabia: A multicenter study". *International Journal of Health Sciences*, vol. 12, pp. 3-8, 2018.

[20] M. Shoaib. "Universal precautions: Knowledge and practices of ICU nurses". *Journal of Medical Sciences*, vol. 10, no. 1, pp. 45-50, 2025.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

A Multi-Account Statistical Evaluation of ChatGPT Proficiency in the Kurdish Sorani Language

Alla Ahmad Hassan[1], Hemin Sardar Abdulla[2], Tara Yousif Mawlood[3], Rebwar Khalid Muhammed[4], Aso M. Aladdin[5]*, Tarik A. Rashid[6]

[1]Department of Database, Computer Science Institute, Sulaimani Polytechnic University, Sulaimani, Iraq, [2]Department of Software Engineering, College of Engineering and Computational Science, Charmo University, Sulaymaniyah, Chamchamal, Iraq, [3]Department of IT, Computer Science Institute, Sulaimani Polytechnic University, Sulaymaniyah, Iraq, [4]Department of Network, Computer Science Institute, Sulaimani Polytechnic University, Sulaimani, Iraq, [5]Computer Engineering Department, Tishk International University, Kurdistan Region, Iraq, [6]Computer Science and Engineering Department, University of Kurdistan Hewler, Erbil, Iraq

ABSTRACT

This research analyzes the strengths and weaknesses of ChatGPT in responding to questions posed in the Kurdish language, specifically its Sorani dialect, by evaluating its responses to a structured dataset of 50 multiple-choice questions across multiple topics such as language, history, culture, and general knowledge. Using four independent user accounts, each subjected to ten repeated testing cycles, the research assesses accuracy, consistency, and variation influenced by account identity, session timing, and model behavior. This study evaluates the multilingual capabilities of ChatGPT by comparing its performance in Kurdish (Sorani) and Arabic languages. The research establishes a framework to examine how artificial intelligence chatbots, such as ChatGPT, function as applied tools for language understanding and educational use. The analysis demonstrates that ChatGPT achieved an overall average accuracy rate of approximately 70%, indicating satisfactory performance in multilingual contexts. However, significant variations were observed across different user accounts, suggesting that factors such as user profile and temporal dynamics can considerably influence output consistency. The comparative findings highlight the developmental challenges in Arabic and Kurdish language processing, emphasizing the need for further refinement of ChatGPT's linguistic performance and its effective integration into academic and technological applications. While ChatGPT exhibited proficiency in answering general knowledge questions, it demonstrated a limited understanding of specialized topics in Kurdish, particularly classical literature and historical content. The research presents the strengths and limitations of ChatGPT for under-resourced languages and provides feedback to developers, educators, and researchers. Observing patterns in accuracy, question difficulty, and error behavior, this research also contributes to ongoing efforts toward improving the linguistic and cultural adequacy of AI models for under-resourced languages.

Index Terms: ChatGPT, Chatbot, AI, Kurdish Language, Sorani Dialect, Statistical Analysis

Access this article online

DOI: 10.21928/uhdjst.v9n2y2025.pp319-334

E-ISSN: 2521-4217
P-ISSN: 2521-4209

Copyright © 2025 Hassan, *et al*. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

1. INTRODUCTION

Artificial intelligence (AI) is being developed across all sectors of life, and it is essential for all individuals to understand its significance, as it plays a crucial role in facilitating and streamlining various tasks [1]. As a result, the most advanced

Corresponding author's e-mail: Aso M. Aladdin, Department of Computer Science, College of Science, Charmo University, Sulaymaniyah, Chamchamal, Iraq. E-mail: aso.aladdin@chu.edu.iq

Received: 12-08-2025 Accepted: 08-11-2025 Published: 30-11-2025

definition describes AI as a collection of technologies [2]. These technologies enable computers to do a very abstract set of high-level things. These operations involve vision perception, language understanding and interpretation, data examination, decision-making, and others [3], [4]. One of the numerous uses of AI is creating software that will determine the best solutions and answers. These software programs assist in performing work and decision-making operations.

According to this, generally, a chatbot is a widely recognized software tool designed to replicate human-like discussions with users, typically using text or voice interactions [5]. Traditional chatbots rely on predefined rules and scripted responses, whereas more modern systems increasingly influence AI, particularly conversational AI approaches such as natural language processing (NLP), to read and respond to user input more effectively. These AI-powered chatbots can interpret user input in a more flexible and context-aware manner, enabling them to generate more accurate and dynamic responses, thereby improving user engagement and automating customer support more effectively [6]. Good chatbots can be incorporated into the organization's existing software to augment its applications. The addition of a chatbot to Microsoft Teams, for example, can make for a lively zone for content, tools, and team members to convene for chats, meetings, and collaboration [7]. Therefore, a major distinction and widely used classification between chatbots is based on the way they support users in conversations and produce responses. The former are rule-based chatbots, which usually present the users with a list of options to choose from [8]. These chatbots are often deployed in simple applications, like for frequently asked questions (FAQs). The second type is AI-based chatbots that are based on an AI, NLP, and machine learning (ML) [9].

AI chatbots are software tools designed to operate using various languages. These tools can process and respond to questions in multiple languages, though they often face limitations in supporting all linguistic variations and dialects, as well as different knowledge about different languages. Several AI programs, such as ChatGPT (by OpenAI), DeepSeek, Gemini, Microsoft Copilot, and Claude, have been developed to perform diverse functions for different purposes. In general, AI refers to computer systems capable of executing tasks traditionally associated with human intelligence in different languages, including learning, reasoning, problem-solving, perception, and decision-making [10]. As a core branch of computer science, AI focuses on developing algorithms that enable machines to interpret their environment, learn from data, and make intelligent decisions to achieve specific goals [11,12].

These AI tools, which rely on chatbots, can understand multiple languages; however, their accuracy varies depending on the language and the level of specificity required for the knowledge involved [13]. Each language presents unique challenges, ranging from grammatical rules and literary conventions to scientific terminology and cultural context. This includes knowledge about notable figures such as scientists, poets, novelists, and historical contributors, as well as the influence of various countries on the language and its development [14]. For the purpose of this study, the research focused on analyzing the fundamental differences in ChatGPT's responses to the same multiple-choice questions when presented in different languages, using separate accounts and at different times. The objective was to evaluate the accuracy and consistency of ChatGPT's answers across these languages and to identify any variations related to linguistic or dialectal differences.

Due to this complexity, achieving complete accuracy in any language is impossible. Therefore, each AI tool must be individually tested and evaluated for its error rate and performance in different languages [15]. Given this problem, our study examines the truthfulness of ChatGPT in answering questions across different domains of expertise in the Kurdish language chatbot with specific emphasis on the Sorani dialect. The motivation for this research stems from the limited quality of information available in the Kurdish Sorani language when using chatbot tools, particularly ChatGPT. During testing, this issue became clear. Moreover, no prior studies have addressed this problem or developed benchmarks for comparison. Therefore, this study proposes a framework to evaluate performance differences across languages using the same methodological approach. Although ChatGPT supports numerous languages, this research specifically focuses on comparing the Kurdish Sorani language with Arabic dialects due to their linguistic proximity. As a result, the aim was to assess the quality and accuracy of responses generated by ChatGPT. The primary objective of this research is to evaluate the degree of improvement in the accuracy of this AI tool, particularly in terms of how effectively it can address both general and language-specific queries.

The satisfaction users experience with an AI tool is strongly influenced by how warm, competent, and socially present the tool appears across different languages. Furthermore, the impact varies depending on the specific dialect, as using dialect-related information makes the chatbot more relevant and useful [16], [17]. In line with the rapid advancement of AI technologies, this study selected a variety of questions

across multiple domains of general knowledge, along with specific questions directly related to the Kurdish language.

In addition, the research also examines how accuracy percentages vary across different conditions. These factors include the user account, geographical location, and query time. It also considers the dynamic learning behavior of the AI when answering the same questions in different languages. As the AI framework stays learning and improving with time, such factors can determine the form in which answers are presented. Based on this context, the research contributes to a better understanding of ChatGPT's performance, providing information about its weaknesses and strengths in responding to questions in Kurdish language with Sorani dialect and wider knowledge domains. To achieve this goal, the study presents several key contributions, which are summarized in the following findings:

- Identified several chatbot categories and examined how dialects are used within multilingual processing. Chatbot systems, including ChatGPT, were tested to evaluate their performance. ChatGPT was used as a well-known AI tool to assess its limitations in answering questions in the Kurdish Sorani dialect. Its responses were compared with those in Arabic for the same queries. The findings showed that ChatGPT produced varied answers across different user accounts and time zones, indicating inconsistency in its output. Thus, the error ratio (ER) ranged from 39.4% to 24.8%, with the lowest value indicating different accuracy and reliable performance.
- ChatGPT showed gaps in knowledge for certain aspects of Kurdish and Arabic, particularly in historical topics, classical poetry, and prominent Sorani poets. To evaluate this, several multiple-choice questions were tested, and some had no correct answers across all accounts, indicating a consistent lack of knowledge in these areas. Statistical analysis examined ChatGPT's performance for each question, both in aggregate across all accounts and individually per account.

The study is structured as follows: Section 2 reviews multilingual language background, Section 3 outlines the methodology, Section 4 presents results and analysis, Section 5 discusses the findings, and Section 6 concludes with key insights and recommendations.

2. BACKGROUND REVIEW

2.1. Performance Evaluation of AI Tools

Although few-shot English learning has been advanced by in-context learning methods that do not update parameters [18], [19], few-shot cross-lingual transfer remains uncharted [20]. While encouraging results in non-English languages are seen [21], [22], [23]. However, it remains unclear whether these in-context learning methods can outperform traditional fine-tuning approaches across different languages and task types. This reflects a chronic lack of balance in NLP research, with English-biased development dominating and limiting development in low-resource and multilingual settings.

The BUFFET benchmark (Benchmark of Unified Format Few-shot Transfer Evaluation) attempts to mitigate this by evaluating few-shot transfer on 15 tasks across 54 languages [24]. Nevertheless, existing challenges such as data paucity, linguistic diversity, and test bias [25] [26] limit unbiased evaluation. Thus, advancement on multilingual benchmarks like ChatGPT still remains crucial to achieve unbiased performance beyond English.

In NLP, chatbots focus on parsing user input and generating responses, particularly by controlling prior knowledge. Several bots analyze the keywords and phrases used by users and gradually learn to respond more accurately and relevantly over time. In a chatbot system, it is not necessary for users to speak using the same dialect or linguistic structure as the chatbot itself [27]. When a user submits a message, the chatbot must interpret and handle the input [28]. However, since the user's language is often natural and unstructured, the chatbot requires a means to translate this input into a structured format that a machine can comprehend and act upon. As a result, this procedure is known as natural language understanding, as detailed under the dialect language component in Table 1. The conversion of a user's communication into structured data is referred to as constructing a semantic frame. This entire operation requires numerous steps, which are depicted in a general workflow in Fig. 1.

Furthermore, hybrid Chatbot [29] is a capable conversational agent which combines the advantages of rule-based systems and AI-driven approaches to develop a stronger and more versatile user experience. It combines accuracy and consistency achieved through the application of predefined rules for dealing with structured queries, with the understanding and responding capabilities to complex unstructured queries made possible by ML and NLP [30]. The combination of the two provides the chatbot with the ability to manage both types of interactions, from static FAQs to a more engaging, responsive, and contextually aware dialog, all while being more efficient, scalable, and providing better user satisfaction.

TABLE 1: Essential elements in chatbot design for multilingual and dialect-specific communication

Component	Purpose	Key applications	
NLU	A specialized branch of NLP that focuses on enabling machines to interpret, comprehend and derive meaning from human language. Unlike basic NLP, which deals with syntax and structure, NLU delves into intent recognition, contextual understanding, and semantic analysis, allowing systems to grasp nuances even across dialects.	AI chatbots and virtual assistants	Understanding user queries.
		Sentiment analysis	Detecting emotions in text.
		Intent recognition	Mapping user input to actionable tasks.
Dialect training data	Training a chatbot involves exposing it to vast amounts of relevant data and leveraging ML algorithms to help it understand, interpret, and respond to user inputs effectively.	Customer support	Handling FAQs, troubleshooting, and ticket routing.
		Virtual assistants	Managing tasks such as scheduling, reminders, and smart home control.
		Electronic commerce	Assisting with product recommendations and order tracking.
		Healthcare	Providing symptom checks and appointment bookings.
		Sentiment analysis	Detecting user emotions for better engagement.
Optional translation layer	Converts dialectal input into a standardized language form, enabling the chatbot to accurately interpret and respond to user queries.	Global customer support	Resolves queries in multiple languages without manual intervention.
		Multinational enterprises	Facilitates internal communications across global teams.
Response generator	Replies in standard or dialectal form, based on user need.	Intent recognition	Maps queries to actions (Dialogflow, Rasa).
		Contextual memory	Retains conversation history
		APIs/Webhooks	Fetches real-time data (inventory, weather) for accurate replies

AI: Artificial intelligence, NLU: Natural language understanding, NLP: Natural language processing, APIs: Application Programming Interface

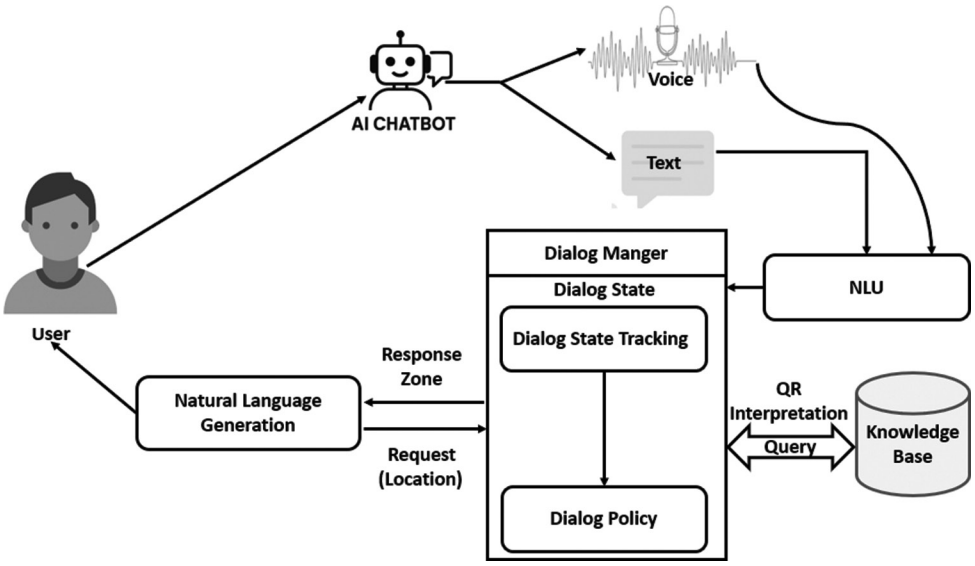


Fig. 1. The key elements involved in chatbot conversation processing.

To develop a multilingual chatbot with dialectal support, the system must include automatic language-detection modules, integrate translation APIs (Application Programming Interface), and be trained on datasets that contain diverse linguistic and dialect-specific samples [31]. Regional idioms and dialectal variations should be integrated through NLP

methods that enable appropriate comprehension and response within the parameters of the idiom or dialect from which they are expressed. To create or at least simulate a chatbot that understands dialects, it must possess the important elements given in Table 1.

As a result, the advantages and disadvantages of chatbot tools across different types have been recently surveyed. These studies highlight their potential to enhance personalized learning and student engagement. However, they also raise concerns, particularly regarding accuracy, cultural relevance, and ethical considerations.

2.2. ChatGPT Applications and Critical Analysis in Literature

The use of ChatGPT, an AI language model created by OpenAI, in Kurdish language education is a double-edged sword. Although ChatGPT performs well in languages with larger online corpora, its performance in less-represented languages, such as Kurdish, has not been as effective. It can serve as an autodidactic device, a dialog partner, and a creator of stimulating learning materials. This said, one should also be mindful of the challenges posed by engaging Chat GPT around human interaction, culture, mistakes and fixes, personal feedback [32]. Leveraging ChatGPT as a tool for learning the Arabic language in post-secondary education would provide fruitful opportunities in terms of learner experience on several platforms and within the existing operating systems and in terms of efficiency overall [33], [34].

It shows how, despite having the potential to improve language teaching by providing personalized feedback and tailored lessons, challenges posed by misinformation, absence of human interaction, and cultural sensitivity need to be solved, especially for less represented languages such as Kurdish [35]. Plus, in addition to the advantages of personalized learning and reduced tasks in education, ChatGPT is also a challenge due to the possibilities of academic dishonesty and issues related to identifying AI-generated text, thus threatening students' abilities to solve problems [36].

ChatGPT is an application of the transformer model for dialog generation and other NLP applications, such as question answering, developed by OpenAI. This is a language model for NLP tasks such as language generation developed by OpenAI [37]. It is capable of producing human-like conversational text. It is automatically generated, without human conversation or participation, chat in natural language, but it does not have human levels of understanding, empathy, or creativity [38]. Hence, ChatGPT is an OpenAI's state-of-the-art natural language model for natural conversational applications [39]. These applications and models, such as customer service, virtual assistants, and comparisons, use deep learning and NLP to provide human-like responses [39], [40]. ChatGPT, a generative AI chatbot, is another tool and resource that can aid language teaching and learning but must be used ethically and effectively with ethical and effective digital skills [41].

While ChatGPT may serve as a promising writing tutor for second language writing, more research is needed on the implications for writing pedagogy and for academic integrity [42]. Conversely, ChatGPT is a humanlike conversational agent based on the GPT architecture. Despite its limitations and ethical issues, it assists scientific studies by providing reviews of the literature, summaries of the content, and decisions in areas such as medicine and surgery [43].

The study examines the artificial language model ChatGPT, which generates human-like discursive text using deep-learning processes, and was developed by OpenAI. It engages as well with the opportunities and dangers associated with its use in an educational context [44]. The other looks into the use of ChatGPT in the item editing and modification process of high-stake testing, highlighting its use as an automated solution to a process within education and testing that essentially leads to enormous efficiency and cost benefits [45]. The model under investigation is ChatGPT, an OpenAI language model used for creative writing in poetry and prose. It explores how it works, what it takes into account when generating text, and the challenges of using it as a tool for creativity [46]. ChatGPT is a large language model for conversation developed by OpenAI. It is also human-like and biased because it has been trained on data generated by humans and then fine-tuned based on human feedback, which leads it to generate responses that are very much in line with what humans prefer and human cognitive biases [47]. The study analyses the application of GPT language models, namely ChatGPT, in higher education, focusing on their application as supports for engaging in theoretical exercises and laboratory practices and the questions of use and assessment in education. Effectiveness declines from theory to exercises to labs, as GPT struggles with specialized tools and software [48].

Critically, these literature studies present a balanced yet fragmented understanding of ChatGPT's role in education. Collectively, the studies highlight both its pedagogical benefits and ethical challenges. They emphasize potential for personalized learning, efficiency, and engagement but also reveal persistent limitations in accuracy, cultural sensitivity,

and academic integrity. Overall, the evaluations remain exploratory, requiring deeper empirical validation across diverse linguistic contexts.

3. METHODS

The present study takes a more systematic approach to gauge the accuracy and consistency of ChatGPT's responses to questions posed in the Kurdish language and specifically to the Sorani dialect. This process consisted of creating a set of multiple-choice questions, populating multiple test accounts with them, and exploring statistics on user performance. This approach also aims for cultural and linguistic relevance, testing reliability, and comprehensive evaluations along a variety of dimensions.

3.1. Data Collection

The study's dataset consists of 50 unique multiple-choice questions taken from a number of Kurdish language books. These questions are about geography, history, language structure, culture, and politics. There are four possible answers to each question, one of which is the actual answer. The data were handpicked to be culturally and linguistically relevant with the purpose of being useful for language assessment and educational use. The questions were hand-picked and screened for clarity, cultural, and linguistic appropriateness by native speakers of Kurdish in the Sorani dialect.

3.2. Data Preparation

Questions were assessed for clarity as well as for consistency in formatting. Response options were identified using the Kurdish alphabet (*a,b,c,d*), and all text was rendered in Unicode to help the display of the Kurdish script properly. The questions were entered into a database in a structured format suitable for processing and analysis.

The present study employs already established assessment frameworks through four distinct test accounts to measure ChatGPT's performance in the Sorani dialect Kurdish language. Each account contains 50 diverse questions that test different aspects of the knowledge of the Kurdish language. To obtain reliable, robust data, each question was asked ten times. The procedures for question formulation and the criteria employed for evaluating the responses are comprehensively outlined in the following section.

Fig. 2 shows the workflow of the experimental design to evaluate ChatGPT's performance in answering Kurdish (Sorani) questions. It starts with four independent ChatGPT accounts (A, B, C, D), which all receive the same set of 50 questions. Each account goes through the processing of the questions over several testing sessions. All the responses are kept in separate data sets and statistical analyses are performed on it. The result is an analysis of ChatGPT's correctness, types of errors, and consistency of performance, which offers new observations regarding how knowledge is represented by the model in the context of a low-resource language.

3.3. Experimental Design

To analyze the consistency and reliability of ChatGPT, four separate ChatGPT accounts (A, B, C, and D) were used in

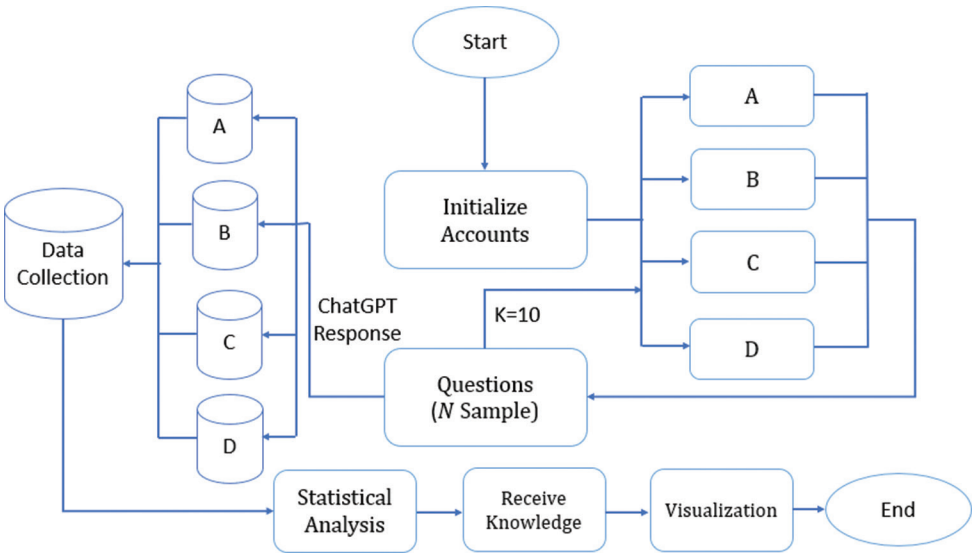


Fig. 2. Multi-account testing and data collection workflow.

the study. Both accounts underwent ten test runs on the complete dataset of 50 questions. This resulted in a total of 2000 responses (4 accounts × 10 sessions [Rounds] × 50 questions). Testing for each iteration occurred at varying times of day for each account over a total period of five days. Tests were either scheduled for morning hours (between 6 AM and 9 AM GMT) or evening hours (between 6 PM and 12 AM GMT). Questions were presented in a randomized order for each session to prevent any learning or memorization effects. The answers were all transcribed in a file for each account and session, which were ChatGPT's responses in each dialect of the Kurdish language. The questions, as noted above, were scientific, literary, and geographical. The process is depicted in Fig. 3. Then, the next step is to conduct the evaluation, which is described in the next sections, in terms of how the responses perform and what they look like.

3.3.1. Question-answer evaluation based on test rounds collections

As determined, the process begins with the initialization of four independent accounts (A, B, C, and D), each configured to perform the same set of tasks under identical conditions. The core of the research involves a loop that repeats ten times, ensuring consistency and reliability throughout the repeated testing process. Within each loop iteration, all four accounts simultaneously process a standardized set of 50 questions, designed to evaluate AI performance. The method reflects a systematic, repeatable testing structure that enables both cross-account comparison and longitudinal performance evaluation, contributing to the methodological rigor of the study.

According to this, to evaluate the responses generated by ChatGPT, a set of fifty questions was used, distributed across 40 test sessions conducted through four different user accounts. These sessions are used to assess the average performance of ChatGPT, with attention given to identifying any variations based on the time of testing or the specific user account. The evaluation results are presented using equation 1 as the average score per session or rounds, which highlights the performance outcomes for each individual test session.

$$Average = \frac{1}{n}\sum_{i=1}^{n} T_i \tag{1}$$

In this evaluation, n represents the total number of test sessions conducted, while T_i refers to the number of correct answers provided by ChatGPT during the *i–th* session. Each session corresponds to one instance of testing, and the value of T_i reproduces the model's accuracy within that specific session.

3.3.2. Question-answer evaluation based on question-level collections

Each question was answered in forty separate test sessions, resulting in a set of forty responses per question. This allowed for a comprehensive tracking of each question's performance across all test cycles. By analyzing these results, it was possible to identify which questions were consistently answered correctly and which were frequently misunderstood. This evaluation supported the classification of questions based on their difficulty and clarity, offering insights into how ChatGPT interprets and adapts its responses, especially when working with variations in question structure across different tests.

This method proved especially useful for assessing ChatGPT's accuracy in responding to questions in the Sorani

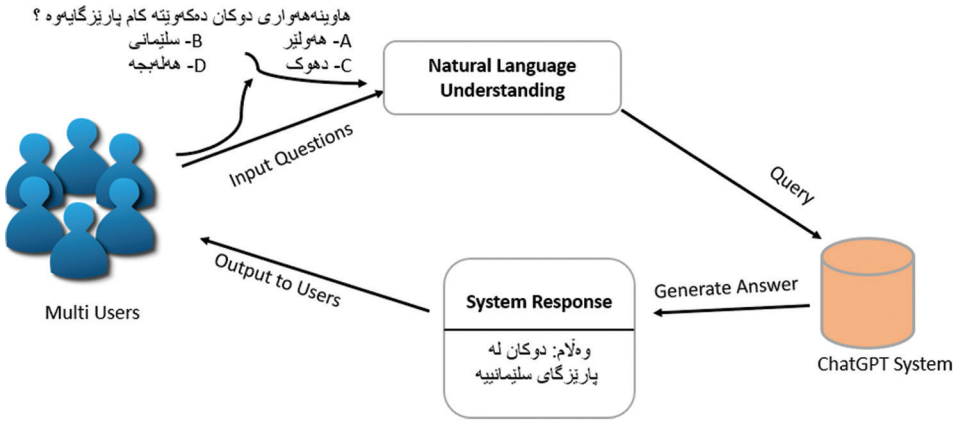


Fig. 3. The multiuser framework for question answering in Sorani dialect with ChatGPT.

dialect. Equation 2 was used to calculate the truth ratio (TR), representing the proportion of correct answers for each question or test session. Following this, Equation 3 was applied to determine the proportion of incorrect responses (ER), providing a balanced view of the model's strengths and limitations in handling Sorani language queries.

$$TR = \frac{T}{N} \tag{2}$$

In this context, T represents the number of correct responses, also referred to as "True" answers. N denotes the total number of responses collected. These two values are essential for calculating the accuracy or TR of the model's performance, as they allow for a straightforward comparison between the number of correct answers and the overall number of attempts.

$$ER = \frac{F}{N} = 1 - TR \tag{3}$$

In this evaluation, F represents the number of incorrect responses, also referred to as "False" answers. This value is used to measure the frequency of errors made by ChatGPT during the test rounds for each question. Question accuracy across all accounts is evaluated to determine how well each individual question performed throughout the test sessions. This is done using Equation 4, which calculates the accuracy of each question by dividing the number of correct responses by the total number of attempts. Specifically, T_j represents the number of correct responses to question j, and R_j denotes the total number of attempts for that same question. This measure provides insight into which questions were consistently understood and answered correctly by ChatGPT, helping to identify patterns in question difficulty and clarity.

$$Accuracy_{questionj} = \frac{T_j}{R_j} \tag{4}$$

3.3.3. Methodological validation framework

The multi-account, multi-round testing framework applied in this study was systematically structured and grounded in established scientific validation principles. Specifically, the design follows concepts of test–retest reliability and cross-account validation derived from psychometric evaluation and computational linguistics. These approaches are commonly employed to assess temporal consistency, reproducibility, and inter-sample stability in language evaluation research. Moreover, the quantitative assessment using TR, ER, and accuracy per session (Equations 1–4) ensures methodological rigor, transparency, and reproducibility by quantitatively verifying the stability and reliability of the model's responses across independent test accounts and repeated trials. This alignment establishes that the present evaluation framework adheres to recognized scientific standards of validity and reliability.

4. RESULTS AND ANALYSES

In response to the rapid advancement of AI technologies, this study systematically selected a comprehensive set of questions encompassing a wide range of general knowledge domains, alongside specialized items specifically pertaining to the Kurdish language. Furthermore, this study contributes to a more nuanced understanding of ChatGPT's performance, highlighting both its capabilities and its limitations in addressing inquiries formulated in the Sorani dialect of Kurdish, as well as in broader domains of general knowledge.

We fully understand the concern regarding the adequacy of using 50 multiple-choice questions to evaluate a large language model like ChatGPT. In our study, these 50 carefully selected questions were tested across four independent ChatGPT accounts, with each account completing ten repeated sessions under different conditions (locations, time zones, and testing rounds), resulting in 2,000 total responses ($4 \times 10 \times 50$). The repetition of the same questions across multiple accounts and sessions was a deliberate methodological choice aimed at examining consistency, reliability, and temporal variation in ChatGPT's responses, rather than solely its breadth of knowledge. Interestingly, the results revealed that even when the same question was repeated, the model did not always provide the same answer across accounts or test rounds. For example, questions related to Kurdish classical poetry were not answered correctly by any of the accounts, highlighting a consistent gap in ChatGPT's knowledge of culturally specific content. This design allowed us to identify not only performance differences between accounts but also systematic weaknesses in certain knowledge domains. We will consider expanding the number and diversity of questions in future studies to provide broader coverage while maintaining cross-account validation.

4.1. Test Rate Evaluation

The comparison study conducted across the four test accounts, each exposed to ten repetitions per question, demonstrates a statistically significant variation in the consistency of response performance. These findings indicate the necessity for focused methodological interventions and model

optimization tactics to enhance the overall dependability and accuracy of the system's outputs.

Test results show that ChatGPT was more correct than incorrect answers, showing that this tool can be relied upon. Table 2 presents each account's average number of true and false responses to 40 test questions, as well as each account's ratio of true to false responses. As for the average of the tests, it is higher than the average of the true answers, being 34.875 and 0.6975, respectively, which indicates a strong tendency toward the correct answers. The lower average of 15.125 with a false response ratio of 0.3025, on the other hand, indicates false answers.

To show the direction of the TR and the ER, 40 tests were performed on a set of four accounts: A, B, C, and D; each of which was tested ten times over the course of 5 days, as outlined in the methods section. The TR starts relatively high at about 70% but then significantly drops between the first and the early tests where it approximates 55% in test 4. The ER, on the other hand, starts at around 30% and reaches its maximum of approximately 50% around there. Beginning with test 7, there is a dramatic turn in that the TR is significantly higher and remains consistent between 70% and 80% with the ER dropping low and stabilizing between 20% and 30%. The analysis maintains this correlation between the two ratios. The ratios are seen to stabilize during the tests, showing that the performance level is reached. The higher TR and lower ER, achieved posttest 7 in Fig. 4, indicate that the subject was overcoming some difficulties.

TABLE 2: The average response to all test questions

Test case	Test average	Ratio
True answer	34.875	0.6975
False answer	15.125	0.3025

Four accounts were compared in terms of their reaction to a series of test questions, and the results differed both in accuracy and the rate of errors. The analysis is based on the average number of true and false answers given by each account, along with the TR and the ER percentages, respectively. There is also an overall average for all four accounts to provide a point of comparison. A segmentation for this breakdown is presented in Table 3.

The performance of four different accounts has been evaluated based on their responses to a series of test questions, revealing notable variations in accuracy and error rates. When examining the true answers, the average number of correct responses increases progressively across the accounts. *Account_A* demonstrates the lowest average with 30.3 correct answers, while *Account_D* achieves the highest with 37.6. This steady improvement indicates a gradual enhancement in performance from *Account_A* to *Account_D*. Conversely, the false answers show an inverse pattern: The average number of incorrect responses decreases from 19.7 for *Account_A* to 12.4 for *Account_D*. This inverse relationship between correct and incorrect answers supports the conclusion that *Account_D* performs the best, while *Account_A* leftovers the most.

Further evidence is provided by the TR and ER, which reinforce these observations. The TR, representing the percentage of correct answers, increases from 60.60% for *Account_A* to 75.20% for *Account_A*. Similarly, the ER decreases from 39.40% to 24.80% over the same range.

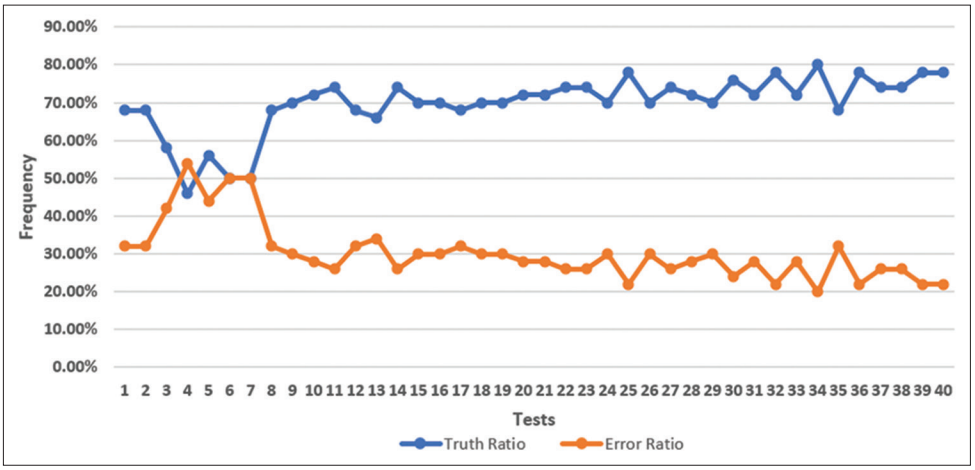


Fig. 4. The frequency of correct and incorrect responses across all tests.

TABLE 3: The average responses to test questions for each account

Test case	*Account_A*	*Account_B*	*Account_C*	*Account_D*	All
True answers	30.3	35.1	36.5	37.6	34.875
False answers	19.7	14.9	13.5	12.4	15.125
Truth ratio	60.60%	70.20%	73.00%	75.20%	0.6975
Error ratio	39.40%	29.80%	27.00%	24.80%	0.3025

These trends confirm the relative performance levels of each account and highlight the potential for targeted improvements. An overall average across all accounts is also included to serve as a baseline for comparison.

The performance of the four accounts has been tested by trial on ten consecutive average test questions and has been concerned with the ratio of correct to wrong answers. The comparison provides us with insight into how well each account performed over time and emphasizes fluctuation in accuracy. A clear visual representation of these trends is provided in Figs. 5 and 6. The primary purpose of these figures is to visually represent the fluctuations in the ratio of correct responses for each account, providing a comparative analysis of their performance over time. The proportion of correct responses is measured between 40.00% and 85.00%, and the testing axis indicates the time-wise arrangement of the ten tests. The performance of each account is plotted separately in a way that an explicit comparison and contrast could be shown for the test period.

Account_A exhibits the highest level of performance variability among the four accounts. It begins with an accuracy of approximately 70%, followed by a sharp decline to below 50% by the fourth test. This low level persists until the eighth test, after which a sudden rebound restores its performance to the initial range. Such fluctuations indicate a lack of stability in *Account_A*'s performance and may reflect underlying issues affecting its ability to generate accurate responses. In contrast, *Account_B* maintains a relatively stable accuracy throughout the testing period, fluctuating within a narrow range of 65% to 75%. Minor rises and dips are observed, but overall, the account demonstrates consistent performance with minimal variation, indicating reliability over time. In addition, *Account_C* also performs at a relatively high level, with accuracy predominantly exceeding 70%. Although it shows some moderate fluctuations, including a slight decline during the middle phase of testing, it recovers toward the end. This pattern suggests a generally resilient performance, with temporary setbacks that do not significantly impair overall accuracy. Finally, *Account_D* consistently achieves the highest performance among all

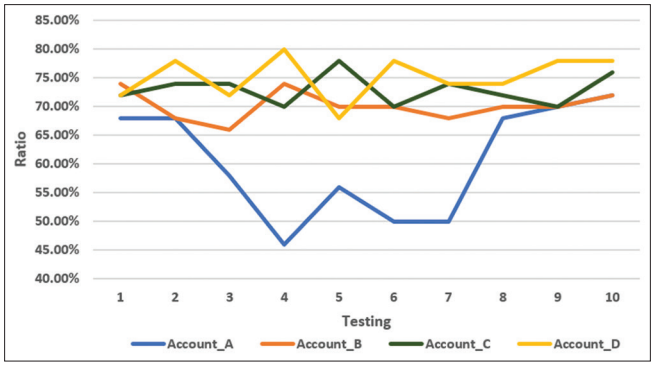


Fig. 5. The ratio of correct responses based on account testing.

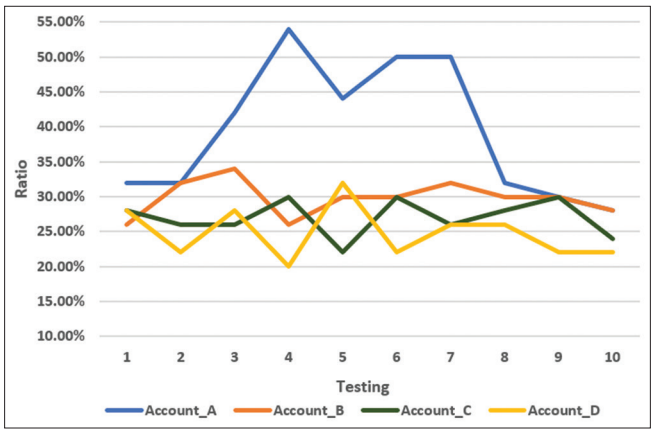


Fig. 6. The ratio of incorrect responses based on account testing.

accounts. Its accuracy remains above 70% across all tests, peaking at approximately 80% during the fourth test. While there are minor variations, its overall trend reflects a stable and robust level of accuracy, indicating dependable performance across the entire testing period.

These results highlight that each account produced different outcomes in response accuracy when evaluated using test questions in the Kurdish Sorani dialect through ChatGPT tools or different Chatbots. This fluctuation implies that AI performance is not uniform across accounts and may be influenced by training discrepancies or disparities in request zones. Consequently, these findings underline the need for

continued refining of AI technologies to ensure uniform and equal performance across languages and user scenarios.

4.2. Questions Rate Evaluation

Nonetheless, the use of interpreters for every dialect significantly impacts the chatbot responses and is an essential factor in training and handling user inputs of multiple languages. As a result, the system occasionally gives different responses to a repeated question of which some might be correct and the others not correct. Inconsistencies typically occur when the dialog manager tries to give responses without enough information about the context of the question.

In particular, ChatGPT exhibits notable limitations in handling the Kurdish language, particularly in the context of Kurdish Sorani literature, poetry, and the knowledge of Kurdish poets. This subsection evaluates the response patterns from four labeled accounts, focusing on how their answers vary, sometimes randomly, when presented with specific questions related to Kurdish Sorani. The discrepancies point toward the reality that ChatGPT does not have adequate knowledge in such domains and therefore generates faulty outputs. Therefore, this analysis provides valuable insights and proposes future research for enhancing AI performance on underrepresented languages. It also requests enrichment of training data sets and language resources, especially for a dialect like Kurdish Sorani, to render future AI responses more accurate and culturally appropriate.

The discrepancies point toward the reality that ChatGPT does not have adequate knowledge in such domains and therefore generates faulty outputs. Therefore, this analysis provides valuable insights and proposes future research for enhancing AI performance on underrepresented languages. It also requests enrichment of training data sets and language resources, especially for a dialect like Kurdish Sorani, to render future AI responses more accurate and culturally appropriate.

Accordingly, the valuation of results highlights how frequently various feature test cases occurred across all accounts, providing a quantitative view of scenario distribution and performance levels. Test case identifies each scenario of counting the number of question-response (QR), frequency shows its occurrence count of responses according to the testes, and ratio reflects its percentage of the total. All summarized data are shown in Table 4. In the feature test, all questions were answered correctly in 18 tests, representing 36% and incorrectly in 3 tests, representing 6%. These three questions, which consistently receive incorrect answers, are related to poems, poet names, and historical knowledge. This significant discrepancy suggests either a high level of accuracy in other areas or a strong bias toward providing correct responses elsewhere.

TABLE 4: The frequency ratio of feature test cases across all accounts for each question.

Test case	Frequency	Cumulative frequency	Ratio (%)	Cumulative percent (%)
No. of all QR truth 40	18	**18**	36	36
No. of all QR false 0	3	21	6	42
No. of QR up to 30 truth	13	34	26	68
No. of QR up to 20 truth	4	38	8	76
No. of QR up to 10 truth	6	44	12	88
No. of QR up to 0 truth	6	50	12	100

QR: Question-response

Moreover, the subsequent segmentation of test cases is based on the number of correct answers within a defined threshold. In particular, cases where the number of correct answers ranged from 30 to 39 occurred 13 times, accounting for 26% of all tests. Although this result based on all questions is considered acceptable, the outcome when considering only the questions with all query-related responses marked as true reaches approximately 61%. This indicates that achieving optimal performance does not necessarily lead to a higher chance of correctly answering the selected query-related questions.

Test cases with a maximum of 20 correct ones were seen 4 times, and these comprised 8% of all tests. Cases with a maximum of 6 and a maximum of 0 correct ones each also appeared 6 times, representing 12%, respectively. These categories enable us to take a more general look at distribution in performance, with differing rates of accuracy across accounts. Notably, earning up to 30 accurate answers showed the second-highest frequency after full-score situations, validating the observation of generally excellent accuracy. The results connected to QRs provide insight into performance trends and accuracy across tests, demonstrating the overall reliability of the tested accounts. However, many restrictions remain, particularly in generating accurate replies in Sorani Kurdish. This reflects limitations in ChatGPT's handling of some Kurdish dialects and domain-specific information, especially when striving for optimal answers across multiple user accounts and geographies.

One of the cases in the study illustrates the TR and ER for all questions across all accounts, based on the frequency of correct and incorrect QR responses over a set of 50 questions. The primary objective is to visually demonstrate the variation in the percentage of correct and incorrect answers across this range, as shown in Fig. 7.

It can be observed that with high TR, ER will be low, and low TR will result in high ER, which models an anticipated reverse relationship between correct and erroneous QR. The measures will normally spike to 100%, which indicates complete accuracy, yet fall to 0%, which indicates complete inaccuracy. Such extreme fluctuation reflects significant variation in performance, featuring normal swings between absolute accuracy and absolute error, consequently implying erratic response quality between accounts.

The analysis provides a comparison of feature test cases in four accounts. The main intention is to describe the distribution of each account's test result, including information about their relative performance and stability. It was found that QR frequencies may differ between accounts, with time zone and domain being potential deciding factors. Thus, the division is by the frequency at which each test case appeared by account or by the frequency at which each given result occurred. These results are summarized in Table 5.

The findings categorized the cases according to the total number of correct answers to QRs per account. *Account_B* has the highest count at 30 correct answers out of a total of 50 questions answered correctly, followed closely by *Account_C* at 28 and *Account_D* at 27 correct answers. *Account_A* has a relatively lower frequency of 18. For cases of all questions being answered wrongly, *Account_B* leads with 7 while *Account_A* indicates 6 and *Account_C* and *Account_D* each indicate 5. This implies that while *Account_B* is the one with the highest number of perfectly correct answers, it also indicates a relatively higher rate of overall failure among the accounts. In addition, Table 5 segment splits tests into various groups based on the number of QRs in given intervals. *Account_D* has at most 5 correct test scores with 9; *Account_A* and *Account_C* have 8 tied, and *Account_B* has 3. *Account_A* and *Account_C* have <5 correct test scores with 10 and 9, respectively, *Account_B* has 9, and *Account_D* has 5, meaning *Account_D* with fewer low scores. For more precise 5 correct answers for examinations, the top frequency is in *Account_A* with 8, then *Account_D* with 4, *Account_B* with 1, and *Account_C* with zero, indicating high heterogeneity across accounts.

The frequency of correct responses across four different accounts over a series of 50 questions is illustrated in Fig. 8. The graph provides a visual comparison of how the number of accurate answers varies for each account, highlighting differences in performance throughout the set. Conversely, Fig. 9 displays the opposite of Fig. 8, indicating the frequency of wrong replies across the QRs for the four different accounts. The frequency reflects the number of correct or

Fig. 7. The rate of correct and incorrect responses for all questions across all accounts.

TABLE 5: The frequency ratio of feature test cases across different accounts

Test cases	*Account_A*	*Account_B*	*Account_C*	*Account_D*
No. of all QR truth 10	18	30	28	27
No of all QR. false 0	6	7	5	5
No. of QR up to 5 truth	8	3	8	9
No of QR less than 5 truth	10	9	9	5
No of QR equal to 5	8	1	0	4

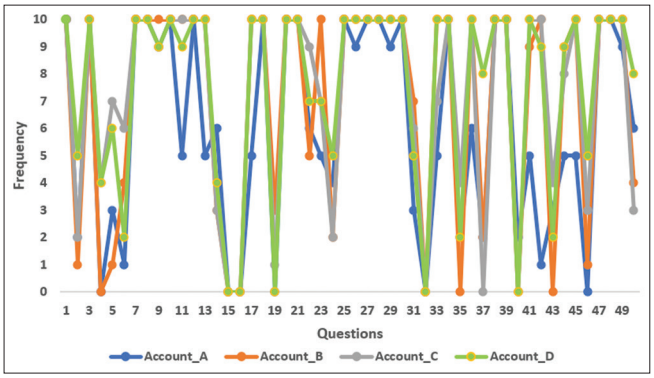


Fig. 8. The frequency of truth responses for each different account.

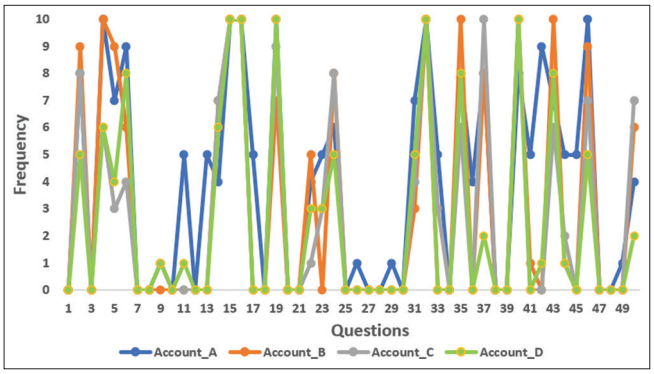


Fig. 9. The frequency of false responses for each different account.

incorrect responses, ranging from 0 to 10 tests per account over 50 distinct questions. Each account's performance is represented independently, allowing for easy separation and comparison throughout the question sequence.

The response patterns of the four accounts are very different in consistency and response correctness. *Account_A* is the most variable with large oscillations between 0 and 10 correct responses, which show extreme inconsistency although its correct responses show fewer errors than others. *Account_B* is also fluctuating but less wildly and in a pattern that indicates a slightly more stable but still fluctuating performance. *Account_C* has only moderate volatility, with most times having a higher proportion of correct responses with fewer instances of zero, indicating more stable accuracy. *Account_D* is characterized by the most stable and consistently high level of performance, staying very close to 10 correct responses on all accounts, indicating solid and consistent ability.

The results clearly show that the interpretation of queries by chatbots using ChatGPT varies significantly, particularly when different dialects are involved. This variation is especially evident in the test queries analyzed during this investigation. The accuracy and consistency of responses differed depending on the dialect used, highlighting challenges in understanding and processing language variations. In addition, external factors like the time zone that the test was aimed at and response time per question also contributed. These could have skewed the performance of the chatbot, leading to variations in results between accounts.

5. DISCUSSION

As demonstrated in the previous section, the analysis focused exclusively on evaluating responses in the Kurdish Sorani language. This area lacks prior research and serves as a new benchmark for assessing ChatGPT's performance compared with earlier versions. The identified limitations highlight the need for deeper analysis by comparing responses in Kurdish with those in Arabic using the same set of questions. Such a comparison provides a structured framework for linguistic evaluation. The findings also support the academic value of this framework in education, showing how AI tools can enhance teaching methods, classroom engagement, and lesson planning.

The comparative model of Kurdish and Arabic results has significant ramifications for the academic use of ChatGPT. Kurdish responses are both less truth-accurate, 69.75% and more error-prone, 30.25% than Arabic, 73.6% truth, 26.4% error on four accounts. This variation reflects language-resource disparities as well as model optimization differences for Kurdish as a low-resource language. Researchers can utilize these findings to understand ChatGPT's disparate performance across languages, with realistic classroom use expectations. For classroom use, Arabic users are able to anticipate more accurate content generation, while Kurdish instructors must utilize critical validation and further clarification to achieve content quality. The results indicate ChatGPT's promise as a complementary pedagogical tool for idea generation, translation, and text support but emphasize language-based testing and ethical revision to prevent misinformation and ensure inclusivity in linguistically diverse classrooms. These results are clearly reflected in the benchmark comparison presented in Table 6.

The statistical comparison between Arabic and Kurdish language answers in Table 7 recognizes that the two languages have similar patterns, although the differences are not conclusive but recognize major challenges for AI optimization when it works in multilingual settings. Kurdish

TABLE 6: A framework comparison between two languages (Kurdish and Arabic) during answer generation across four selected ChatGPT accounts

Test case	*Account_A*		*Account_B*		*Account_C*		*Account_D*		All	
	Kurish Lang	Arabic Lang	Kurish Lang	Arabic Lang	Kurish Lang	Arabic Lang	Kurish Lang	Arabic Lang	Kurish Lang	Arabic Lang
True answers	30.3	31.8	35.1	36.2	36.5	38.9	37.6	40.3	**34.875**	**36.8**
False answers	19.7	18.2	14.9	13.8	13.5	11.1	12.4	9.7	**15.125**	**13.2**
Truth ratio	60.60%	63.60%	70.20%	72.40%	73.00%	77.80%	75.20%	80.60%	**0.6975**	**0.736**
Error ratio	39.40%	36.40%	29.80%	27.60%	27.00%	22.20%	24.80%	19.40%	**0.3025**	**0.264**

TABLE 7: ChatGPT response accuracy and frequency comparison between Kurdish-Sorani and Arabic languages

Test case	Kurdish Lang			Arabic Lang.		
	Frequency	Ratio (%)	Cumulative percent (%)	Frequency	Ratio (%)	Cumulative percent (%)
No. of all QR truth 40	18	36	36	19	38	38
No. of all QR false 0	3	6	42	4	8	46
No. of QR up to 30 truth	13	26	68	15	30	76
No. of QR up to 20 truth	4	8	76	5	10	86
No. of QR up to 10 truth	6	12	88	3	6	92
No. of QR up to 0 truth	6	12	100	4	8	100

answers were negligibly lower in TRs in some cases, while Arabic answers underwent mild changes, which depicts ChatGPT's less consistent management of low-resource languages like Kurdish. These variations cause high-impact problems in learning use cases with content requirements for precision and reproducibility. To counteract this, future optimization of ChatGPT needs to focus on enhancing its understanding of Kurdish morphology, syntax, and context semantics with maintenance Arabic accuracy. Providing the AI with the same standardized questions across each language, alongside properly constructed multilingual data sets, can aid in tuning performance and reproducibility. In education, these multilingual AI materials can be used to prepare lessons, classroom lectures, and homework, offering equal learning chances for every language. Table 7 indicates comparative results, showing highly distinguishable differences in truth frequency and ratio that result in targeted improvements.

6. CONCLUSION

This work presents a comparative analysis of ChatGPT's performance in responding to multiple-choice questions (QR sets) in the Kurdish Sorani language within a methodologically controlled testing environment. The research also compares these responses with Arabic-language QR sets, rendering it feasible to critically test ChatGPT's performance across different multilingual settings. The technique provides valuable suggestions to users, particularly lecturers in higher learning institutions, on how best to utilize ChatGPT and also reveals potential limitations for developers to address. Through the experimentation of 50 expertly drafted questions on four user accounts over a number of sessions, the study reveals the strengths and weaknesses of ChatGPT as a tool for underrepresented languages. Some of the questions had no proper answers in all accounts, reflecting systematic deficiencies in the model's understanding. Interestingly, wrong answers were not randomly produced; instead, the AI reused the same mistakes again and again, exhibiting recognizable patterns in its performance instead of guessing.

The overall accuracy rate of around 70% reveals an effective baseline of performance, demonstrating that ChatGPT, when employed through chatbots in multiple interpretation situations, displays a core comprehension of the Sorani dialect and general knowledge themes. However, the results also highlight substantial differences across different accounts and test sessions, suggesting that ChatGPT's effectiveness can be influenced by factors such as time, session context, and user-specific activities. These findings emphasize the importance of evaluating AI systems not only in terms of average accuracy and statistical comparisons between languages, such as Kurdish and Arabic, but also with respect to their consistency and reliability across different testing conditions.

Furthermore, the study highlights significant limitations of the model in handling culturally specific or historically nuanced

content, especially in areas such as classical Kurdish literature and regional history. The gap itself is a general issue in AI design around the underrepresentation of minority cultures and languages in training datasets. While the model did not make completely random mistakes and would replicate very similar errors, it did not manage to be as profound in sense as it needed to be for more specialized fields.

Even within such constraints, a steady increase in trends of the same type of response rates in repeated testing holds promise for adaptive learning or higher convergence toward user input in the long term. The study provides a balanced framework of appraisal through its multi-layered analysis, consisting of session trends, account comparisons, and QR assessments, and provides an accurate measure to examine AI performance under low-resource language conditions. In conclusion, while ChatGPT shows promise as a tool for language learning and information retrieval in Kurdish or Arabic, its current limitations warrant cautious integration, especially in educational or culturally sensitive applications. Future work enhancements need to go toward incorporating linguistic datasets, enhanced cultural training, and context-aware reasoning.

The conclusions drawn from this study provide valuable insights not only for developers and educators working with Kurdish but also for the broader discourse on ethical and inclusive AI development for marginalized languages and cultures. While the chatbot operates with a limited dataset and within a security framework utilizing symmetric algorithms, several limitations remain, highlighting areas for further improvement. Despite these constraints, the findings demonstrate the potential of ChatGPT in educational contexts, supporting teaching and learning activities through structured QR interactions [49]. Furthermore, the study underscores its applicability in evaluating the effectiveness and impact of AI chatbots in other domains, such as healthcare, particularly when assessed through QR-based testing. Overall, these results reinforce the importance of adapting AI systems to low-resource languages while maintaining ethical, reliable, and culturally inclusive practices. In addition, the differences between ChatGPT-4.0 and the newly released ChatGPT-5 can be evaluated through a future comparative study.

REFERENCES

[1] N. Rane. "ChatGPT and similar generative artificial intelligence (AI) for smart industry: Role, challenges and opportunities for industry 4.0, industry 5.0 and society 5.0". *SSRN Electronic Journal*, vol. 2, no. 1, pp. 10–17, 2024.

[2] K. Ofosu-Ampong. "Artificial intelligence research: A review on dominant themes, methods, frameworks and future research directions". *Telematics and Informatics Reports*, vol. 14, p. 100127, 2024.

[3] M. J. Sousa, S. Pani, F. dal Mas and S. Sousa. "*Incorporating AI Technology in the Service Sector*. Apple Academic Press, New York. 2024.

[4] M. M. Maas. "Concepts in advanced AI governance: A literature review of key terms and definitions". *AI Foundations Report,* vol. 3, 2023.

[5] A. Casheekar, A. Lahiri, K. Rath, K. S. Prabhakar and K. Srinivasan. "A contemporary review on chatbots, AI-powered virtual conversational agents, ChatGPT: Applications, open challenges and future research directions". *Computer Science Review*, vol. 52, p. 100632, 2024.

[6] A. Mary Sowjanya and K. Srividya. ividyaSrividyaSrividya24. em*Conversational Artificial Intelligence*. Wiley, New Jersey. pp. 713-725, 2024.

[7] D. Das, C. Prasad and J. Geetha. "Intelligent Conversational AI for Microsoft Teams with Actionable Insights". In: *2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS)*. IEEE. pp. 1-5, 2024.

[8] L. Jain, R. Ananthasayam, U. Gupta and R. Radha. "Comparison of rule-based chat bots with different machine learning models". *Procedia Computer Science*, vol. 259, pp. 788-798, 2025.

[9] E. Yurdakurban, K. G. Topsakal and G. S. Duran. "A comparative analysis of AI-based chatbots: Assessing data quality in orthognathic surgery related patient information". *Journal of Stomatology Oral and Maxillofacial Surgery*, vol. 125, no. 5, p. 101757, 2024.

[10] A. M. Saghiri, S. M. Vahidipour, M. R. Jabbarpour, M. Sookhak and A. Forestiero. "A survey of artificial intelligence challenges: Analyzing the definitions, relationships, and evolutions". *Applied Sciences*, vol. 12, no. 8, p. 4054, 2022.

[11] Y. Dong, J. Hou, N. Zhang and M. Zhang. "Research on how human intelligence, consciousness, and cognitive computing affect the development of artificial intelligence". *Complexity*, vol. 2020, pp. 1-10, 2020.

[12] I. H. Sarker. "AI-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems". *SN Computer Science*, vol. 3, no. 2, p. 158, 2022.

[13] A. P. Chaves, J. Egbert, T. Hocking, E. Doerry and M. A. Gerosa. "Chatbots language design: The influence of language variation on user experience with tourist assistant chatbots". *ACM Transactions on Computer-Human Interaction*, vol. 29, no. 2, pp. 1-38, 2022.

[14] K. Mageira, D. Pittou, A. Papasalouros, K. Kotis, P. Zangogianni and A. Daradoumis. "Educational AI Chatbots for content and language integrated learning". *Applied Sciences*, vol. 12, no. 7, p. 3239, 2022.

[15] A. Bewersdorff, K. Seßler, A. Baur, E. Kasneci and C. Nerdel. "Assessing student errors in experimentation using artificial intelligence and large language models: A comparative study with human raters". *Computers and Education: Artificial Intelligence*, vol. 5, p. 100177, 2023.

[16] N. Haristiani. "Artificial intelligence (AI) chatbot as language learning medium: An inquiry". *The Journal of Physics: Conference Series*, vol. 1387, no. 1, p. 012020, 2019.

[17] A. S. E. AbuSahyon, A. Alzyoud, O. Alshorman and B. Al-Absi. "AI-driven technology and chatbots as tools for enhancing english language learning in the context of second language acquisition:

A review study". *International Journal of Membrane Science and Technology*, vol. 10, no. 1, pp. 1209-1223, 2023.

[18] I. Beltagy, A. Cohan, R. Logan IV, S. Min and S. Singh. "Zero- and Few-Shot NLP with Pretrained Language Models". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, Stroudsburg, PA, USA: Association for Computational Linguistics. pp. 32-37, 2022.

[19] T. Shin, Y. Razeghi, R. L. Logan IV, E. Wallace and S. Singh. "Autoprompt: Eliciting knowledge from language models with automatically generated prompts". *arXiv preprint arXiv:2010.15980*, 2020.

[20] X. V. Lin, T. Mihaylov, M. Artetxe, T. Wang, S. Chen, D. Simig, M. Ott, N. Goyal, S. Bhosale, J. Du, R. Pasunuru,… & X. L*i*. "Few-shot learning with multilingual language models," *arXiv preprint arXiv:2112.10668*, 2021. doi: 10.48550/arXiv.2112.10668

[21] H. Naveed, A. U. Khan, S. Qiu, M. Saqib, S. Anwar, M. Usman, N. Akhtar, N. Barnes and A. Mian. "A comprehensive overview of large language models". *ACM Transactions on Intelligent Systems and Technology*, vol. 16, no. 5, pp. 1-72, 2025.

[22] B. K. Arif and A. M. Aladdin. "A comparative analysis of ChatGPT and traditional machine learning algorithms on real-world data". *Kurdistan Journal of Applied Research*, vol. 10, no. 2, pp. 93-118, 2025.

[23] E. A. Alomari. "Unlocking the potential: A comprehensive systematic review of ChatGPT in natural language processing tasks". *Computer Modeling in Engineering and Sciences*, vol. 141, no. 1, pp. 43-85, 2024.

[24] A. Asai, S. Kudugunta, X. Yu, T. Blevins, H. Gonen, M. Reid, Y. Tsvetkov, S. Ruder and H. Hajishirzi. "BUFFET: Benchmarking Large Language Models for Few-shot Cross-lingual Transfer". In: *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Vol. 1: Long Papers. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 1771-1800, 2024.

[25] P. Efimov, L. Boytsov, E. Arslanova and P. Braslavski. "*The Impact of Cross-Lingual Adjustment of Contextual Word Representations on Zero-Shot Transfer*". Springer, Berlin. pp. 51-67, 2023.

[26] E. Razumovskaia, I. Vulić and A. Korhonen. "Data augmentation and learned layer aggregation for improved multilingual language understanding in dialogue". In: *Findings of the Association for Computational Linguistics: ACL 2022*, Stroudsburg, PA, USA. pp. 2017-2033, 2022.

[27] W. Huang, K. F. Hew and L. K. Fryer., 2022.l Linguistics: ACL 2022tilingual language understanding in dialogue". In: sks". ta". uage acqguage acqe acqag*Journal of Computer Assisted Learning,* vol. 38, no. 1, pp. 237-257, 2022.

[28] G. A. Santos, G. G. de Andrade, G. R. S. Silva, F. C. M. Duarte, J. P. J. Da Costa and R. T. de Sousa. "A conversation-driven approach for chatbot management". *IEEE Access*, vol. 10, pp. 8474-8486, 2022.

[29] W. Maeng and J. Lee. "Designing a Chatbot for survivors of sexual violence: Exploratory study for hybrid approach combining rule-based chatbot and ML-based Chatbot". In: *Asian CHI Symposium 2021*. ACM, New York, NY, USA. pp. 160-166, 2021.

[30] A. Sakshi, T. Mehrotra, P. Tyagi, and V. Jain, "Emerging trends in hybrid information systems modeling in artificial intelligence". In: *Hybrid Information Systems*. De Gruyter, Germany. pp. 115-152, 2024.

[31] M. Orosoo, I. Goswami, F. R. Alphonse, G. Fatma, M. Rengarajan and B. Kiran Bala. "Enhancing Natural Language Processing in Multilingual Chatbots for Cross-Cultural Communication". In: *2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*. IEEE. pp. 127-133, 2024.

[32] J. Guerrero-Ibáñez, S. Zeadally and J. Contreras-Castillo. "Sensor technologies for intelligent transportation systems". *Sensors*, vol. 18, no. 4, p. 1212, 2018.

[33] H. Leunard, R. Rachmawati, B. N. Zani and K. Maharjan. "GPT Chat: Opportunities and challenges in the learning process of Arabic language in higher education". *Journal International of Lingua and Technology,* vol. 2, no. 1, p. 10, 2023.

[34] A. M. Aladdin, Y. N. Bakir and S. I. Saeed. "The effects to trend the suitable OS platform". *Journal of Advances in Natural Sciences*, vol. 5, no. 1, pp. 342-351, 2018.

[35] V. İnci Kavak, D. Evis and A. Ekinci. "The use of ChatGPT in language education". *Experimental and Applied Medical Science*, vol. 5, no. 2, pp. 72-82, 2024.

[36] F. Mosaiyebzadeh, S. Pouriyeh, R. Parizi, N. Dehbozorgi, M. Dorodchi and D. Macêdo Batista. "Exploring the Role of ChatGPT in Education: Applications and Challenges". In: *The 24th Annual Conference on Information Technology Education*. ACM, New York, NY, USA, pp. 84-89, 2023.

[37] V. Goar, N. S. Yadav and P. S. Yadav. "Conversational AI for natural language processing: An review of ChatGPT". *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 3s, pp. 109-117, 2023.

[38] G. Sharma and A. Thakur. "ChatGPT in drug discovery". *ChemRxiv*, 2023.

[39] F. Fui-Hoon Nah, R. Zheng, J. Cai, K. Siau and L. Chen. "Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration". *Journal of Information Technology Case and Application Research*, vol. 25, no. 3, pp. 277-304, 2023.

[40] A. M. Aladdin, R. K. Muhammed, H. S. Abdulla and T. A. Rashid. "ChatGPT: Precision Answer Comparison and Evaluation Model". 2024. DOI: 10.36227/techrxiv.172833414.47483047/v1

[41] L. Kohnke, B. L. Moorhouse and D. Zou. "ChatGPT for language teaching and learning". *RELC Journal*, vol. 54, no. 2, pp. 537-550, 2023.

[42] J. S. Barrot. "Using ChatGPT for second language writing: Pitfalls and potentials". *Assessing Writing*, vol. 57, p. 100745, 2023.

[43] P. Zangrossi, M. Martini, F. Guerrini, P. De Bonis and G. Spena. "Large language model, AI and scientific research: Why ChatGPT is only the beginning". *The Journal of Neurosurgical Sciences*, vol. 68, no. 2, 2024.

[44] B. F. Gonçalves and V. Gonçalves. "*Artificial Intelligence Language Models: The Path to Development or Regression for Education*?" Springer, Berlin. pp. 56-65, 2024.

[45] D. S. M. Pereira, F. Mourão, J. C. Ribeiro, P. Costa, S. Guimarães and J. M. Pêgo. "ChatGPT as an item calibration tool: Psychometric insights in a high-stakes examination". *MedTeach*, vol. 47, no. 4, pp. 677-683, 2025.

[46] M. K. Audichya and J. R. Saini. "ChatGPT for Creative Writing and Natural Language Generation in Poetry and Prose". In: *2023 International Conference on Advanced Computing Technologies and Applications (ICACTA)*. IEEE. pp. 1-7, 2023.

[47] A. Azaria. "ChatGPT: More Human-Like Than Computer-Like, but Not Necessarily in a Good Way". In: *2023 IEEE 35th International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE.

pp. 468-473, 2023.

[48] H. Al-Khatri, N. Al-Azri, H. Hassan, R. A. Maamari and D. A. Khatri, "ChatGPT applications in active learning in higher education: A restricted systematic review," In: *Proceedings of the 11th International Conference Higher Education Advances* (HEAd), 2025. Available from: https://ocs.editorial.upv.es/index.php/HEAD/HEAd25/paper/view/20029

[49] D. O. Hasan, A. M. Aladdin, A. A. H. Amin, T. A. Rashid, Y. H. Ali, M. Al-Bahri, J. Majidpour, I. Batrancea and E. S. Masca. "Perspectives on the impact of E-Learning pre- and post-COVID-19 pandemic-the case of the Kurdistan Region of Iraq". *Sustainability*, vol. 15, no. 5, p. 4400, 2023.

RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Address Bar Spoofing in Contemporary Web Browsers: A Taxonomy, Exploitation Study, and Mitigation Guidelines

Renua Hiwa Ismael, Jaza Mahmood Abdullah

Department of Information Technology, College of Commerce, University of Sulaimani, Sulaymaniyah, Iraq

ABSTRACT

The browser address bar is the cornerstone of user trust and web security. Despite advancements, address bar spoofing remains a persistent threat, enabling attackers to make malicious URLs appear legitimate. This paper presents an extensive investigation into address bar spoofing vulnerabilities across modern desktop and mobile browsers. We introduce a comprehensive taxonomy classifying over 15 distinct spoofing techniques, many of which are novel. Across systematic testing, over 70 vulnerabilities were identified and responsibly disclosed, resulting in patches across more than 15 browsers. These findings are enumerated in this paper for verification. This research analyzes the root causes of these vulnerabilities, highlighting common pitfalls in URL parsing, display logic, and UI state management. Based on our findings, we propose a robust mitigation framework and best practices for browser developers, alongside actionable advice for users. Our findings underscore the ongoing challenge of maintaining address bar integrity and the critical need for continuous vigilance in browser security. A public repository documents these findings to aid further research.

Index Terms: Address Bar Spoofing, Web Browser Security, Phishing, URL Spoofing, UI Security, Vulnerability Taxonomy, Cyber Security

1. INTRODUCTION

Phishing attacks continue to be one of the most prevalent and damaging cyber threats, costing individuals and organizations billions of dollars annually [1]. Attackers employ various social engineering tactics, often relying on deceptive URLs to trick victims. While URL masking or cloaking techniques are common, a more insidious form of deception is address bar spoofing, where a vulnerability within the web browser itself is exploited to display a trusted origin (e.g., bank.com) while the user is actually interacting with a malicious site. This undermines the fundamental trust users place in the address bar as an authenticator of web identity [2].

Despite significant efforts by browser vendors to harden this critical UI component, our research demonstrates that address bar spoofing is not a solved problem. Modern browsers, with their complex feature sets and support for diverse URI schemes and UI states, introduce new attack surfaces. This paper details an extensive, multi-year investigation into these vulnerabilities across prevalent desktop (Windows, macOS) and mobile (iOS, Android) browsers, including Chrome, Firefox, Safari, Edge, Opera, Brave, Arc, and others.

A comprehensive dataset, including detailed descriptions, Proof-of-Concept (PoC) code, and replication videos for all 74 vulnerabilities discussed, is publicly available in our

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp335-345

E-ISSN: 2521-4217
P-ISSN: 2521-4209

Copyright © 2025 Ismael and Abdullah. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Renua Hiwa Ismael, Department of Information Technology, College of Commerce, University of Sulaimani, Sulaymaniyah, Iraq. E-mail: renwax23@gmail.com

Received: 12-07-2025 Accepted: 25-09-2025 Published: 30-11-2025

repository [3]. For direct verifiability within this paper, a summary of these findings is provided in Appendix A.

The primary contributions of this paper are:

1. Demonstrated Persistence and Novelty: We uncovered over 70 distinct address bar spoofing vulnerabilities, many leveraging novel or refined exploitation techniques, leading to direct security improvements in over 15 browsers.
2. Comprehensive Vulnerability Taxonomy: We developed a structured taxonomy classifying diverse spoofing techniques, moving beyond isolated bug reports to provide a systematic understanding of the attack surface.
3. Actionable Mitigation Framework: Based on identified root causes, we propose concrete mitigation strategies for browser developers and user education guidelines.

2. THE EVOLVING THREAT OF ADDRESS BAR SPOOFING

Address bar spoofing is not a new phenomenon. Chromium's guidelines for URL display [4] acknowledge many such scenarios. Early instances date back to vulnerabilities in browsers, such as Internet Explorer 6, exploiting dialog boxes or character encoding quirks [5]. Over time, attacks evolved to leverage weaknesses in handling HTTPS indicators with delayed responses [6], Unicode homograph attacks [7], race conditions in mobile browsers [8], manual spoofing tests [9], and flawed server-side redirect handling [10]. However, as browsers add new features (e.g., Picture-in-Picture, varied URI scheme support) and UIs evolve, new vectors for spoofing continually emerge, often due to subtle implementation errors or incomplete parsing and validation logic. Our research found that even browsers built on mature, security-focused engines, such as Chromium and Gecko are susceptible to these attacks when specific UI choices or feature integrations are made.

3. RESEARCH METHODOLOGY

Our research methodology involved a multi-staged approach conducted over a multi-year period (2019–2024). This section details our testing environment and the systematic process used for vulnerability discovery, analysis, and classification.

3.1. Threat Model and Assumptions

For this research, we assume a standard web-based threat model. The attacker is capable of hosting a malicious website and can persuade a victim to click a link leading to it (e.g., through a phishing email or social media message). However, the attacker cannot install malware on the victim's machine or compromise the underlying operating system. The victim is assumed to be a non-expert user who relies on the browser's UI, particularly the address bar, as a primary indicator of a website's identity. The goal of the attacker is to exploit a browser vulnerability to make their malicious site's address appear as a trusted, legitimate domain.

3.2. Testing Environment and Scope

Our investigation targeted a wide range of modern web browsers across major desktop and mobile platforms. We primarily focused on the latest stable and beta versions available during the testing periods to ensure relevance. The scope of our test environment is summarized in Table 1.

3.3. Analysis Stages

Our process consisted of five distinct stages:

1. Systematic Testing of Browser Features: We systematically tested browser features known to be complex or historically problematic. For each feature, we developed a test plan to probe edge cases. For instance, to test data: URI handling, over 50 permutations were crafted, varying the encoding, position of hostname-like strings, and special characters to observe parsing and display logic. Features tested included search engine integration, Address bar freezing, internal URI schemes, full screen/PiP APIs, blob/data/javascript URI schemes, RTL character display, and Unicode parsing.
2. Reverse-Engineering Browser Behavior: For identified anomalies, we analyzed browser behavior to determine the root cause. This sometimes-involved decompiling Android applications or examining open-source code. For example, to diagnose a search query spoof in Opera Mini (ID: 11), we used jadx-gui [11] to decompile the Android application package (APK). By analyzing the Java class responsible for parsing search provider URLs,

TABLE 1: Overview of the testing environment and scope

Browser (engine)	Versions tested	Operating systems (OS)
Google Chrome (Chromium)	130–139	Windows 10, macOS 15, Android 14, iOS 15-26
Mozilla Firefox (Gecko)	135–144	Windows 10, macOS 15, Android 14, iOS 15-26
Apple Safari (WebKit)	15–26	macOS 15-26, iOS 15-26
Microsoft Edge (Chromium)	130–139	Windows 10, macOS 12-14, Android 14, iOS 15-26
Opera/Opera GX (Chromium)	120–130	Windows 10, Android 14
Arc Browser (Chromium/WebKit)	1.0–1.10.1	Windows 10, macOS 15–26, iOS 15–26

we identified a flawed regular expression that failed to correctly validate the origin.
3. Exploitation and PoC Development: Once a potential vulnerability was identified, we developed a minimal PoC to confirm its exploitability and demonstrate its security impact.
4. Responsible Disclosure: All vulnerabilities were reported to the respective browser vendors through bug bounty programs (e.g., HackerOne, Bugcrowd) or direct security contacts. Public disclosure of details was withheld until patches were made available or a reasonable timeframe had passed.
5. Taxonomy Creation: Based on the diverse attack vectors and root causes, we categorized all confirmed vulnerabilities into the comprehensive taxonomy presented in Section 4.

4. A COMPREHENSIVE TAXONOMY OF ADDRESS BAR SPOOFING TECHNIQUES

Our research identified numerous ways attackers can manipulate the browser's address bar. The following taxonomy categorizes the primary techniques observed. For brevity, only a selection of categories and illustrative examples are presented.

To provide a more structured overview, Table 2 summarizes the key characteristics of the primary spoofing classes identified in our research, including their typical root cause and user interaction requirements.

4.1. UI State Manipulation

Exploiting browser states like Full-screen mode or features, such as Picture-in-Picture (PiP) to obscure or replace the authentic address bar.

- Full-screen Spoofing: An attacker's site requests full-screen mode upon user interaction. The browser UI, including the address bar, disappears. The attacker then renders a fake address bar image within the web content, tricking the user (Fig. 1). While some browsers warn upon entering full screen, many do not, or the warning is easily missed.
- Window without Address Bar (e.g., via PiP): Certain APIs, such as Picture-in-Picture, can create new windows or views without a native address bar. Attackers can then draw a fake address bar within this content area. We found Microsoft Edge on Mac vulnerable to this via PiP.

4.2. Non-Standard Scheme Handling

Abusing the browser's often less-rigorous parsing, validation, and display logic for URIs beyond standard http and https.

- Data: URI Spoofing: data: URIs embed content directly. We found that by crafting a data: URI with a spoofed hostname-like string before the actual comma-separated data (e.g., data://google.com/,<malicious_content>), some mobile browsers (such as Arc Mac or Opera GX Android) would misleadingly display google.com in the address bar. See Fig. 2 for an example in Arc Mac.
- Blob: URI Spoofing: blob: URIs represent raw data. Some browsers, when navigating to a blob: URI, display the full blob URL. By appending a hash fragment containing the spoofed URL (e.g., blob: https://real-origin.com/GUID#https://spoofed.com), browsers, such as Safari would prioritize displaying the fragment, effectively hiding the true blob: origin.
- JavaScript: URI Spoofing: Redirecting to or opening a JavaScript: URI can lead to confusion. In PlayStation 5's browser, redirecting to a legitimate site, then immediately opening a new window to a JavaScript: URI caused the address bar to retain the legitimate site's URL while executing an attacker-controlled script that wrote to the new window.
- Internal URI Schemes: Browsers use internal schemes (e.g., internal://, arcmobile-internal://) for error pages or special functions. Redirecting to these schemes, often with parameters controlling the displayed URL, allowed spoofing. For example, Firefox for iOS allowed redirection to internal://local/errorpage?url=https://

TABLE 2: Characteristics of spoofing technique classes

Taxonomy class	Primary root cause	Typical user interaction	Common dependency
UI State Manipulation	Hiding/replacing browser chrome	One click (e.g., to enter full screen)	Browser/OS specific APIs (PiP, Fullscreen)
Non-Standard Scheme Handling	Lax parsing of non-HTTP (S) URIs	None (on link click)	Browser-specific scheme implementation
Navigation Timing	Premature address bar update	None (on link click)	General browser navigation logic
URL Parsing and Display	Flawed truncation or display logic	None	Logic, especially on mobile UI
Unicode & RTL	Incorrect text rendering	None	Unicode/BiDi rendering engine
Internal Component Interaction	Privilege boundary violation	None (on link click)	Browser-specific internal pages/APIs

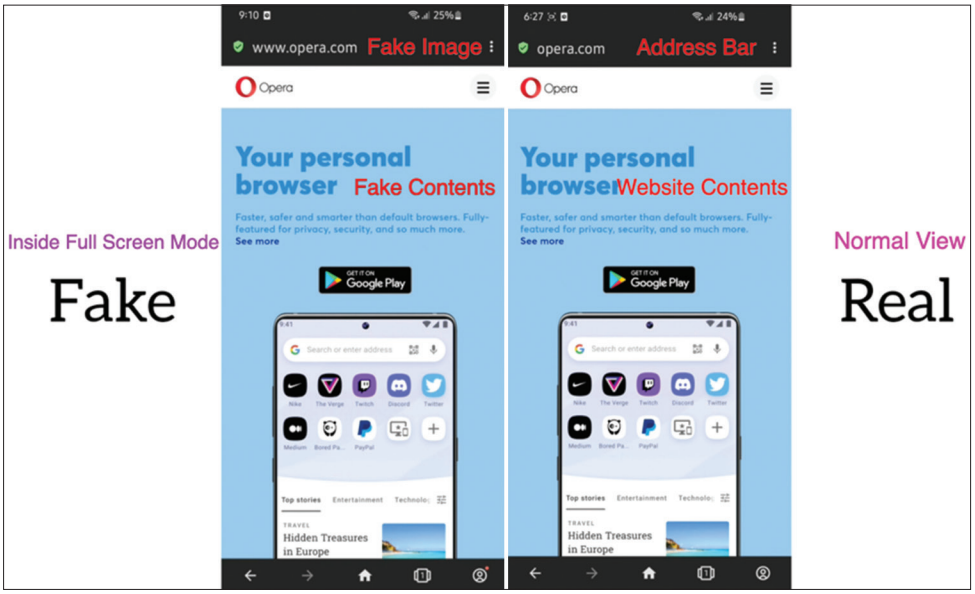


Fig. 1. Demonstration of Fullscreen Spoofing in Opera Mobile. (Left) The attacker's page (attacker.com) enters fullscreen, hides the real UI, and renders a fake address bar image showing opera.com. (Right) The legitimate opera.com site shown in the standard browser view for comparison.

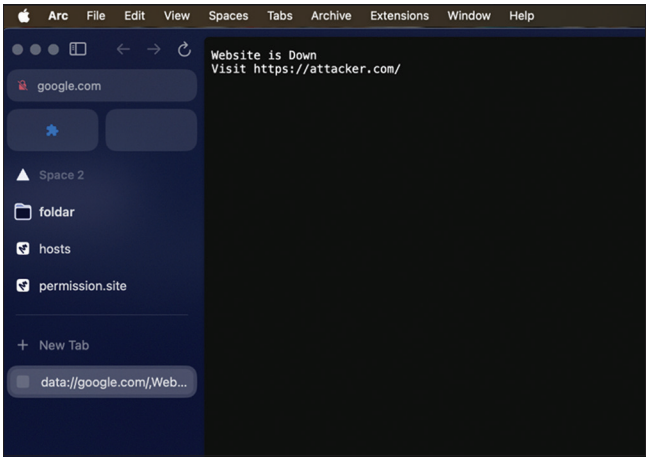


Fig. 2. Exploitation of data: URI parsing in Arc for macOS. The address bar misleadingly displays google.com, a string parsed from the URI path while the page content is fully controlled by the attacker via the data: URI itself.

google.com/., which would display google.com while allowing content.

4.3. Navigation Timing and Race Conditions

Capitalizing on the temporal gap between initiating a navigation request and receiving a definitive server response.

- URL Freeze (Change Before Response): The most common vulnerability. An attacker redirects to an unreachable port on a legitimate domain (e.g., google.com:8081). Many browsers immediately update the address bar to google.com *before* realizing the port is unreachable. The page content remains attacker-controlled, leading to a spoof (Fig. 3). This was observed in numerous browsers, including Arc Android and Safari.
- HTTP 204 No Content: This technique is distinct from URL Freeze: while URL Freeze exploits an unreachable destination, the HTTP 204 attack uses a valid but empty server response. The browser correctly updates the address bar to the target URL, but because no new content is sent, the attacker's original page content remains visible. This was observed in Arc for Android.

4.4. URL Parsing and Display Logic Flaws

Leveraging ambiguities in how browsers parse, truncate, and display complex URLs.

- Long Domain/Subdomain to Hide Origin: On small screens (especially mobile), long subdomains (e.g., very.long.subdomain.target.com.attacker.com) can cause the browser to truncate the URL, potentially only showing very.long.subdomain.target.com and hiding attacker.com. Firefox for Android/iOS was vulnerable, prioritizing the beginning of the URL, PoC in Fig. 4.
- Search Engine Integration Quirks: Some mobile browsers display the search query in the address bar instead of the search engine's URL. Flaws in validating the search engine's origin or path allowed spoofing. For example, a malformed regex in Opera GX allowed google.pwn.wtf to be treated as a Google search, displaying the query from ?q= as the URL. The Arc browser for iOS was similarly vulnerable.

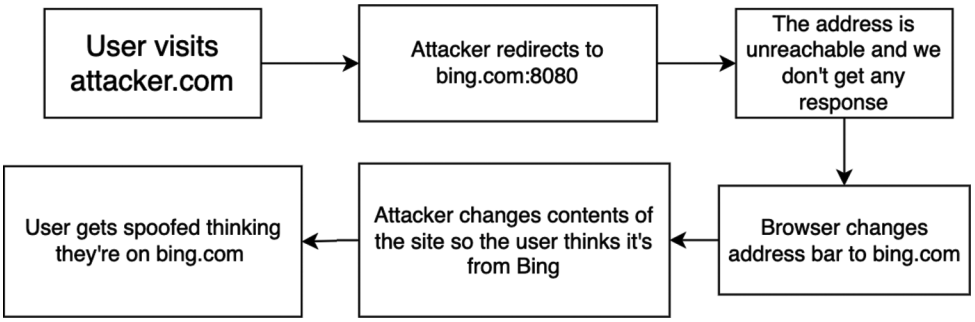


Fig. 3. Flowchart of a "URL Freeze" timing attack. The browser pre-maturely updates the address bar to a legitimate domain (bing.com) upon redirection, but the destination port is unreachable, causing the navigation to fail while leaving the spoofed address and attacker-controlled content visible.

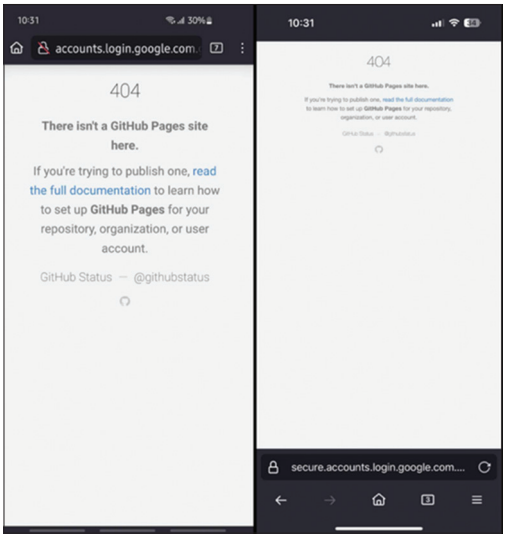


Fig. 4. Long Subdomain Spoof in Firefox Mobile. The browser truncates the long URL, prioritizing the display of the fake subdomain (accounts.login.google.com…) while completely hiding the true, malicious origin (…github.io).

4.5. Unicode and Right-to-Left (RTL) Manipulation

Utilizing visually confusable Unicode characters or exploiting complex rendering rules for mixed RTL/LTR text.

- Unicode Whitespace-like Characters: Characters like U+08FC can appear as whitespace. Bentkowski [12] showed this could be used in domains, such as important-domain.google.com<U+08FC><U+08FC>evil.com, where Chrome/Firefox would display only the important-domain.google.com part.
- RTL/LTR Confusion: Mixing RTL (e.g., Arabic, Kurdish) and LTR (e.g., English) scripts in a domain can confuse the browser's display logic, causing parts of the URL to be reordered or hidden, leading to spoofing. Brave for iOS was vulnerable when displaying domains like https://xn--llb.login.accounts.google.com.xn--llb.attacker.com, misrendering the true origin.

4.6. Browser Internal Component Interaction

Exploiting vulnerabilities within or between privileged browser components.

- Internal Pages XSS: An XSS in a privileged internal browser page (e.g., arcmobile-internal://in Arc iOS) responsible for rendering UI elements or handling errors could be used to manipulate the address bar display by controlling parameters passed to this internal URI.

5. RESULTS AND DISCUSSION

Our process involved identifying and classifying each vulnerability, reporting it to the vendor for patching, documenting the entire procedure along with the root causes that are shared among them, and then assigning severity ratings of each spoof.

5.1. Statistical Overview

Our research initiated in 2019 and intensified recently, uncovered over 70 unique address bar spoofing vulnerabilities across 15+ different web browsers and browser-based applications on various platforms. This led to over $80,000 in bug bounty reward. Table 3 summarizes the number of unique vulnerabilities found in selected major browsers (mobile versions included).

5.2. Key Root Causes

Several recurring themes emerged as root causes:

1. Premature Address Bar Update: Changing the address bar value *before* navigation is complete and a response is received (most common), URL freezing and navigation timing.
2. Window/View Without Address Bar: Features creating content views (PiP, certain extension APIs) without mandating a secure, native address bar.

TABLE 3: Number of vulnerabilities found in major web browser vendors

Desktop+Mobile	Number of unique vulnerabilities found
Chrome	4
Firefox	9
Safari	9
Brave	5
Edge	6
Opera/Opera GX	12
Yandex	4
Arc	10

A "unique vulnerability" is defined as a distinct root cause, even if it manifested in multiple browsers based on the same engine (e.g., a single WebKit bug affecting both Safari and Arc for iOS)

3. Flawed URL Truncation/Display Logic: Incorrectly prioritizing the beginning/end of long URLs or failing to clearly indicate the true registrable domain.
4. Insufficient Prompts/Notifications: Lack of clear user notification when entering states, such as full screen, where the address bar is hidden.
5. Insecure Handling of Search Engine Queries: Weak validation of search engine origins/paths when displaying queries in the address bar.
6. Lax Parsing of Non-HTTP(S) Schemes: Treating schemes, such as data:, blob:, or internal URIs, with less rigor than http(s), leading to misinterpretation of origin.
7. Incorrect RTL/Unicode Character Handling: Failure to correctly parse and display mixed-direction text or confusable Unicode characters in domain names.

5.3. Exploitability and Severity

Vulnerabilities ranged in severity:

To move beyond qualitative descriptions, we developed a bespoke scoring rubric to consistently assess the severity of each vulnerability. This framework, summarized in Table 4, was inspired by CVSS but adapted for UI spoofing, scores vulnerabilities on two primary factors: Visual Fidelity and Interaction Complexity.

Severity is assigned based on the sum of the two scores (ranging from 2 to 6):

- High severity (Score 5–6): These vulnerabilities achieve near-perfect mimicry with minimal or no user interaction. They pose the most significant threat. Example: The Safari about: blank spoof (Fig. 5) was pixel-perfect (Score 3) and required a single click (Score 2), for a total score of 5.
- Medium severity (Score 3–4): These spoofs are convincing but have minor visual flaws or require more obvious user interaction. Example: The Brave iOS blob: URI spoof required one click (Score 2) but had minor visual inconsistencies (Score 2), for a total score of 4.
- Low severity (Score 2): These attacks are less realistic, rely on uncommon browser states, or require significant user interaction, making them less likely to succeed. Example: The Opera GX Unicode freeze required specific text selection (Score 1) and was visually flawed (Score 1), for a total score of 2.

Browsers, such as Opera and Arc, which implement many new features and adjust the address bar more toward UX and esthetic look, are more prone to spoofing attacks, even "imperfect" spoofs pose a real risk to unsuspecting users. Responsible disclosure timelines varied, with some vendors patching within days and others taking several months.

6. MITIGATION STRATEGIES

A multi-layered approach is necessary to effectively mitigate address bar spoofing, along with the browser vendor and user knowledge in the area.

6.1. Proposed Fixes and Recommendations for Browsers

1. Always Show Top-Level Domain Clearly:
   - If a URL is too long, prioritize displaying the TLD and the end of the URL path. Do not hide the TLD.
   - For multiple subdomains, ensure the registrable domain (e.g., example.com in sub.example.com) is always visible and clearly distinguished (Fig. 6).
2. Don't Change Address Value Before Response: Only update the address bar *after* a successful response from the destination server is received and content begins to render. For unreachable destinations or HTTP 204, either revert to the previous URL or clear content and display the new URL with an empty/error page (Fig. 7).
3. Show Prompts for UI State Changes: Clearly notify users when entering states, such as full-screen or landscape mode if the address bar is hidden.
4. Address Bar is for URLs, Not Search Queries: Ideally, always display the search engine's URL. If queries must be shown, use robust origin/path validation for the search engine.
5. Don't Expose Internal URIs to Web Pages: Prevent web content from redirecting to or interacting with privileged internal browser URIs.
6. Every Window Must Have a Secure Address Bar: All windows/tabs/popups/PiP displaying web content must have a non-spoofable, browser-controlled address bar

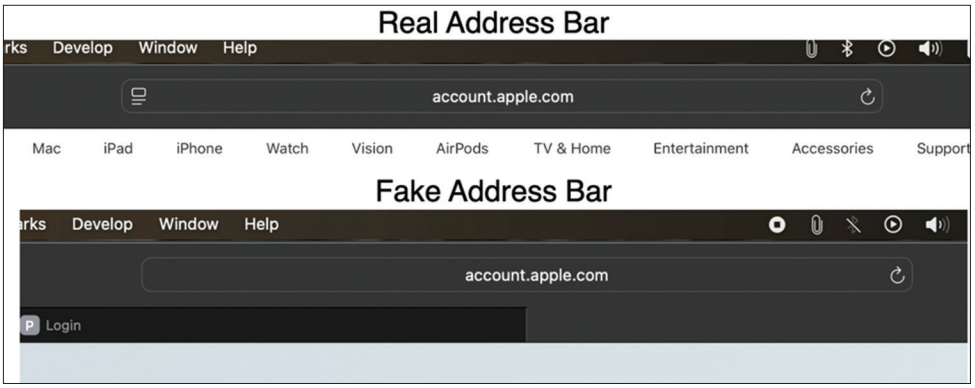


Fig. 5. High-Severity Spoof in Safari. A comparison shows an attacker-controlled page rendering a fake address bar (bottom) that is pixel-perfect and visually indistinguishable from the legitimate browser UI (top), demonstrating a highly convincing attack.

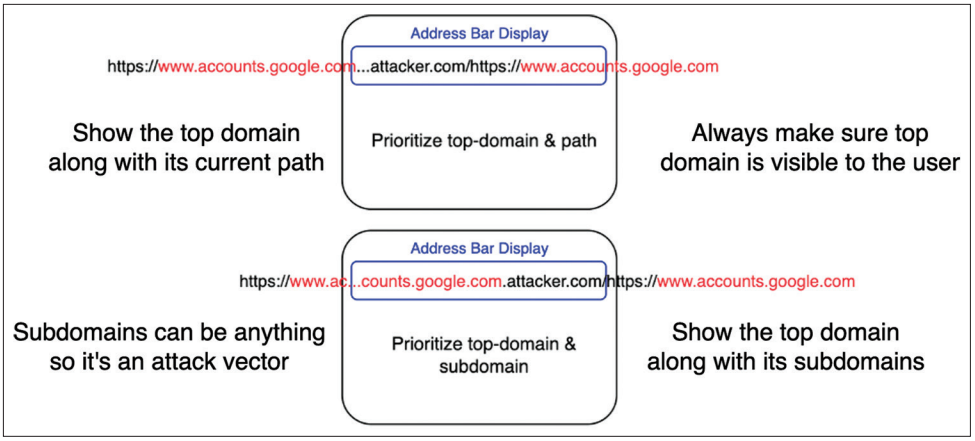


Fig. 6. Diagram illustrating the correct principle for displaying long URLs. The browser UI must always prioritize displaying the true registrable domain (attacker.com) rather than potentially deceptive subdomains, which can be arbitrarily long.

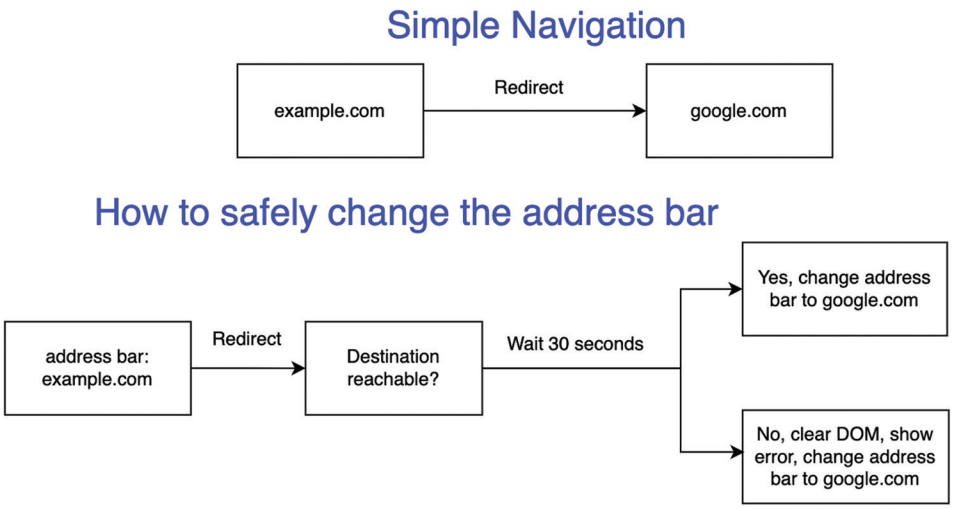


Fig. 7. Flowchart for a secure address bar update process. The UI should only change the displayed origin after receiving a successful response from the destination server, preventing timing attacks where the navigation fails.

TABLE 4: Severity Scoring Rubric for Address Bar Spoofing

Score	Visual Fidelity (How convincing is the spoof?)	Interaction Complexity (How much user action is needed?)
3	Pixel-Perfect: Indistinguishable from the legitimate browser UI to a typical user.	Zero-Click: Occurs automatically upon navigation or page load.
2	Convincing: Minor visual flaws or artifacts that an attentive user might notice.	Single-Click: Requires one simple, non-obvious user click (e.g., "Enter Fullscreen").
1	Flawed: Obvious visual defects, incorrect fonts, or missing UI elements.	Complex Interaction: Requires multiple clicks, text entry, or non-standard actions (e.g., copy-paste).

displaying the present origin.

7. Correctly Parse and Display Non-HTTP Schemes: For about: blank, blob:, data:, etc., clearly indicate the scheme and its nature (e.g., "insecure," "local data:", "blob:" URI). Do not allow hash fragments to obscure the scheme itself.
8. Correctly Display RTL Domains: Implement strict parsing and display rules for mixed RTL/LTR domains to prevent visual reordering that hides the true origin.
9. Maintain the "Line of Death": Strictly enforce separation between browser chrome (trusted UI) and web content. Prevent web content from overlaying or interacting with the address bar area, a principle core to Browsers and Applications' security UI design [13].

6.2. User Education

1. Type Sensitive Addresses Manually or Use Trusted Bookmarks: Especially for banking or critical services.
2. Use Developer Tools to Verify Origin: In case of suspicion, console.log (location.origin) reveals the true origin or Inspect site identity via browser padlock icon/certificate details.
3. Keep Browsers Updated: Apply security patches promptly.
4. Be Wary of Unexpected Behavior: If a site looks slightly off or behaves unusually after clicking a link, exercise caution.

7. DISCUSSION AND CONCLUSION

This research has systematically demonstrated that address bar spoofing remains a significant and evolving threat in contemporary web browsers. By developing a comprehensive taxonomy, discovering and disclosing over 70 vulnerabilities, and analyzing their root causes, we have provided a deep insight into the present state of address bar security.

Our analysis of the findings in Table 3 reveals notable trends. Chromium-based browsers that implement heavy UI customizations, such as Opera, Opera GX, and Arc, accounted for a disproportionately high number of vulnerabilities. These often stemmed from unique features, such as custom search engine integrations or non-standard UI state handling, suggesting that deviations from the core Chromium security model introduce new attack surfaces. In contrast, vulnerabilities in WebKit (Safari) were more frequently tied to the nuanced handling of non-standard URI schemes, such as blob:, while Gecko (Firefox) showed susceptibility to URL truncation on mobile displays. This highlights that while engine-level security is mature, browser-specific feature implementations remain a primary source of spoofing vulnerabilities.

The findings underscore a persistent between usability and security. Features designed to create a cleaner UX – such as fast address bar updates, hiding URL components, and novel UI states, such as Picture-in-Picture – are frequently the source of the security flaws we identified. While this research was extensive, it could not cover every browser or operating system, and niche browsers may harbor similar flaws.

In conclusion, maintaining the integrity of the address bar requires continuous, meticulous attention to detail across navigation logic, protocol handling, and UI rendering. Our proposed mitigation strategies offer a multi-layered defense to help vendors fortify this critical UI element. For future work, we recommend two key directions: (1) Expanding testing to include embedded web views and browser extensions, which represent growing attack surfaces, and (2) developing standardized, automated testing tools to systematically probe for these complex UI-based vulnerabilities at scale. The insights from this research aim to significantly advance the understanding and fortification of this crucial element of web security, ultimately working toward a safer online experience.

REFERENCES

[1] *StationX, "Phishing Statistics"*, 2024. Available from: https://www.stationx.net/phishing-statistics [Last accessed on 2025 Jul 10].

[2] "*Trend Micro. Address Bar Spoofing.*" Available from: https://www.trendmicro.com/vinfo/us/security/definition/address-bar-spoofing [Last accessed on 2025 Jul 10].

[3] R. Ismael. "*Address Bar Spoofing. GitHub Repository*," 2025. Available from: https://github.com/renwax23/address_bar_

spoofing [Last accessed on 2025 Jul 10].

[4] *Chromium. "Guidelines for URL Display*." Available from: https://chromium.googlesource.com/chromium/src/+/HEAD/docs/security/url_display_guidelines/url_display_guidelines.md [Last accessed on 2025 Jul 10].

[5] Y. Koster. "*Address Bar Spoofing flaw in Internet Explorer,"* 2004. Available from: https://www.akitasecurity.nl/advisory/AK20040801/address_bar_spoofing_flaw_in_internet_explorer.html [Last accessed on 2025 Jul 10].

[6] L. Treiber. "*Google Chrome HTTPS Address Bar Spoofing*," 2012. Available from: https://blog.acrossecurity.com/2012/01/google-chrome-https-address-bar.html [Last accessed on 2025 Jul 10].

[7] C. Weber. "*Unicode Security Guide - Visual Spoofing*," 2014. Available from: https://cweb.github.io/unicode-security-guide/visual-spoofing

[8] R. Baloch. "*Bypassing Mobile Browser Security for Fun and Profit. In: Presented at BlackHat Asia*," 2016.

[9] *"Chromium. Chromium Trickuri GitHub Repo*," 2019. Available from: https://github.com/chromium/trickuri [Last accessed on 2025 Jul 10].

[10] R. Baloch. "*Multiple Address Bar Spoofing Vulnerabilities in Mobile Browsers,"* 2020. Available from: https://www.rafaybaloch.com/2020/10/multiple-address-bar-spoofing-vulnerabilities.html

[11] *Skylot, "Jadx GitHub Repo*." Available from: https://github.com/skylot/jadx [Last accessed on 2025 Jul 10].

[12] M. Bentkowski. *Address Bar Spoofing in Chrome and Firefox*, 2017. Available from: https://research.securitum.com/address-bar-spoofing-in-chrome-and-firefox-description-of-cve-2017-5089-and-cve-2017-7763 [Last accessed on 2025 Jul 10].

[13] E. Law. "*The Line of Death*," 2017. Available from: https://textslashplain.com/2017/01/14/the-line-of-death [Last accessed on 2025 Jul 10].

[14] A. N. Joinson, U. D. Reips, T. Buchanan and C. B. P. Schofield. "Privacy, trust, and self-disclosure online." *Human-Computer Interaction*, vol. 25, no. 1, pp. 1-24, 2010.

[15] K. Yee. "Aligning security and usability." *IEEE Security and Privacy*, vol. 2, no. 5, pp. 85-88, 2004.

APPENDIX A: SUMMARY OF DISCLOSED VULNERABILITIES

The following table summarizes the 74 unique vulnerabilities identified and disclosed during this research. The full dataset, including PoC codes, image/video PoCs, bounty amount, timeline and full bug description, is available at the repository cited in Ismael [3]. The "Taxonomy Class" refers to the categories defined in Section 4. Note: Not all browsers assign CVEs, which is why some are missing.

ID	Title	Technique	Browser	CVE
1	Bypassing CVE-2019-10875 or Xiaomi's Mint Browser's URL Spoofing patch	Search-Engine	Xiaomi Mint	0
2	0-day Alert: URL Spoofing Bypassed for the latest Mint Browser 1.6.4	Search-Engine	Xiaomi Mint	0
3	Opera Mini Browser Address Bar Spoof	Internal-URI	Opera Mini	0
4	Full Address Bar Spoofing in Opera Touch and Opera GX Mobile	Search-Engine	Opera GX Mobile	0
5	Opera GX For Android URL Spoof/Freeze	URL-Freeze	Opera GX Mobile	0
6	XSS in play.gx.games to RCE, Full Address Bar Spoof in Opera GX	Window-Without-Address-Bar	Opera GX	0
7	Opera for Android Address Bar Spoof	Fullscreen	Opera Android	0
8	Opera Mini for Android Address Bar Spoof	Fullscreen	Opera Mini	0
9	Full Address Bar Spoof and Browser Takeover Inside Opera GX	Window-Without-Address-Bar	Opera GX	0
10	Opera GX Android - Address Bar Spoof with intent://vtp.operagx.gg/?url=	Internal-URI	Opera GX Android	0
11	Opera Mini - Address Bar Spoof with opera-mini://open?url=data:	Internal-URI	Opera Mini	0
12	Iframe Injection in GX Corner to Address Bar Spoofing in Opera GX	Window-Without-Address-Bar	Opera GX	0
13	Opera Mini - Full Address Bar Spoof	End-of-URL-Shown	Opera Mini	0
14	Opera Mac/Windows - New Window Address Bar Spoof using Picture-in-Picture	Window-Without-Address-Bar	Opera	0
15	Brave iOS Address Bar Spoofing Using blob: URI	blob:-URI	Brave	0
16	Brave iOS Address Bar Spoofing Using about: blank URI	about: blank-URI	Brave	0
17	Brave iOS Address Bar Spoof Using 2 RTL (Arabic Characters) Subdomains	RTL-Character	Brave	0
18	Brave iOS Address Bar Spoof Using RTL Domain with RTL+LTR Subdomains	RTL-Character	Brave	0
19	Firefox Desktop Address Bar Spoofing Using Arabic Punycode and a long URL	RTL-Character	Firefox	CVE-2024-11695
20	Edge - Mac - Address Bar Spoofing using Picture-in-Picture	Window-Withour-Address-Bar	Edge Mac	CVE-2024-38093
21	Edge Mobile (Android-iOS) Address Bar Spoof Using Bing Search	Search-Engine	Edge Mobile	CVE-2024-38083
22	Zerion Wallet App For iOS Address Bar and Wallet Confirmation Origin Spoof	Long-Subdomain	Zerion Wallet	0
23	Coinbase Wallet App For iOS Address Bar Origin Spoof	Long-Subdomain	Coinbase Wallet	0
24	Address Bar Spoofing Using about: blank URI	about: blank-URI	Arc Mac	0
25	Address Bar Spoofing using data: URI	data:-URI	Arc Mac	0
26	Address Bar Spoofing using documentPictureInPicture	Window-Without-Address-Bar	Arc Mac	0
27	Address Bar Spoof Using Fullscreen	Fullscreen	Arc Mac	0
28	Address Bar Spoof in Arc Search iOS	Invalid-Scheme	Arc iOS	0
29	Arc Search iOS Full Address Bar Spoof Using Blob: URI	blob:-URI	Arc iOS	0
30	Arc Search iOS Address Bar Spoof Using Search Functionality	Search-Engine	Arc iOS	0
31	Arc Search Android Full Address Bar Spoof Using Custom Port Website Hanging	URL-Freeze	Arc Android	0
32	Arc Searc Android Address Bar Spoof Using Long Subdomain	Long-Subdomain	Arc Android	0
33	Arc Search Android Address Bar Spoof Using Fullscreen	Fullscreen	Arc Android	0
34	Arc Search Android Full Address Bar Spoof by Redirecting to a 204 No Content Page	URL-Freeze	Arc Android	0

ID	Title	Technique	Browser	CVE
35	Arc Browser address bar spoof using view-source:	view-source:-URI	Arc Mac	0
36	Address Bar Spoofing Using blob: URI	blob:-URI	Arc Mac	0
37	Arc Browser address bar spoof using file://URI	file:-URI	Arc Mac	0
38	XSS in arcmobile-internal://using Server-Side Redirect to JavaScript: URI to Full Address Bar Spoof in Arc Search iOS	Internal-XSS	Arc iOS	0
39	Stored XSS in search.arc.net to Address Bar Spoof in Arc iOS and Potential UXSS	Search-Engine	Arc iOS	0
40	Firefox Mobile Address Bar Spoof Using Long Subdomain	Long-Subdomain	Firefox Mobile	0
41	Firefox Mobile iOS Full Address Bar Spoof Using Open in New Tab and JavaScript: URI	JavaScript:-URI	Firefox iOS	CVE-2025-23108
42	Firefox Mobile iOS Full Address Bar Spoof Using Server-Side Redirect to internal://local/errorpage	Internal-URI	Firefox iOS	CVE-2025-27426
43	Firefox Mobile iOS Address Bar Spoof Using Server-Side Redirect to non-http Scheme	Invalid-Scheme	Firefox iOS	CVE-2025-27424
44	Yandex iOS Address Bar Spoofing Using blob: URI	blob:-URI	Yandex iOS	0
45	Yandex Browser iOS Address Bar Spoofing Using about: blank UR	about: blank-URI	Yandex iOS	0
46	Yandex Browser iOS Address Bar Spoof Using 2 RTL (Arabic Characters) Subdomains	RTL-Character	Yandex iOS	0
47	Google App iOS Full Address Bar Spoof	about: blank-URI	Google iOS	0
48	Brave iOS Address Bar Spoof and Potential UXSS Using brave://open-url?url=	Internal-URI	Brave iOS	0
49	OKX App iOS-Android dApp Address Bar Origin Soof	URL-Freeze	OKX	0
50	Coinbase Wallet App For iOS Address Bar Origin Spoof	Long-Subdomain	Coinbase Wallet iOS	0
51	PlayStation 5 Browser Full Address Bar Spoofing	JavaScript:-URI	PlayStation 5 Browser	0
52	Safari New Window Address Bar Spoof With a Subdomain or Long Domain	Long-Subdomain	Safari	0
53	Address Bar Spoof Using Invalid Protocol	Invalid-Scheme	Naver WHALE	0
54	Safari iOS Full Address Bar Spoof Using Tab Hanging	URL-Freeze	Safari	CVE-2025-24128
55	Safari Mac/iOS Full Address Bar Spoof Using Tab Hanging and Invalid URLS	URL-Freeze	Safari	CVE-2025-24128
56	Safari Mac and iOS Address Bar Spoof	about: blank-URI	Safari	CVE-2025-24113
57	Safari Mac/iOS Browser Address Bar Spoof Using Search Engines	Search-Engine	Safari	CVE-2025-30467
58	Safari Mac Address Bar Spoof Using Cursor Overlap	overlap	Safari	
59	Safari Address Bar Spoof Using blob: URI	blob:-URI	Safari	CVE-2025-24113
60	Safari New Window Address Bar Spoof With a Subdomain or Long Domain	Long-Subdomain	Safari	CVE-2025-31266
61	Safari Browser Address Bar Spoof Using Redirect	URL-Freeze	Safari	CVE-2025-43327
62	Firefox Desktop Address Bar Spoof Using Search Query (bypass fix in bug 1970997)	Search-Engine	Firefox	CVE-2025-9183
63	Rabby Wallet iOS Browser Address Bar Origin Spoof to Account Takeovers	Tab-Hanging	Rabby Wallet	0
64	Arc iOS Full Address Bar Spoof Using Timing Attack	URL-Freeze	Arc iOS	0
65	Arc iOS Full Address Bar Spoof Using Search Functionality and Race Condition	Search-Engine	Arc iOS	0
66	MetaMask iOS Address Bar Spoof Using Drag-Drop	URL-Freeze	MetaMask iOS	0
67	Brave iOS Address Bar Spoof Using about: blank	about: blank-URI	Brave iOS	0
68	MetaMask Wallet iOS - Browser Full Address Bar Spoof	Tab-Hanging	MetaMask iOS	0
69	Brave iOS Address Bar Spoof with open-url and data: URI	data:-URI	Brave iOS	0
70	Brave iOS Drag-Drop Address Bar Spoof	URL-Freeze	Brave iOS	0
71	Address Bar Spoof and UXSS Using Shortcuts Chrome iOS	End-of-URL-Shown	Chrome iOS	0
72	Firefox Address Bar Spoof with Long Subdomain and Popup	Long-Subdomain	Firefox	0
73	Edge Mobile (Android-iOS) Full Address Bar Spoof Using Bing and Path Traversal	Search-Engine	Edge Mobile	0
74	Edge iOS Address Bar Spoof Using 2 RTL (Arabic Characters) Subdomains	RTL-Character	Edge iOS	0

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Knowledge and Attitude toward Human Papillomavirus among Students at the University of Sulaimani/City Campus

Khanda Mustafa Ahmed

University of Sulaimani/College of Nursing, Adult Nursing Department, Sulaymaniyah, Iraq

ABSTRACT

Human papillomavirus (HPV) is a sexually transmitted infection, which poses a significant global health concern. HPV affects sexually active individuals worldwide and is the primary cause of cancers in the cervix, anus, vagina, penis, and/or oropharynx. The aim of the study was to assess knowledge and attitude toward HPV among students at the University of Sulaimani City Campus. A cross-sectional study was conducted between December 08, 2024, and February 20, 2025, involving 210 students from the colleges of nursing, pharmacy, education, and Islamic science. Participants were predominantly aged 19–23 years. Female participants (55.7%) outnumbered male participants (44.3%). The most significant representation came from the pharmacy college (34.3%), followed by nursing (28.1%), education (21%), and Islamic science (16.7%). Most participants (86.7%) were in their fourth year of study. Rural residency was reported by 64.8%. Only 29.5% of students demonstrated good knowledge of HPV, and 24.8% showed positive attitudes. While 31% strongly agreed on the importance of vaccination, just 35.7% recognized its role in cancer prevention. Significant associations were found between knowledge levels and age, gender, and father's education ($P < 0.005$). Moreover, there was a significant association between age, gender, father's education, and level of knowledge ($P < 0.005$). Findings reveal substantial gaps in HPV awareness and vaccine attitudes among university students. Targeted educational interventions, particularly within nursing programs, are essential to equip future healthcare professionals with the knowledge needed to promote HPV prevention and reduce cervical cancer risk.

Index Terms: Human Papillomavirus, University Students, HPV Prevention, Vaccination, Knowledge, and Attitude

1. INTRODUCTION

Human papillomavirus (HPV) is a common infection that primarily affects the skin and mucous membranes [1], [2]. Its prevalence varies significantly across regions. Historically, the Middle East and North Africa (MENA) region has reported lower rates of HPV infection, largely attributed to more conservative sexual practices [3]. However, recent findings by Obeid *et al.* [4] indicate a rising trend in HPV infections among women in the region, particularly among individuals aged 10–24 years. Although comprehensive regional data remain limited, approximately 81% of people in the MENA region are considered to have HPV overall. Countries including Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, the United Arab Emirates, and Yemen are among the nations that make up the Gulf Cooperation Council.

Recent studies by Jensen *et al.* and Petca *et al.* [5], [6] confirmed that HPV is a prevalent genital virus. There are more than 100 types of HPV that are transmitted through skin-to-skin

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp346-356

E-ISSN: 2521-4217
P-ISSN: 2521-4209

Copyright © 2025 Ahmed. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Khanda Mustafa Ahmed, University of Sulaimani/College of Nursing, Adult Nursing Department, Sulaymaniyah, Iraq. E-mail: khanda.mustafa@univsul.edu.iq

Received: 21-09-2025 Accepted: 29-10-2025 Published: 01-12-2025

contact, including sexual contact and other genital contact, such as touching an infected person's genitalia.

However, HPV cannot be transmitted through contact with inanimate objects, such as toilet seats. The Centers for Disease Control and Prevention [7] confirmed that the majority of HPV infections may resolve spontaneously; however, they can occasionally lead to genital warts, skin lesions, and malignancies, including cervical cancer.

A persistent HPV infection can lead to the development of abnormal cells on the cervix, which, if left untreated, may progress to cervical cancer [8]. While cervical cancer is the most common cancer caused by HPV, both men and women can also develop fewer common cancers, including anal, vulvar, vaginal, oral/throat, and penile cancers, as a result of a persistent infection [9]. The appearance of warts varies depending on the specific type of HPV infection [8]. A genital wart may present as a stem-like protrusion, a flat lesion, or small, cauliflower-shaped growths. Furthermore, multiple viruses or serotypes are associated with HPV, with high-risk HPV serotypes being strongly linked to cervical cancer [10]. High-risk serotypes, such as HPV 16 and HPV 18, are frequently detected and have the potential to cause cancer, as reported by Kombe *et al.* [11].

Furthermore, most sexually active men and women can become infected with HPV at some time in their lives, and some people may get infected again. Most people acquire the virus early in life, shortly after becoming sexually active [1], [12].

Cervical cancer is the fourth most common cancer in women and the fourth major cause of cancer-related mortality, according to Sung *et al.* and Bongaerts *et al.* [13], [14]. It is regarded as the third most prevalent malignancy in women under the age of 45 in almost 79% of nations. The estimated global incidence of cervical cancer in 2018 was 13.1 cases/10,000 women, with significant variations between countries, as reported by Arbyn *et al.* [15].

The World Health Organization (WHO) recommends HPV vaccination for girls aged 9–14 as a primary preventive measure against cervical cancer [16]. To achieve global elimination, the WHO aims to reduce incidence to fewer than 4 cases/100,000 women [16] through a life-course strategy encompassing primary, secondary, and tertiary prevention. Secondary interventions focus on screening women aged 30, while tertiary measures include palliative care and cancer treatment [17]. By 2030, the goal is for 90% of girls under 15 to be vaccinated, 70% of women aged 35–45 to be screened, and 90% of those with precancerous lesions to receive treatment [18].

Assessing knowledge and attitudes (KAs) about HPV among university students is crucial for several reasons. First, one of the main causes of cervical cancer and other conditions is HPV. Understanding students' knowledge can help identify gaps and guide the design of educational campaigns to promote awareness about HPV and its prevention. Second, attitudes toward HPV vaccination significantly influence its acceptance. By assessing these attitudes, health authorities can address misconceptions and encourage vaccination, which is essential in reducing HPV-related diseases.

Moreover, university students are forming lifelong health behaviors at a crucial moment. Assessing their attitudes and knowledge about HPV supports larger public health campaigns meant to lower the prevalence of diseases linked to HPV. As university students are at an age where they make independent health decisions, they can significantly benefit from education about HPV. This education encourages healthier lifestyle choices, promotes regular health check-ups, and empowers them to make informed decisions about their well-being, including vaccination, safe sex practices, and regular screenings.

This research provides a unique contribution to the literature by examining university students' attitudes toward HPV within the context of Sulaimani City, offering culturally specific insights that are currently underrepresented in global research. The central hypothesis is that students' attitudes toward HPV are influenced by their level of knowledge, cultural beliefs, and exposure to health education, which in turn affects their willingness to engage in preventive behaviors such as vaccination and screening.

These assessments not only help identify gaps in awareness but also inform the development of targeted educational programs and culturally sensitive materials focused on HPV prevention, transmission, and management. By understanding students' perceptions, the study aims to reduce stigma and misconceptions, ultimately fostering a more supportive environment for individuals affected by HPV.

2. MATERIALS AND METHODS

2.1. Study Design and Setting

This cross-sectional study was conducted between December 08, 2024, and February 20, 2025, involving 210 students from

the colleges of nursing, pharmacy, education, and Islamic science. A convenient sampling technique was used. Students in their last years of study (4th and 5th years) who expressed an interest to take part in the study were included in the inclusion criteria. As final-stage university students were intentionally selected for this study due to their advanced academic standing and increased exposure to health-related content. By the 4th or 5th year of study, students, particularly those enrolled in health disciplines, are more likely to have encountered formal education on infectious diseases, vaccination, and public health principles. Their maturity and critical thinking skills enhance the reliability of self-reported data, especially in studies assessing KAs. Moreover, as these students transition into professional roles, their perceptions and behaviors regarding HPV prevention carry practical implications for future healthcare delivery and peer education. This cohort's inclusion improves the significance of the results and assists in the creation of focused educational programs targeted to raise young adults' knowledge of HPV and encourage their vaccination uptake. On the other hand, 1st, 2nd, and 3rd-year students were not included.

The sample size was determined based on practical feasibility and methodological considerations. In the absence of prior local data on HPV-related KAs among university students in Sulaimani, a conservative estimate of 50% prevalence was assumed to maximize sample size requirements. Using a 95% confidence level and a 5% margin of error, the ideal sample size was calculated to be approximately 384 participants. However, due to logistical constraints and limited access to multiple institutions, the study was conducted at a single university, and a final sample of 210 students was recruited using convenience sampling. While this number falls below the statistically ideal threshold, it provides valuable preliminary insights and serves as a foundation for future multicenter studies with broader representation.

2.2. Ethical Approval

The University of Suleiman's College of Nursing's Ethical Committee provided its approval to the study. (Meeting number 24 on April 21, 2025). All study participants gave verbal consent. Furthermore, the confidentiality of all personal information was ensured throughout the study to provide the participants' anonymity.

2.3. Data Collection

Participants were randomly selected, and questionnaires were distributed to 20–30 students in the final stage of each of the aforementioned colleges. Before questionnaire administration, participants were verbally informed of the study's objectives and provided with assurances regarding the confidentiality and anonymity of their responses. A self-administered, anonymous questionnaire was provided to each participant, requiring approximately 15 min to complete.

Following a comprehensive examination of the literature, a questionnaire was developed for the study. It consisted of three parts – Part I: Includes information, including details such as age, gender, college, year of study, marital status, and the parents' educational attainment; Part II: Includes a knowledge assessment about HPV, containing thirteen questions evaluating participants' understanding of HPV infection, the HPV vaccine, and transmission; and Part III: Includes six statements measuring participants' attitudes regarding HPV.

2.4. Reliability

To evaluate the instrument's reliability, a pilot study was conducted involving a sample of 15 students. These participants were excluded from the main study to mitigate potential bias. A test-retest procedure was implemented over a 2-week interval, and the resulting scores were analyzed to calculate a reliability coefficient. Cronbach's alpha was employed to assess internal consistency, yielding a value of 0.84 for the full instrument, indicating excellent reliability. Data entry was performed on the same day as collection, allowing for immediate quality control and verification of completeness.

2.5. Pilot Study

A small trial study was conducted in the mentioned colleges at the University of Sulaimani. It took place from November 20 to November 27, 2024, involving 15 students as the sample for the pilot study. These students were excluded from the main study sample.

2.6. Validity

To ensure content validity, the questionnaire was reviewed by a panel of six experts representing diverse fields relevant to the study. Each expert received a preliminary version of the instrument and was asked to evaluate its clarity, relevance, and adequacy in capturing the intended constructs. The majority consensus affirmed that the questionnaire was appropriately designed and effectively structured to assess the phenomenon under investigation.

2.7. Statistical Analysis

The data preparation process involved data entry, coding of questions and variables, data cleaning to identify and correct illogical values, and data computation. The analysis

was conducted using the Statistical Package for the Social Sciences (SPSS) version 22. Descriptive statistics were used to analyze numerical data, providing useful summaries and facilitating the identification of average values, including the mean, median, and standard deviation. Frequency distribution represents categorical data. A $P < 0.05$ was considered statistically significant. Sociodemographic characteristics were presented as frequencies and percentages. Inferential statistical tests, including one-way analysis of variance and t-tests, were used where appropriate to compare the means of numeric variables. Upon identifying significant differences, students' KA scores were categorized into three groups according to their percentage of the maximum possible score: Poor (0–50%), fair (51–75%), and good (76–100%).

The level of HPV knowledge was assessed using a set of closed-ended and multiple-choice questions. Nine of these questions had "Yes" (3 points), "No" (2 points), and "Do not know" (1 point) as response options. The remaining questions were multiple-choice, allowing participants to select more than one answer from a list of provided options.

Attitudes toward HPV were assessed using a 5-point Likert scale: Strongly Agree (5 points), Agree (4 points), Neutral (3 points), Disagree (2 points), and Strongly Disagree (1 point). The attitude section comprised 6 statements, resulting in a total possible score ranging from 6 to 30 per respondent. To facilitate interpretation, total scores were converted into a standardized scale ranging from 1 to 5, where a score closer to 5 indicated a more positive attitude toward HPV. Based on the standardized score, participants' attitudes were classified as positive (≥4), neutral (3–3.9), or negative.

3. RESULTS

The study sample comprised university students with a mean age of 22.1 years (SD = 1.72), reflecting a relatively homogenous cohort typical of senior undergraduate populations. The majority (85.2%) were aged between 19 and 23 years, suggesting that the findings predominantly represent early adulthood, a developmental stage characterized by increased autonomy and evolving health-related decision-making. A smaller proportion (14.8%) was aged 24–28, indicating limited representation from older students who may have different life experiences or health priorities. Gender distribution showed a slightly higher proportion of female participants (55.7%) compared to males (44.3%). This may reflect enrollment trends in health-related disciplines and could be relevant given the gendered nature of HPV-related health risks and screening behaviors.

Participants were drawn from four academic colleges, with the largest representation from pharmacy (34.3%), followed by nursing (28.1%), education (21%), and Islamic science (16.7%). The predominance of students from health-related fields (pharmacy and nursing) may enhance the relevance of their responses, as these students are more likely to have received formal education on infectious diseases and preventive health measures. A substantial majority (86.7%) were in their 4th year of study, suggesting a cohort with advanced academic exposure and potential familiarity with public health concepts. This may contribute to more informed perspectives on HPV and related interventions.

Residency data indicated that 64.8% of participants lived in rural areas. This variable is particularly salient, as rural residency may be associated with reduced access to healthcare services and health education, potentially influencing KAs toward HPV prevention and vaccination. Regarding marital status, most participants were single (78.1%), while 21.9% were married. Marital status may intersect with sexual health behaviors and perceptions, which are relevant to HPV-related research, particularly in culturally sensitive contexts.

Finally, parental education levels were relatively high, with 71% of mothers and 82.9% of fathers reported as educated. Parental education is often linked to greater health literacy and may indirectly shape participants' attitudes and knowledge through familial influence and early exposure to health-promoting behaviors (Table 1).

The results indicate that over half of the participants (51.4%) were unaware of HPV. In addition, there is a lack of understanding about HPV, with the majority of participants (94.3%) unsure of what it is or incorrectly associating it with bacterial infections. The most common source of information regarding HPV among participants was health professionals (23.8%), followed by the Internet (18.6%), friends (13.8%), and television (4.8%). Regarding HPV transmission, the majority of participants (30.5%) reported that it can be transmitted through sexual contact. In comparison, 28.6% believed that it could be transmitted through skin contact, while 12.4% thought that it could be transmitted through sharing personal items.

Over half of the participants (52.4%) were unaware that HPV is a causative agent of cancer. However, 60% recognized its link to oral cancer, while 26% identified its association

TABLE 1: Distribution of participants according to sociodemographic characteristics

Variables	Frequency (F)	Percentage
Age groups		
19–23 years	179	85.2
24–28 years	31	14.8
Mean±SD 22.1±1.72 Median 22		
Gender		
Female	117	55.7
Male	93	44.3
College		
Nursing	59	28.1
Pharmacy	72	34.3
Education	44	21
Islamic science	35	16.7
Years of study		
4th years	182	86.7
5th years	28	13.3
Residency		
Urban	74	35.2
Rural	136	64.8
Marital status		
Single	164	78.1
Married	46	21.9
Mother education		
Educated	149	71
Non educated	61	29
Father education		
Educated	174	82.9
Non educated	36	17.1
Total	210	100

with anal cancer. In terms of vaccination awareness, 34.3% of respondents correctly indicated that the HPV vaccine is recommended for females and can be administered prior to the age of 12. However, only 29.5% knew that the vaccine can be administered after the age of 18 years old.

With respect to preventive knowledge, there is a lack of awareness about prevention methods. For example, 86.7% of participants demonstrated awareness of the Pap smear test as a preventive screening tool, and 83.3% knew that using condoms can help prevent HPV. Interestingly, the proportion of participants who believed that HPV infection could be asymptomatic was evenly split at 50%. More than two-thirds (68.6%) knew that HPV can affect both genders. Finally, 64.8% of participants expressed interest in receiving more information about HPV, indicating a strong demand for educational resources (Table 2).

The current survey shows that 37.1% of participants strongly agree that receiving an HPV vaccination is necessary, indicating a more positive viewpoint regarding HPV vaccination. Nonetheless, a notable proportion of participants (36.2%) expressed neutrality regarding the recommendation of the

TABLE 2: General knowledge toward human papillomavirus of the study sample

Knowledge about human papillomavirus	Yes F (%)	No F (%)
1-Have you heard of Human Papillomavirus (HPV)?	102 (48.6)	108 (51.4)
2-What is HPV?		
Virus that causes warts	69 (32.9)	141 (67.1)
A virus that can cause certain cancers	37 (17.6)	173 (82.4)
A bacterial infection	12 (5.7)	198 (94.3)
Not sure	98 (46.7)	112 (53.3)
3-Where did you first hear about HPV? (Select all that apply)		
Social media	37 (17.6)	173 (82.4)
Internet	39 (18.6)	171 (81.4)
Television	10 (4.8)	200 (95.2)
Health professionals	50 (23.8)	160 (76.2)
Friends or family	29 (13.8)	181 (86.2)
Others (please specify)	49 (23.3)	161 (76.7)
4-How is HPV transmitted? (Select all that apply)		
Skin-to-skin contact	60 (28.6)	150 (71.4)
Sexual contact	64 (30.5)	146 (69.5)
Sharing personal items	26 (12.4)	184 (87.6)
Airborne transmission	20 (9.5)	190 (90.5)
Not sure	83 (39.5)	127 (60.5)
5-Can HPV cause cancer?	110 (52.4)	100 (47.6)
6-Is there a vaccine available for HPV?	133 (63.3)	77 (36.7)
7-Who should get the HPV vaccine? (Select all that apply)		
Female only	72 (34.3)	138 (65.7)
Male only	9 (4.3)	201 (95.7)
Both	62 (29.5)	148 (70.5)
Do not know	51 (24.3)	159 (75.7)
8-At what age is the HPV vaccine recommended?		
Before age 12	72 (34.3)	138 (65.7)
Between ages 12 and 18	9 (4.3)	201 (95.7)
After age 18	62 (29.5)	148 (70.5)
Do not know	51 (24.3)	159 (75.7)
9-Which health conditions do you associate with HPV? (Select all that apply)		
Cervical cancer	47 (22.4)	163 (77.6)
Genital warts	60 (28.6)	150 (71.4)
Oral cancer	21 (10.0)	189 (90)
Anal cancer	26 (12.4)	184 (87.6)
None	13 (6.2)	197 (93.8)
Not sure	99 (47.1)	111 (52.9)
10-Do you think HPV infection can be asymptomatic?	105 (50.0)	105 (50.0)
11-Are you aware that HPV can affect both men and women?	144 (68.6)	66 (31.4)
12-Do you know how HPV can be prevented?		
Vaccine	72 (34.3)	138 (65.7)
Pap-smear test	28 (13.3)	182 (86.7)
Using condom	35 (16.7)	175 (83.3)
Do not know	92 (43.8)	118 (56.2)
13-Would you be interested in getting more information on HPV?	136 (64.8)	74 (35.2)
Total	210	100

HPV: Human papillomavirus

HPV vaccine to friends and family. Regarding attitudes toward vaccination, 35.7% of participants believe that

getting vaccinated against HPV can prevent certain types of cancer, while a significant number of participants (37.6%) remain neutral. Nearly one-quarter (39.5%) of participants agree or strongly agree that it is crucial for students to be educated about HPV and its potential health risks, reflecting strong support for HPV education. In addition, 45.7% of participants believe that increased awareness of HPV could help reduce its spread. Furthermore, 32.4% of participants agree that discussing HPV openly among students might help reduce the stigma associated with sexually transmitted diseases (STIs), showing a positive attitude toward open communication and education (Table 3).

The findings reveal that less than half of the participants (45.7%) possessed minimal knowledge about HPV, while only 29.5% possess good knowledge. Furthermore, more than a third of the participants (36.7%) exhibit a poor attitude toward HPV, and 24.8% demonstrate a good attitude. In terms of encouraging positive attitudes toward HPV prevention and education, it shows that although attitudes are slightly better than knowledge levels, considerably more can be performed (Table 4).

The findings from Table 5 indicate that there is an association between certain sociodemographic. Factors such as age, gender, years of study, marital status, and father's education were statistically significant in relation to knowledge about HPV among the study participants ($P < 0.005$). Nevertheless, no statistically significant association between mother's education, residency, and college was found.

Table 6 reveals statistically significant associations between participants' attitudes toward HPV and the variables of age, gender, college affiliation, years of study, and father's educational level ($P < 0.005$). In contrast, no significant associations were observed with residency status, marital status, or mother's education.

4. DISCUSSION

The purpose of the current study was to evaluate university students' attitudes and knowledge regarding HPV.

These findings are consistent with those reported by Khan *et al.* [19] in Pakistan, where the majority of respondents were female (74.9%) and single (79%). The demographic parallels with Khan *et al.* [19] suggest that young, unmarried women across South Asia may share similar vulnerabilities and openness to HPV prevention. This regional consistency underscores the urgency of culturally sensitive, gender-focused interventions; this population should be a key target for public health efforts. Tailored awareness campaigns and vaccination outreach initiatives are essential to effectively promote HPV prevention within this group. In addition, while pharmacy students comprised the largest academic group in both studies (46.2% in Khan *et al.*'s study; 34.3% in the present study), the limited representation from non-health disciplines such as arts and humanities underscores a gap in HPV-related education across the broader university population. These findings highlight the need for interdisciplinary educational interventions that extend beyond health faculties, integrating HPV awareness into general curricula and student wellness programs. Such efforts could enhance vaccine uptake, reduce stigma, and promote informed decision-making among diverse student populations.

TABLE 3: Level of Attitude toward human papillomavirus among study participants

Attitude toward human papillomavirus	Answer					
	Strongly agree F (%)	Agree F (%)	Neutral F (%)	Disagree F (%)	Strongly disagree F (%)	Mean
1-Do you think that it is important to get vaccinated against HPV?	65 (31)	78 (37.1)	50 (23.8)	8 (3.8)	9 (4.3)	3.86
2-Would you recommend the HPV vaccine to your friends and family?	67 (31.9)	33 (15.7)	76 (36.2)	19 (9)	15 (7.1)	3.56
3-Do you believe that getting vaccinated against HPV can prevent certain types of cancer?	75 (35.7)	22 (10.5)	79 (37.6)	23 (11)	11 (5.2)	3.60
4-Do you think students need to be educated about HPV and its potential health risks?	83 (39.5)	83 (39.5)	30 (14.3)	8 (3.8)	6 (2.9)	4.09
5-Do you believe that increased awareness of HPV could help reduce its spread?	76 (36.2)	96 (45.7)	25 (11.9)	5 (2.4)	8 (3.8)	4.08
6-Do you think that discussing HPV openly among students might help reduce stigma associated with STIs?	66 (31.4)	68 (32.4)	62 (29.5)	10 (4.8)	4 (1.9)	3.86

Scores MD=24.0, Mean±SD 23.0±3.55, HPV: Human papillomavirus

TABLE 4: Distribution of knowledge and attitude scores among study participants

Variables *n*=(210)	Scores	Frequency (F)	Percentage
Knowledge human papillomavirus	Poor	96	45.7
	Fair	52	24.8
	Good	62	29.5
Attitude to human papillomavirus	Poor	77	36.7
	Fair	81	38.6
	Good	52	24.8

TABLE 5: Association between knowledge of human papillomavirus and sociodemographic characteristics among study participants

Sociodemographic characteristics	*n*	Knowledge of human papillomavirus scores	Statistical analysis	
		Mean SD	t/F	*P*-value
Age groups				
19–23 years	179	7.89±2.45	−11.131	0.000
24–28 years	31	7.96±1.51		H. Sig
Gender				
Female	117	7.76±2.11	−6.036	0.000
Male	93	8.08±1.89		H. Sig
College				
Nursing	59	8.11±2.00	4.217	0.912
Pharmacy	72	7.98±2.16		N. Sig
Education	44	7.96±2.14		
Islamic science	35	7.31±1.51		
Years of study				
4th year	182	7.92±2.03	−11.076	0.005
5th year	28	7.75±1.97		Sig
Residency				
Urban	74	7.78±1.85	−2.848	0.375
Rural	136	7.97±2.11		N. Sig
Marital status				
Single	164	8.04±2.12	−9.228	0.000
Married	46	7.41±1.51		H. Sig
Mother education				
Educated	149	8.04±2.08	−7.954	0.125
Non-educated	61	7.57±1.82		N. Sig
Father education				
Educated	174	8.03±2.11	−9.959	0.000
Non-educated	36	7.27±1.34		H. Sig

Another study by Liu *et al.* [20] conducted in Beijing, China, reported that 56.3% of participants were female and 43% were male. Similarly, Shetty *et al.* [21] in South India found that 72.3% of participants were female, while 27.7% were male. In Shetty's study, 50.8% of participants were between the ages of 20 and 22. The largest proportion of participants were enrolled in medical programs (43.5%), followed by those studying dentistry (27. 9%), nursing students (21.1%), and medical laboratory students (7.5%). Regarding marital status, 91% of the participants were single. Most participants were 1st-year students (32.2%). More than half of the students (55.8%) were from urban areas. In addition, the percentage

TABLE 6: Association between attitude of human papillomavirus and sociodemographic characteristics among study participants

Sociodemographic characteristics	*n*	Attitude of human papillomavirus scores	Statistical analysis	
		Mean±SD	t/F	*P*-value
Age groups				
19–23 years	179	22.9±3.57	−12.844	0.000
24–28 years	31	23.8±3.40		H. Sig
Gender				
Female	117	23.1±3.23	−6.885	0.000
Male	93	22.9±3.92		H. Sig
College				
Nursing	59	23.0±3.10	4.269	0.000
Pharmacy	72	23.0±3.89		H. Sig
Education	44	23.2±3.77		
Islamic science	35	22.8±3.37		
Years of study				
4th year	182	22.8±3.64	−13.549	0.029
5th year	28	24.7±2.35		Sig
Residency				
Urban	74	22.7±3.73	−3.822	0.335
Rural	136	23.2±3.45		N. Sig
Marital status				
Single	164	23.3±3.36	−10.365	0.067
Married	46	22.5±4.16		N. Sig
Mother education				
Educated	149	23.0±3.46	−9.817	0.304
Non-educated	61	23.1±3.80		N. Sig
Father education				
Educated	174	23.1±3.33	−11.705	0.044
Non-educated	36	22.5±4.50		Sig

of fathers' professions in various fields was higher compared to mothers' professions, which included healthcare providers, self-employed individuals, and retirees.

The findings of this study were consistent with those of a previous study conducted by Khan *et al.* [19], which also examined KAs toward HPV among university students in Pakistan. The study confirmed that 57% of students reported having heard about HPV. The participants' knowledge of HPV infection was 43.6%, whereas their knowledge of the HPV vaccine was 25.3%. Furthermore, almost the same percentage of students said that HPV infects both men and women, and 55% said that HPV causes cervical cancer. According to the vast majority of students (71%), HPV is not a rare disease in Pakistan. Nearly 68% of students, however, disagreed that HPV can cause no symptoms. About 47% of respondents said that HPV causes anal and penile malignancies, while 52% claimed that it causes external genital warts.

Regarding the mechanism of transmission, 40.51% of respondents answered that they were unsure, 34.10% stated

that body fluid exchange is the most common way, followed by skin-to-skin contact (44.87%), and coughing or sneezing (5.64%). Less than half (44.87%) of students said that vaccination is the best way to prevent HPV, followed by sexual abstinence (30%), condom use (21.54%), and antibiotic use (5.38%). Unlike prior studies conducted in South and East Asia, this research provides the first comparative insight into HPV knowledge among Iraqi university students. The persistence of misconceptions and low vaccine awareness highlights a critical gap in sexual health education, suggesting the need for national curriculum reform and targeted outreach.

Moreover, reports from previous studies by Khan *et al.*, [19]; Shetty *et al.*, [21]; Patel *et al.*, [22]; Khatiwada *et al.*, [23]; Mekonnen *et al.*, [24] confirmed that the sources of information were varied, including the Internet (37.8%), teachers (31.7%), television (22.1%), healthcare providers (21.9%), brochures (16.7%), posters (14.2%), family (9.4%), social media (67.3%), college education (48.0%), families or friends (41.7%), doctors or health-related workers (37.3%), friends (9.6%), radio (6%), media (49.12%), healthcare professionals (30.97%), and primary care physicians (2.9%). These findings suggest that social media campaigns are effective tools that hold contemporary relevance, indicating that social media campaigns may serve as effective tools for enhancing public knowledge about HPV. In addition, traditional media continues to have a significant impact, and healthcare providers continue to be reliable sources of information. This research highlights how important it is for medical professionals to be informed and initiative-taking in teaching people about HPV. Finally, family and community-based interventions can be very effective in raising awareness about HPV and fostering positive attitudes.

Regarding knowledge on HPV transmission, Chanprasertpinyo *et al.* [25] confirmed that approximately 65% of participants knew that HPV spreads through sexual contact and causes infection in both men and women. In addition, 52.4% of participants recognized that early sexual activity increases the risk of HPV infection, and 52.8% were aware of its role in causing cervical cancer. However, only 49.7% understood that condom use can help prevent HPV transmission. Notably, over half of the respondents (57%) were unaware that the HPV vaccine offers protection for both males and females In addition, Shetty *et al.* and Varer *et al.* [21], [26] found that a small percentage of participants (24.4%) stated that a woman who had previously been infected could receive the HPV vaccine; 21.8% were aware that the HPV vaccine could not cure a current infection. Nonetheless, 64.4% of respondents said that both males and females should have the vaccination, and 61% knew that cervical cancer screening was required for individuals who had already had the vaccination.

Likewise, Kim *et al.* [27] conducted a study in China assessing knowledge, attitudes, and perceptions associated with HPV vaccination among female Korean and Chinese university students. They found that 21.1% of Chinese female college students and 25.4% of Korean female students had heard of HPV. 15.8% of Korean students thought the immunization was unnecessary in the absence of infection signs, while 17.2% said that it should only be given to females. 52.7% of Chinese students said that HPV vaccinations should be given to both men and women, 31% answered that they were not required if there were no signs of infection, and 16.3% answered that only women should have them.

Furthermore, a study conducted by Shetty *et al.* [21] in South India revealed that most students (78%) knew the sexual route of HPV transmission, and 25.8% believed that HPV infection could be asymptomatic. While 62.6% of students demonstrated awareness that cervical cancer results from persistent infection with high-risk HPV strains and acknowledged that HPV can affect males, only 37.2% recognized its link to oropharyngeal cancer. In addition, 49.5% of participants reported that condom use offers protection against HPV transmission.

Moreover, Khatiwada *et al.* [23] in Indonesia confirmed that 90% of students reported that HPV is the leading cause of cervical cancer, and 86.3% agreed that sexual transmission is crucial for the spread of HPV infections among individuals. The participants' knowledge of the HPV vaccines was slightly greater than two-thirds. Regarding the administration of the HPV vaccine, 35% of students aged 15–20 years had received it. In contrast, 26% of students believed that 9–14 years was the optimal age range for receiving the vaccine, while 28% considered 21–26 years to be ideal.

Similarly, Cheema *et al.* [28] claimed that the majority (67.6%) of students knew that HPV can infect both males and females, and 58.3% stated that HPV can cause health effects in both men and women. Over half of the participants (53.8%) recognized that HPV is primarily transmitted through sexual intercourse, and 54% acknowledged that practicing safe sex – such as using condoms – can reduce the risk of transmission. In contrast, only 12.6% were aware that HPV infections often resolve spontaneously without medical intervention, and just 11.5% correctly identified the recommended vaccination age as between 11 and 12 years.

A significant portion of participants believed that HPV could be transmitted through skin contact or sharing items, which mirrors misconceptions reported in studies from India and China. However, the persistence of these myths in our Iraqi sample may be influenced by limited sexual health education and cultural sensitivities surrounding (STIs). Iraq lacks coordinated national efforts to address HPV knowledge gaps. This suggests a need for culturally tailored educational interventions that respect local norms while promoting accurate information.

Similarly, a study by Khan *et al.* [19] revealed that approximately 53% of students stated that there is no vaccine against HPV, and 64% disagreed that a vaccine can prevent HPV. In addition, Khatiwada *et al.* [22] reported that over 70% of participants perceived the HPV vaccine as both safe and highly effective. However, approximately 21% believed that vaccination is unnecessary for individuals who are otherwise healthy.

Furthermore, Cheema *et al.* [28] reported that 71.5% of students would accept HPV vaccination if recommended by a physician, while Patel *et al.* [22] found that 90% of participants believed that vaccines were safe and expressed willingness to receive and promote them. Despite this general receptiveness, vaccine hesitancy remains pronounced among nursing students, with 20% expressing safety concerns and fewer than 3% reporting prior vaccination. This discrepancy suggests that professional exposure to healthcare information does not uniformly translate into vaccine confidence, highlighting the need for targeted educational interventions within nursing curricula.

Regarding peer and family influence, Di Giuseppe *et al.* [29] and Cheema *et al.* [28] noted that 54.5% of students desired more information before recommending the vaccine to others. This underscores a broader issue: Negative perceptions often stem from misinformation, particularly through social media, and limited understanding of HPV risks and vaccine benefits. Despite robust evidence supporting vaccine safety, concerns about side effects persist, exacerbated by cultural stigma surrounding sexually transmitted infections. Myths such as claim that the vaccine causes infertility or promotes early sexual activity further deter uptake, especially in conservative societies

Significantly, because HPV is spread through sexual contact, conversations about the virus and its vaccine are sometimes frowned upon in some cultures, which causes stigma and discourages people from being vaccinated. Myths and misinformation such as claim that the vaccine causes infertility or encourages early sexual behavior. Apart from safety concerns, other reasons that contribute to vaccine hesitation include cultural beliefs, misinformation, and poor access to reliable health education.

In terms of KA, Chowdhury *et al.* [30] found that 56.71% of participants had poor HPV knowledge, while Abdelaliem *et al.* [31] reported moderate attitudes among 57% of nursing students. Cheema *et al.* [27] similarly observed low knowledge levels, with only 22.4% demonstrating good understanding. Voidazan *et al.* [32] revealed that 6th-year medical students in Romania had significantly higher HPV knowledge (62%) compared to 1st-year students (55.1% with minimal understanding), suggesting that academic progression correlates with awareness. These findings emphasize the importance of integrating HPV education early in university programs to address knowledge gaps and foster informed attitudes.

Sociodemographic factors also influence attitudes. Abdelaliem *et al.* [30] confirmed a significant association with marital status, while academic year showed a negative correlation with nursing students' attitudes toward vaccination. In contrast, age and marital status were positively linked to favorable perceptions. These trends suggest that maturity and life experience may enhance receptivity to preventive health measures.

This study has several limitations. First, its findings may not be generalizable beyond the sampled university in Sulaimani city. Future research should adopt multicenter designs to capture broader educational contexts. Second, the cross-sectional nature of the study limits causal inference; longitudinal or experimental approaches are needed to establish temporal relationships. Finally, due to the absence of a standardized global tool for assessing HPV KAs, the questionnaire was adapted from previously validated instruments, which may affect comparability across studies.

5. CONCLUSIONS

The results suggest a gap in KAs regarding HPV among university students. As future healthcare professionals, it is essential for students to possess a foundational understanding of HPV and its vaccine to effectively educate the public and promote preventive health behaviors. Therefore, it is crucial to develop community-based initiatives with a focus on integrating HPV education into school and university

health curricula, organizing peer-led workshops, and engaging local healthcare professionals to culturally deliver sensitive and evidence based information. In addition, media campaigns should be strategically designed to correct prevalent misconceptions, particularly the myth that HPV is transmitted through airborne routes or casual contact. Messaging should clearly communicate that since sexual contact is the main way that HPV spreads, it is important to highlight the vaccine's role in cancer prevention and foster open dialog to reduce stigma associated with sexually transmitted infections.

Moreover, practical outreach efforts could include setting up information booths in public spaces staffed by trained healthcare professionals, distributing educational brochures, and leveraging digital platforms to reach diverse audiences. These interventions should be inclusive and gender sensitive, ensuring that individuals of all backgrounds are equipped with accurate knowledge to make informed decisions about HPV prevention and risk management.

6. ACKNOWLEDGMENT

Gratitude is extended to all of the study participants for their participation and special thanks are given for their great assistance and collaboration in making the data collection process easier.

7. CONFLICTS OF INTEREST

The author affirms that they have no conflicts of interest.

REFERENCES

[1] World Health Organization. "*Human Papillomavirus and Cervical Cancer*". Available from: https://www.who.int/news-room/fact-sheets/detail/human-papilloma-virus-and-cancer [Last accessed on 2025 Nov 19].

[2] J. Wolf, L. F. Kist, S. B. Pereira, M. A. Quessada, H. Petek, A. Pille, J. G. Maccari, M. P. Mutlaq and L. A. Nasi. "Human papillomavirus infection: Epidemiology, biology, host interactions, cancer development, prevention, and therapeutics". *Reviews in Medical Virology*, vol. 34, no. 3, p. e2537, 2024.

[3] H. B. Baghi, B. Yousefi, M. A. Oskouee and M. Aghazadeh. "HPV vaccinations: A Middle Eastern and North African dilemma". *Lancet Infectious Dieases*, vol. 17, no.1, pp. 18-19, 2017.

[4] D. A. Obeid, S. A. Almatrrouk, M. B. Alfageeh, M. N. A. Al-Ahdal, F. S. Alhamlan. "Human papillomavirus epidemiology in populations with normal or abnormal cervical cytology or cervical cancer in the Middle East and North Africa: A systematic review and meta-analysis". *Journal of Infection and Public Health*, vol. 13, no. 9, pp. 1304-1313, 2020.

[5] J. E. Jensen, G. L. Becker, J. B. Jackson and M. B. Rysavy. "Human papillomavirus and associated cancers: A review". *Viruses*, vol. 16, no. 5, p. 680, 2024.

[6] A. Petca, A. Borislavschi, M. E. Zvanca, R. C. Petca, F. Sandru and M. C. Dumitrascu. "Non-sexual HPV transmission and role of vaccination for a better future". *Experimantal and Theraputic Medicine*, v*ol.* 20, no. 6, p. 1, 2020.

[7] Centers for Disease Control and Prevention. "*Genital HPV Infection: CDC Fact Sheet*". Atlanta GA, 2017. Available from: https://www.cdc.gov/std/hpv/stdfact-hpv.html [Last accessed on 2025 Nov 19].

[8] P. V. Chin-Hong, G. E. Reid and AST Infectious Diseases Community of Practice. "Human papillomavirus infection in solid organ transplant recipients: Guidelines from the American Society of transplantation infectious diseases community of practice". *Clinical Transplantation*, vol. 33, p. e13590, 2019.

[9] K. Tanaka, G. Kogure, M. Onuki, K. Matsumoto, T. Iwata, D. Aoki and I. Kukimoto. "Ancient evolutionary history of human papillomavirus type 16, 18 and 58 variants prevalent exclusively in Japan". *Viruses*, vol. 14, no. 3, p. 464, 2022.

[10] Y. Lu, J. Xue, T. Deng, X. Zhou, K. Yu, L. Deng, M. Huang, X. Yi, M. Liang, Y. Wang and H. Shen. "Safety and feasibility of CRISPR-edited T cells in patients with refractory non-small-cell lung cancer". *Nature Medicine*, vol. 26, no. 5, pp. 732-740, 2020.

[11] A. J. K. Kombe, B. Li, A. Zahid, H. M, Mengist, G. A Bounda, Y. Zhou and T. Jin. "Epidemiology and burden of human papillomavirus and related diseases, molecular pathogenesis, and vaccine evaluation". *Frontiers in Public Health*, vol. 8, p. 552028, 2021.

[12] A. Ardekani, M. Sepidarkish, A. Mollalo, P. Afradiasbagharani, S. Rouholamin, M. Rezaeinejad, M. Farid Mojtahedi, S. Mahjour, M. Almukhtar, M. Nourollahpour Shiadeh and A. Rostami. "Worldwide prevalence of human papillomavirus among pregnant women: A systematic review and meta-analysis". *Reviews in Medical Virology*, vol. 33, no. 1, p. e2374, 2023.

[13] H. Sung, J. Ferlay, R. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal and F. Bray. "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries". *A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209-249, 2021.

[14] O. Bongaerts, C. Clevers, M. Debets, D. Paffen, L. Senden, K. Rijks, L. Ruiten, D. Sie-Go, P. J. Van Diest and M. Nap. "Conventional microscopical versus digital whole-slide imaging-based diagnosis of thin-layer cervical specimens: A validation study". *Journal of Pathology Informatics*, vol. 9, no. 1, p. 29, 2018.

[15] M. Arbyn, E. Weiderpass, L. Bruni, S. De Sanjosé, M. Saraiya, J. Ferlay and F. Bray. "Estimates of incidence and mortality of cervical cancer in 2018: A worldwide analysis". *The Lancet Global Health*, vol. 8, no. 2, pp. e191-e203, 2020.

[16] World Health Organization. "*Vaccines and Immunization: What is Vaccination*?" Geneva, 2021. Available from: https://www.who.int/news-room/questions-and-answers/item/vaccines-and-immunization-what-is-vaccination [Last accessed on 2025 Nov 19].

[17] World Health Organization. "*Global Strategy to Accelerate the Elimination of Cervical Cancer as a Public Health Problem*", 2020. Available from: https://www.who.int/publications/i/item/9789240014107 [Last accessed on 2025 Nov 19].

[18] Z. Q. Toh, F. Zhao, L. Zhang, L. Wei and H. Zou. "HPV natural history, immunological responses and vaccination strategies:

Challenges and opportunities". *Frontiers in Immunology*, vol. 14, p. 1349664, 2024.

[19] T. M. Khan, M. A. Buksh, I. U. Rehman and A. Saleem. "Knowledge, attitudes, and perception towards human papillomavirus among university students in Pakistan". *Papillomavirus Research*, vol. 2, pp. 122-127, 2016.

[20] Y. Liu, N. Di and X. Tao. "Knowledge, practice and attitude towards HPV vaccination among college students in Beijing, China". *Human Vaccines and Immunotherapeutics*, vol. 16, no. 1, pp. 116-123, 2020.

[21] S. Shetty, S. Prabhu, V. Shetty and A. K. Shetty. "Knowledge, attitudes and factors associated with acceptability of human papillomavirus vaccination among undergraduate medical, dental and nursing students in South India". *Human Vaccines and Immunotherapeutics*, vol. 15, no. 7-8, pp. 1656-1665, 2019.

[22] N. H. Patel, H. S. Parikh, M. R. Jasrai, P. J. Mewada and N. Raithatha. "The study of the prevalence of knowledge and vaccination status of HPV vaccine among healthcare students at a tertiary healthcare center in Western India". *The Journal of Obstetrics and Gynecology of India*, vol. 74, no. 4, pp. 326-333, 2024.

[23] M. Khatiwada, C. Kartasasmita, H. S. Mediani, C. Delprat, G. Van Hal and C. Dochez. "Knowledge, attitude and acceptability of the human papilloma virus vaccine and vaccination among university students in Indonesia". *Frontiers in Public Health*, vol. 9, p. 616456, 2021.

[24] B. A. Mekonnen, Y. K. Anagaw, B. A. Kassahun and M. C. Worku. "Evaluation of female university students knowledge, attitudes, and practices toward human papillomavirus infection and vaccination. Multicenter cross-sectional study". *BMC Womens Health*, vol. 24, no. 1, p. 437, 2024.

[25] W. Chanprasertpinyo and C. Rerkswattavorn. "Human papillomavirus (HPV) vaccine status and knowledge of students at a university in rural Thailand". *Heliyon*, 6, no. 8, p. e04625, 2020.

[26] C. Varer Akpinar and S. Alanya Tosun. "Knowledge and perceptions regarding human papillomavirus (HPV) and willingness to receive HPV vaccination among university students in a north-eastern city in Turkey". *BMC Womens Health*, vol. 23, no. 1, p. 299, 2023.

[27] H. W. Kim, E. J. Lee, Y. J. Lee, S. Y. Kim, Y. J. Jin, Y. Kim and J. L. Lee. "Knowledge, attitudes, and perceptions associated with HPV vaccination among female Korean and Chinese university students". *BMC Women's Health*, vol. 22, no. 1, p. 51, 2022.

[28] S. Cheema, A. Abraham, P. Maisonneuve, A. Jithesh, K. Chaabna, R. Al Janahi, S. Sarker, A. Hussain, S. Rao, A.B. Lowenfels and R. Mamtani. "HPV infection and vaccination: A cross-sectional study of knowledge, perception, and attitude to vaccine uptake among university students in Qatar". *BMC Public Health*, vol. 24, no. 1, p. 2316, 2024.

[29] G. Di Giuseppe, S. Angelillo, A. Bianco, F. Gallè, F. Licata, G. Liguori, F. Napolitano, C. G. A Nobile, M. Pavia, C. P. Pelullo and I. F. Angelillo. "Evaluating knowledge, attitudes, and behaviors toward HPV infection and vaccination among university students in Italy". *Vaccines (Basel)*, vol. 11, no. 10, p. 1517, 2023.

[30] S. Chowdhury, R. Ara, S. Roy, S. M. S. Tanvir, F. N. Eva, T. M. Neela, A. A. Moonmoon, S. Sifat, M. Zamila and M. D. H. Hawlader. "Knowledge, attitude, and practices regarding human papillomavirus and its' vaccination among the young medical professionals and students of Bangladesh". *Clinical and Experimental Vaccine Research*, vol. 11, no. 1, p. 63, 2022.

[31] S. M. F. Abdelaliem, A. M. Kuaia, A. A. Hadadi, A. K. Alhujayri, A. A. Al Anazi, A. A Hajar, R. S. AlShareda and S. M. Amri. "Knowledge and attitudes toward human papillomavirus and vaccination: A survey among nursing students in Saudi Arabia". *Healthcare*, vol. 11, no. 12, p. 1766, 2023.

[32] S. Voidazan, S. H. Morariu, M. Tarcea, H. Moldovan and M. Dobreanu. "Human papilloma virus (HPV) infection and HPV vaccination: Assessing the level of knowledge among students of the University of medicine and pharmacy of Tirgu Mures, Romania". *Acta Dermatovenerologica Croatica*, vol. 24, no. 3, p. 193, 2016.

ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Forecasting of the Infant Mortality Rate in Iraq

Awayi Ghazy Abdulkareem[1], Lana Abdul Hamed Muhamed Nury[2], Soran Husen Mohamad[3]

[1]Department of Pediatric Nursing, College of Nursing, University of Sulaimani, City of Sulaimani, Iraq, [2]Department of Maternal Neonate Nursing, College of Nursing, University of Sulaimani, City of Sulaimani, Iraq, [3]Department of Statistics and Informatics, College of Administration and Economic, University of Sulaimani, City of Sulaimani, Iraq.

ABSTRACT

This study investigates applying the GM(1,1) model to forecast the infant mortality rate (IMR) in Iraq from 2025 to 2034, utilizing historical data spanning 2015–2024. The findings indicate a consistent decline in IMRs during the analyzed period, reflecting effective public health interventions. The model's parameters were estimated using the Ordinary Least Squares method, revealing an intercept of 0.0278 and a slope of 26.7693. The forecasting accuracy of the GM(1,1) model was exceptional, demonstrated by a Mean Absolute Percentage Error of only 0.2869% and a precision rate of 99.7131%, categorizing the forecasts as highly accurate. Projected IMRs show a continued decline, decreasing from 20.55 deaths per 1000 live births in 2025 to approximately 16.00 by 2034. These results underscore the utility of the GM(1,1) model in providing reliable forecasts to inform health policy and intervention strategies aimed at improving maternal and child health in Iraq.

Index Terms: Forecasting, Infant, Mortality Rate, Iraq

1. INTRODUCTION

The infant mortality rate (IMR) is a crucial indicator of a nation's health status and reflects the effectiveness of health systems and social conditions. In Iraq, the IMR has experienced fluctuations due to various factors, including healthcare access, conflict, and socioeconomic conditions. Historically, Iraq has faced significant challenges in improving health outcomes, particularly following decades of war and instability. World Bank [1] reported that while there have been improvements in healthcare services, disparities remain between urban and rural areas, impacting the overall IMR. Understanding and forecasting IMR is essential for effective health policy planning and intervention strategies.

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp357-362

E-ISSN: 2521-4217
P-ISSN: 2521-4209

Copyright © 2025 Abdulkareem, *et al*. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Advancements in statistical methods and data analytics have provided new opportunities for forecasting IMR in Iraq. Time series analysis has become popular, allowing researchers to examine historical trends and predict future rates. A study done in 2022 utilized time series data spanning several decades to demonstrate how statistical models can effectively forecast IMR trends in Iraq. This analysis highlighted that while there has been a general decline in IMR, periods of instability have caused significant fluctuations, indicating the need for adaptive health policies that respond to changing circumstances [2].

Integrating machine learning techniques offers an even more refined approach to forecasting IMR. Recent research has shown that artificial intelligence and machine learning models, such as back propagation neural networks, can enhance predictive accuracy beyond traditional statistical methods [3]. Also investigated the use of machine learning to predict IMR in the Kurdistan Region of Iraq, finding promising results that could aid policymakers in anticipating healthcare needs and resource allocation effectively. These advanced forecasting methods provide valuable insights influencing public health strategies to reduce infant mortality [2].

Corresponding author's e-mail: Lana Abdul Hamed Muhamed Nury, Department of Maternal Neonate Nursing ,College of Nursing, University of Sulaimani, City of Sulaimani, Iraq. E-mail: lana.muhamednury@univsul.edu.iq

Received: 24-03-2025 Accepted: 14-08-2025 Published: 01-12-2025

Furthermore, addressing the factors contributing to IMR is vital for improving overall child health in Iraq. Critical determinants include maternal education, access to healthcare, and nutritional status. Therefore, efforts to reduce IMR should encompass multidimensional strategies that address these underlying factors. Wang *et al.* [2] emphasize the importance of comprehensive maternal and child healthcare programs prioritizing education and access to services. By forecasting IMR based on these determinants, stakeholders can develop targeted interventions that facilitate improved health outcomes for mothers and infants.

Enhancing forecasting models has significant implications for public health planning in Iraq. Accurate forecasts of IMR can guide resource allocation, inform interventions, and ultimately contribute to achieving global health targets, such as the Sustainable Development Goals. Addressing IMR contributes to individual health outcomes and plays an essential role in national development, emphasizing health as a critical component of socio-economic growth [4]. Therefore, continuous efforts to refine forecasting techniques and implement evidence-based policies are crucial for advancing maternal and child health in Iraq.

2. MATERIALS AND SUBJECTS

2.1. Study Design

This study employed a quantitative descriptive forecasting design using secondary data to predict future trends in Iraq's IMR. The study design was selected to enable the application of time-series forecasting models, specifically the GM(1,1) model from Grey System Theory, to evaluate and project patterns based on historical data.

2.2. Sample of the Study

This study investigates applying the GM(1,1) model to forecast the infant mortality rate in Iraq from 2025 to 2034, utilizing historical data spanning 2015 to 2024 by using secondary data to predict future trends in Iraq's infant mortality rate (IMR) (appendix).

2.3. Setting of the Study

The setting of this study is virtual and data-driven, relying on publicly accessible online databases. The IMR data for Iraq from 2015 to 2024 was sourced from Macrotrends.net, a reputable database compiling global health and development indicators. Forecasting was conducted in a simulated environment using statistical software capable of applying Grey System models.

2.4. Method of Data Collection

Grey System Theory, introduced by Deng in 1989, addresses systems lacking complete information, such as structure, operation mechanisms, and behaviour documentation, termed Grey Systems. Examples include the human body, agriculture, and the economy. Grey System Theory aims to bridge the social and natural sciences gap, fostering interdisciplinary collaboration across various fields. It has demonstrated enduring relevance since its inception, particularly in China, where it is widely known and applied across sectors such as agriculture, ecology, economics, meteorology, medicine, geography, and industry [15].

In Grey System Theory, the GM (m, n) model, where m represents the order of the difference formula and n indicates the number of factors, is employed to predict future system outputs with relatively high accuracy, even without a complete mathematical model of the actual system. The GM (1, 1) model is particularly prominent due to its computational efficiency, with researchers often focusing on it for predictions. This model utilises first-order differential equations to match data generated by the accumulation generating operation (AGO).

Grey System Theory encompasses five main groups: gray generating, gray relational analysis, gray forecasting, gray decision-making, and gray control. Among these, gray forecasting plays a significant role in addressing uncertainty and information insufficiency within systems. The general form of the gray forecasting model is GM (n, m), where n and m denote the order of ordinary differential equations and the number of gray variables. Various forecasting models are proposed based on different ordinary differential equations and the number of gray variables utilized. The advantage of the GM(1,1) model lies in its computational efficiency, offering accurate predictions despite minimal data requirements (Li *et al.*, 2015). This study presents the concepts of the GM(1,1) model. More specifics are given as follows:

2.5. Gray Model GM(1,1)

The commonly used gray prediction model is often the GM(1,1), indicating the utilization of a single variable within the model (Liu *et al.*, 2015). The calculation process of GM(1,1) involves six steps (Ahmed *et al.*, 2023).

2.6. Evaluate Precision of Forecasting Models

2.6.1. Mean absolute percentage error (MAPE)

Several statistical tests and measurements assess the accuracy and performance of the proposed model, among them MAPE. To evaluate the reliability and performance of the

TABLE 1: Categorizing the grade of predicting accuracy.

Grade level	Highly accurate	Good	Reasonable	Inaccurate
MAPE (%)	<10	10–20	20–50	>50

forecasting technique in this study (Ahmed *et al*., 2023), the forecasting accuracy level can be classified into four grades based on the MAPE of each model, as indicated in Table 1:

A lower MAPE indicates higher precision in the forecasting model. Typically, an MAPE below 10% signifies an accurate model, while an MAPE between 10% and 20% denotes a good model with acceptable accuracy.

2.6.2. Precision rate (p)

Precision Rate, which measures the level of the closeness of the statement of forecast quantity and the actual value, p is defined as follows:

Table 2 shows a higher precision rate, which indicates greater precision in the forecasting model. Typically, a precision rate >99% signifies an accurate model, while a precision rate between 98.0% and 95.0% suggests a good model with acceptable accuracy.

2.7. Ethical Considerations

This study used secondary data that are publicly available and are not contain any personal or identifiable information. Therefore, ethical approval was not required. However, appropriate attribution to the data source (Macrotrends.net) has been made to ensure transparency and integrity. All data were handled in accordance with ethical research standards concerning the use of open-access information.

3. RESULT OF THE STUDY

In this study, time series data regarding the IMR in Iraq, spanning the years 2015 to 2024, were analyzed. This study evaluates the GM(1,1) model for forecasting Iraq's Infant Mortality from 2025 to 2034 using data spanning from 2015 to 2024.

The IMR from 2015 to 2024 is displayed in Table 3, which demonstrates a steady decrease over time. In 2015, the rate was 26.737 fatalities per 1000 live births; by 2024, it had dropped to 20.661. This pattern shows that within the past 10 years, maternity care, healthcare services, and child survival circumstances have improved.

TABLE 2: Categorizing the grade of predicting accuracy.

Precision rank	Highly accurate	Good	Reasonable	Inaccurate
Precision rate (p) (%)	p≥99.0	p≥95.0	p≥90.0	p≤90.0

TABLE 3: The infant mortality rate

Years	Infant mortality rate	Years	Infant mortality rate
2015	26.737	2020	22.918
2016	25.86	2021	22.324
2017	24.982	2022	21.731
2018	24.105	2023	21.137
2019	23.511	2024	20.661

Deaths per 1000 live births

3.1. Estimated Model Forecasting of GM(1,1)

Using the Ordinary Least Squares (OLS) method, the parameters of the GM(1,1) model have been estimated as follows: the intercept (*a*) is 0.027813794, and the slope (*b*) is 26.7692722.

Table 4 represents the actual value, forecasted value, and percentage error (PE%) of the GM(1,1) model. It is seen that the MAPE value for the GM(1,1) is (0.2869%), and it can be said that the efficiency and accuracy values obtained are (P = 100−MAPE(%)=100−0.2869% = 99.7131%). This means that the forecasting value is considered a highly accurate prediction because the value of the precision rate in $P \geq 99.0\%$. The forecasted values for 2025–2035 are presented in Table 5.

The projected IMR s using the GM(1,1) model for 2025–2034 are shown in Table 5. According to the data, the number of deaths per 1000 live births decreased steadily from 20.55 in 2025 to 15.99 in 2034. If present trends continue, this prediction points to further advancements in newborn health outcomes.

4. DISCUSSION

The analysis of the IMR in Iraq from 2015 to 2024 presents significant insights into the utility of the GM(1,1) model in forecasting future trends. The data indicate a consistent decline in IMRs over the observed period, highlighting effective public health interventions and evolving healthcare access in Iraq. This trend is crucial, as it reflects improvements in maternal and infant health services and aligns with global objectives to reduce mortality rates among vulnerable populations [5]. The application of the GM(1,1) model,

TABLE 4: Predicted value and rate error value of the GM (1,1) model

Year	No. k	Real value	Forecasted value	Percentage error (PE%)
2015	k=0	26.737	26.737	0.0000
2016	k=1	25.86	25.66701172	0.7463
2017	k=2	24.982	24.96295143	0.0762
2018	k=3	24.105	24.2782039	0.7185
2019	k=4	23.511	23.61223937	0.4306
2020	k=5	22.918	22.96454261	0.2031
2021	k=6	22.324	22.33461254	0.0475
2022	k=7	21.731	21.7219618	0.0416
2023	k=8	21.137	21.1261164	0.0515
2024	k=9	20.661	20.54661537	0.5536
Results			MAPE (%)	0.2869%
	Precision Rate (p)		99.7131%	

TABLE 5: The forecasted value by GM (1,1) model

Years	Infant mortality rate	Years	Infant mortality rate
2025	20.54661537	2030	17.87900743
2026	19.98301037	2031	17.38857639
2027	19.43486537	2032	16.91159815
2028	18.90175629	2033	16.44770368
2029	18.38327069	2034	15.99653409

Deaths per 1000 live births

noted for its computational efficiency, has proven effective in predicting future values despite the inherent complexities and uncertainties associated with health data [6].

The estimated parameters from the GM(1,1) model–an intercept (a) of 0.0278 and a slope (b) of 26.7693–afford a robust framework for understanding the dynamics influencing infant mortality in Iraq. Such parameters reinforce the model's capability to accurately reflect historical data patterns, ensuring that forecasting for 2025–2034 captures expected trends for continued decline. This is mirrored in similar studies, which have demonstrated that time series models, particularly those grounded in Grey System Theory, can provide significant predictive power even in fields characterized by limited data [7].

The results demonstrate a highly accurate forecasting performance with a MAPE of 0.2869%, suggesting exceptional precision in the GM(1,1) model's predictions. This performance aligns with established thresholds for model accuracy, indicating that forecasts developed through this approach are reliable and strategically beneficial for policy formulation. A precision rate of 99.7131% indicates a model that can significantly guide health policy decisions, targeting interventions effectively to reduce the IMR further [11], [12].

Moreover, the forecasted values for 2025–2034 reveal a trend toward a continual reduction in IMRs, with estimates showing values consistently declining from 20.55 deaths/1000 live births in 2025 to approximately 16.00 by 2034. This outlook is aligned with the global health agenda, which emphasizes the importance of sustained efforts to improve healthcare systems and children's outcomes [13]. The substantial decrease projected by the GM(1,1) model underscores the potential impacts of ongoing health initiatives and resource allocation strategies to improve maternal and child health, particularly in developing regions [14].

Applying the GM(1,1) model provides valuable insights into predicting infant mortality trends in Iraq, backed by a strong statistical foundation and high forecasting accuracy. Continued research in this area could explore further enhancements to the model by incorporating additional variables, such as socioeconomic factors and environmental influences, to enrich the predictive capabilities and policy implications [15]. The findings of this study reinforce the significance of effective health prediction models as tools for enhancing public health strategies and advancing the broader objectives of global health initiatives.

5. CONCLUSION

This study demonstrated the effectiveness of the GM(1,1) model from Grey System Theory in forecasting Iraq's IMR using data from 2015 to 2024. The model showed high accuracy, with an MAPE of 0.2869% and a precision rate of 99.7131%, projecting a continued decline in IMR from 2025 to 2034. These results highlight the model's value in supporting health policy and planning. The findings underscore the need for continued investment in maternal and child health, and future research could improve the model by integrating socioeconomic and environmental factors.

6. DISCLOSURE

The authors report no conflicts of interest in this work.

REFERENCES

[1] World Bank. "*Mortality Rate, Infant (Per 1,000 Live Births) - Iraq*," 2023. Available from: https://data.worldbank.org [Last accessed on 2025 Mar 24].

[2] H. Wang, X. Li and J. Li. "Predicting infant mortality using machine learning algorithms: A comparative study". Journal of Biomedical Informatics, vol. 108, pp. 103501, 2020.

[3] S. O. Salih. "Utilising backpropagation networks for predicting infant mortality rates in Northern Iraq." *International Journal of Forecasting*, vol. 38, no. 2, pp. 123-135, 2022.

[4] World Health Organization. *"Maternal and Child Health: A Global Challenge*." 2021. Available from: https://www.who.int [Last accessed on 2025 Mar 24].

[5] S. P. Nyoni. "Forecasting trends in infant mortality rates in Iraq using time series analysis." *Journal of Health Policy and Management*, vol. 29, no. 4, pp. 147-156, 2022.

[6] World Health Organization. "*Global Health Observatory Data Repository*." 2021. Available from: https://www.who.int/data/gho [Last accessed on 2025 Mar 24].

[7] J. L. Deng. "Grey system theory: New theory and practice." *Journal of Grey System*, vol. 1, no. 1, pp. 1-24, 1988.

[8] X. Wang, Y. Xiao and Q. Xu. "Application of grey prediction models in health and epidemiology: A literature review." *Health Informatics Journal*, vol. 25, no. 1, pp. 26-36, 2019.

[9] R. Zhang, T. Liu, and J. Wang. "Evaluating forecast accuracy of models: A focus on infant health." *International Journal of Public Health*, vol. 62, no. 3, pp. 433-444, 2017.

[10] United Nations. "*Sustainable Development Goals*." 2020. Available from: https://sdgs.un.org/goals [Last accessed on 2025 Mar 24].

[11] P. Li, T. Wang, and L. Zhao. "Integrating socio-economic indicators with forecasting models in health studies." *Statistical Modelling in Healthcare*, vol. 5, no. 3, pp. 34-49, 2021.

[12] D. H. Ahmed, S. H. Mohamad and R. H. R. Karim. Using single exponential smoothing model and grey model to forecast corn production in Iraq during the period (2022-2030)." *University of Kirkuk Journal For Administrative and Economic Science*, vol. 13, no. 3, pp. 261-271, 2023.

[13] G. D. Li, S. Masuda and M. Nagai. "Predictor design using an improved grey model in control systems." *International Journal of Computer Integrated Manufacturing*, vol. 28, no. 3, pp. 297-306, 2015.

[14] J. Liu, S. Liu and Z. Fang. "Fractional-order reverse accumulation generation GM(1,1) model and its applications." *Journal of Grey System*, vol. 27, no. 4, pp. 52-62, 2015.

[15] J. Deng. "Introduction to grey system theory." *The Journal of Grey System*, vol. 1, no. 1, pp. 1-24, 1989.

APPENDIX

1. GREY MODEL GM(1,1)

The GM(1,1), which uses a single variable in the model, is frequently the most widely used gray prediction model (Liu *et al*., 2015). There are six steps in the GM(1,1) computation procedure (Ahmed *et al*., 2023), which are as follows:

Step 1: The non-negative original sequence data is given by $X^{(0)}$

$$X^{(0)} = \{x^{(0)}(1), x^{(0)}(2),\ldots,x^{(0)}(n)\}, n \geq 4 \quad (1)$$

Step 2: Using the non-negative original sequence data $X^{(0)}$, $X^{(1)}$ is constructed by a one-time accumulated generating operation (1-AGO), represented as

$$X^{(1)} = \{x^{(1)}(1), x^{(1)}(2),\ldots,x^{(1)}(n)\} \quad (2)$$

Where:

$$x^{(1)}(1) = x^{(0)} \quad (1)$$

$$X^{(1)}(k) = \sum_{j=1}^{k} x^{(0)}(k = 1,2,\ldots,n)$$

Step 3: Determine the background value z through mean generating operation (MGO).

$$z^{(1)}(k) = 0.5\, x^{(1)}(k+1) + 0.5 x^{(1)}(k)\ k = 23 \quad (3)$$

Step 4: The result of 1-AGO produces a continuously rising sequence, resembling the solution curve of a first-order linear differential equation. Thus, the solution curve of the following differential equation approximates the 1-AGO data:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \quad (4)$$

The parameters a and b are called the development coefficient and gray input, respectively. $x^{(1)}(1) = x^{(0)}(1)$ represents the corresponding initial condition

Step 5: The parameters a and b can be calculated using the OLS method

$$\begin{pmatrix} a \\ b \end{pmatrix} = (\beta^T \beta)^{-1} \beta^T Y_n \quad (5)$$

Where β and Y_n are defined as follows:

$$\beta = \begin{bmatrix} x^{(1)}(2) & 1 \\ x^{(1)}(3) & 1 \\ \vdots & \ldots \\ x^{(1)}(n) & 1 \end{bmatrix} \text{ and } Y_n = [x^{(1)}(2), x^{(1)}(3),\ldots,x^{(1)}(n)]^T,$$

Solving Eq. (5) along with the initial condition yields the solution.

$$\hat{x}^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{b}{a}\right] e^{-ak} + \frac{b}{a}\ k = 1,2,3.... \quad (6)$$

Step 6: Utilizing the inverse accumulated generating operation (I-AGO) on $\hat{x}^{(1)}(k)$, the predicted data of $\hat{x}^{(0)}(k)$ Can be estimated as follows:

$$\hat{x}^{(0)}(k+1) = (1 - e^{-a})\left[x^{(0)}(1) - \frac{b}{a}\right] e^{-ak} \quad (7) \text{ or}$$

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \quad (8)$$

2. EVALUATE PRECISION OF FORECASTING MODELS

2. 1. MAPE

The correctness and performance of the suggested model are evaluated using a number of statistical tests and metrics, including Mean Absolute Percentage Error (MAPE). The MAPE index was used to assess the forecasting technique's performance and dependability in this study (Ahmed *et al*., 2023). It has the following definition.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} |PE_k| * 100\% =$$

$$\frac{1}{n}\sum_{i=1}^{n} \left|\frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)}\right| * 100\% \quad (9)$$

2. 2. Precision Rate (p)

Precision Rate, which measures the level of the closeness of the statement of forecast quantity and the actual value, p is defined as follows:

$$P = 100 - \text{MAPE}(\%) \quad (10)$$

p-ISSN 2521-4209
e-ISSN 2521-4217

جامعة التنمية البشرية
UNIVERSITY OF HUMAN DEVELOPMENT

UHD Journal of Science and Technology

A Scientific periodical issued by University of Human Developement

Vol.9 No.(2) December 2025

2025 2725

e.mail:jst@uhd.edu.iq