

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



Ismael Abdulkareem Ali¹, Sozan Abdulla Mahmood²

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, funduscopy, 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

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

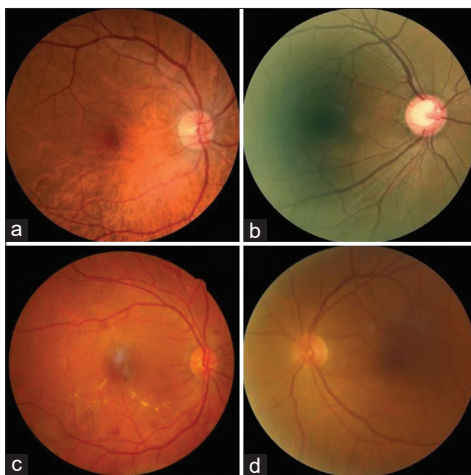


Fig. 1. Retinal fundus images in various conditions sourced from the eye disease classification dataset [45]: (a) normal; (b) glaucoma; (c) diabetic retinopathy; (d) cataract.

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 Albelaïhi 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

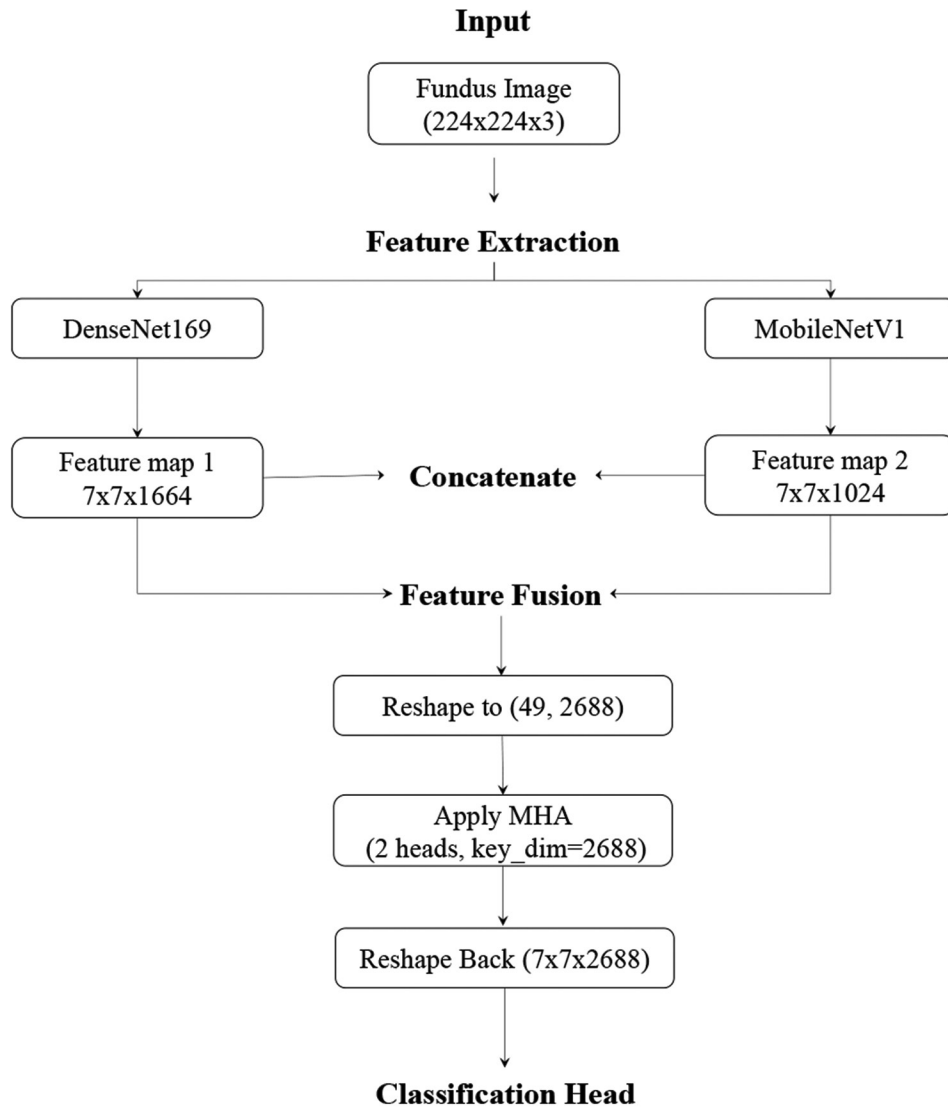


Fig. 2. Implementation of the feature fusion and attention mechanism in the proposed hybrid model.

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

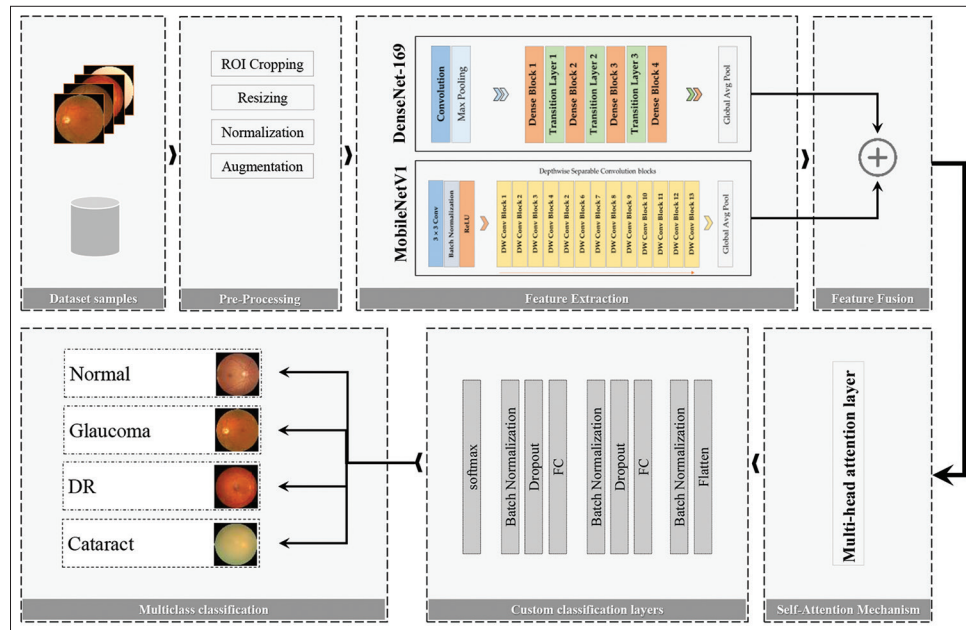


Fig. 3. Overall proposed system.

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 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 $(49,$

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

Augmentation	Feature/range
Horizontal flip	100%
Vertical flip	100%
Rotation	$0^\circ, 90^\circ, 180^\circ, \text{ or } 270^\circ$
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	$[0,1]$

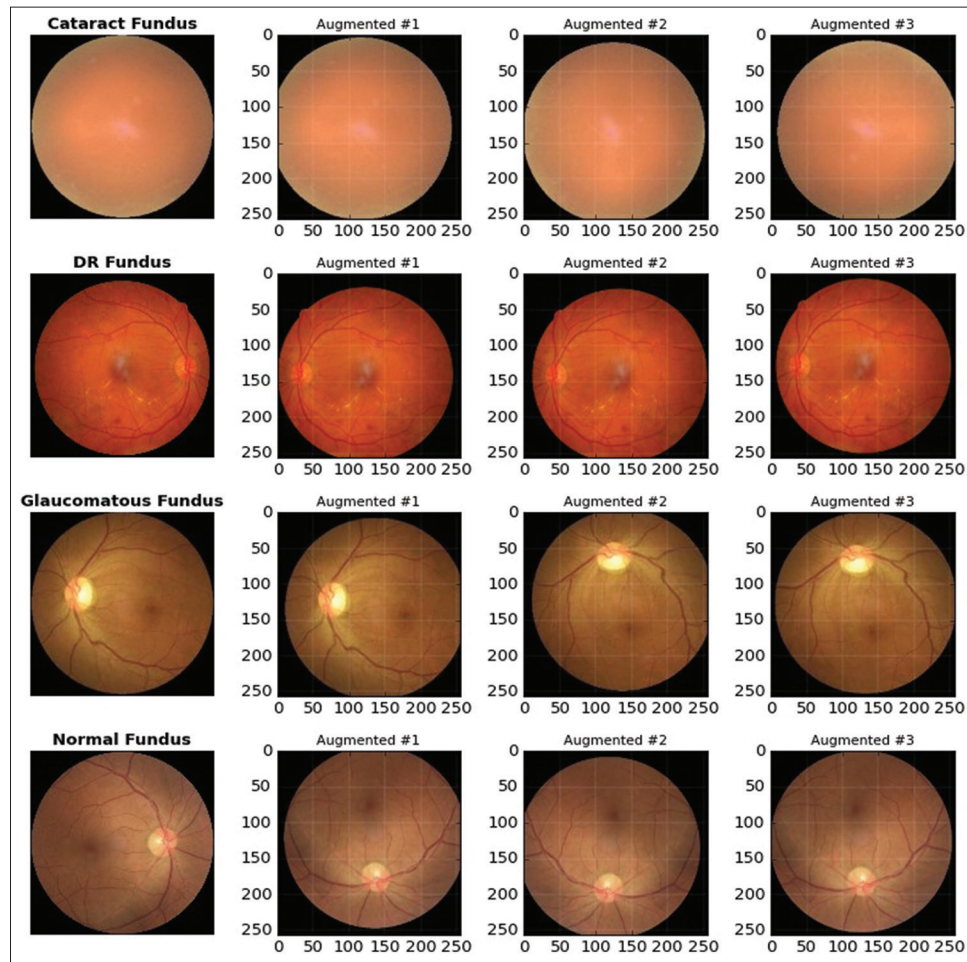


Fig. 4. Retinal fundus images with various augmented versions generated through random transformations.

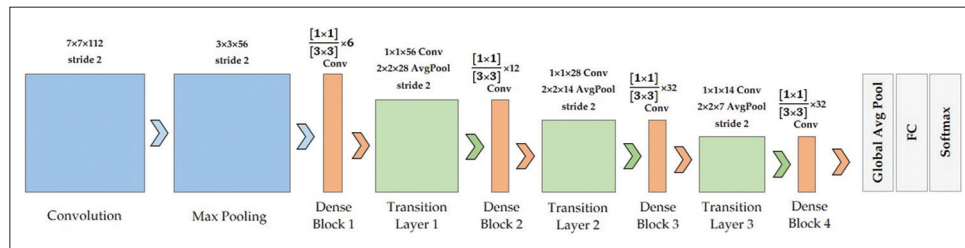


Fig. 5. DenseNet-169 architecture including layers.

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.

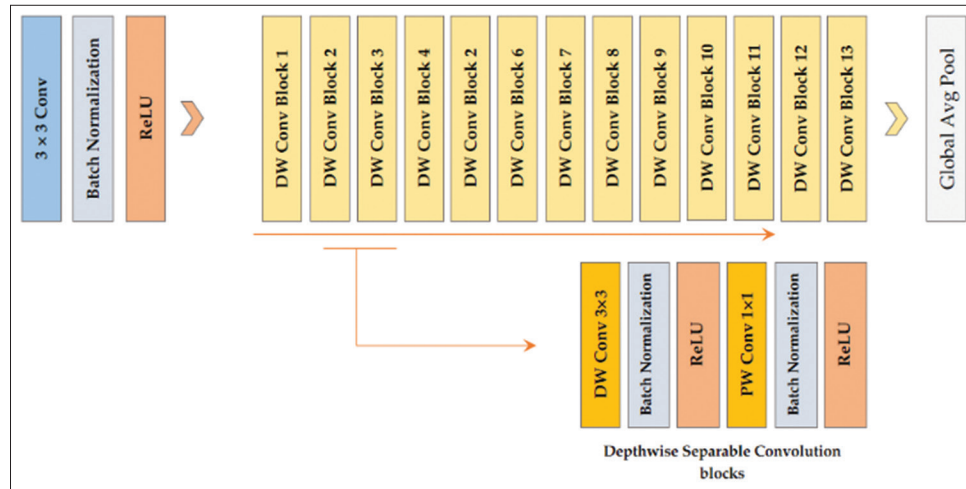


Fig. 6. MobileNetV1 architecture.

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	112×112	3×3 × 32 DW, stride 1
Dense Block 1	56×56	[1×1, 3×3]×6 (Conv), Dropout 0.2		112×112	1×1 × 32×64, stride 1
Transition Layer 1	56×56	1×1 Conv	Block 2	56×56	3×3 × 64 DW, stride 2
	28×28	2×2 AvgPool, stride 2		56×56	1×1 × 64×128, stride 1
Dense Block 2	28×28	[1×1, 3×3]×12 (Conv), Dropout 0.2	Block 3	56×56	3×3 × 128 DW, stride 1
Transition Layer 2	28×28	1×1 Conv2×2 AvgPool, stride 2		56×56	1×1 × 128×128, stride 1
	14×14		Block 4	28×28	3×3 × 128 DW, stride 2
Dense Block 3	14×14	[1×1, 3×3]×32 (Conv), Dropout 0.2		28×28	1×1 × 128×256, stride 1
Transition Layer 3	14×14	1×1 Conv	Block 5	28×28	3×3 × 256 DW, stride 1
	7×7	2×2 AvgPool, stride 2		28×28	1×1 × 256×256, stride 1
Dense Block 4	7×7	[1×1, 3×3]×32 (Conv), Dropout 0.2	Block 6	14×14	3×3 × 256 DW, stride 2
Classifier	1×1	7×7 GlobalAvgPool		14×14	1×1 × 256×512, stride 1
	1000	1000D FC, Softmax	Block 7–11	14×14	3×3 × 512 DW, stride 1
				14×14	1×1 × 512×512, stride 1
			Block 12	7×7	3×3 × 512 DW, stride 2
				7×7	1×1 × 512×1024, stride 1
			Block 13	7×7	3×3 × 1024 DW, stride 2
				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.

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 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

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

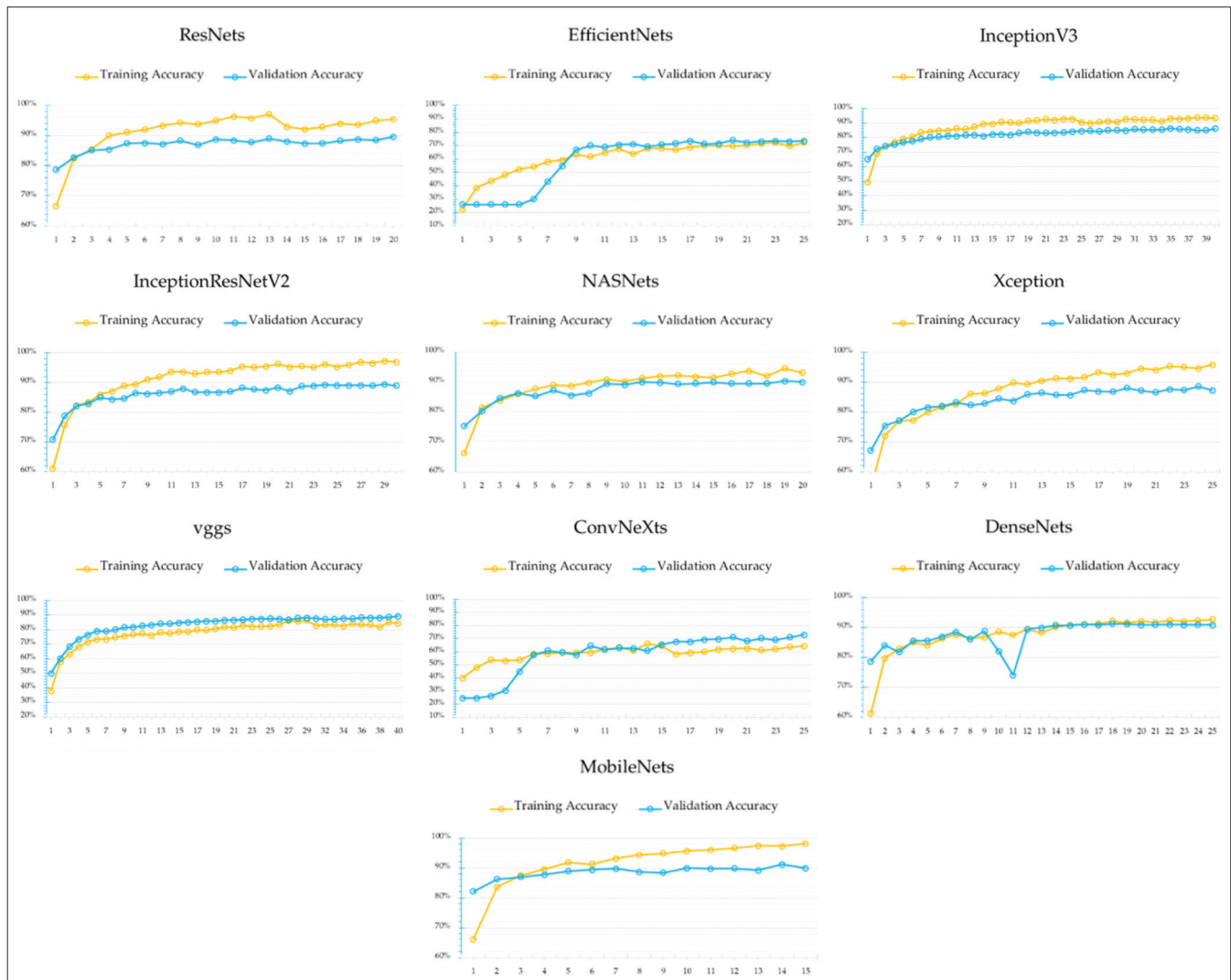


Fig. 7. Different transfer learning model's performance.

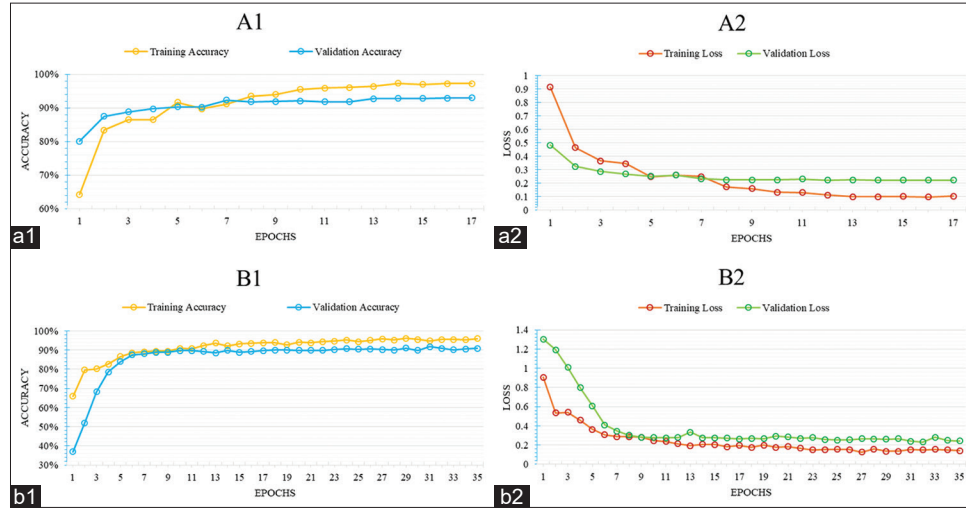


Fig. 8. Accuracy and loss of the proposed hybrid model.

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).

$$Accuracy = \frac{Tp + TN}{Tp + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

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

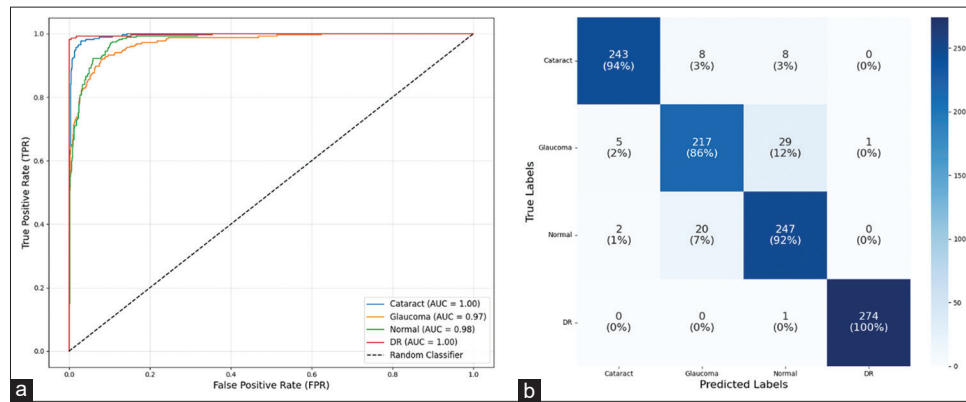


Fig. 9. Receiver operating characteristic curve (a) and confusion matrix (b) produced by the proposed model.

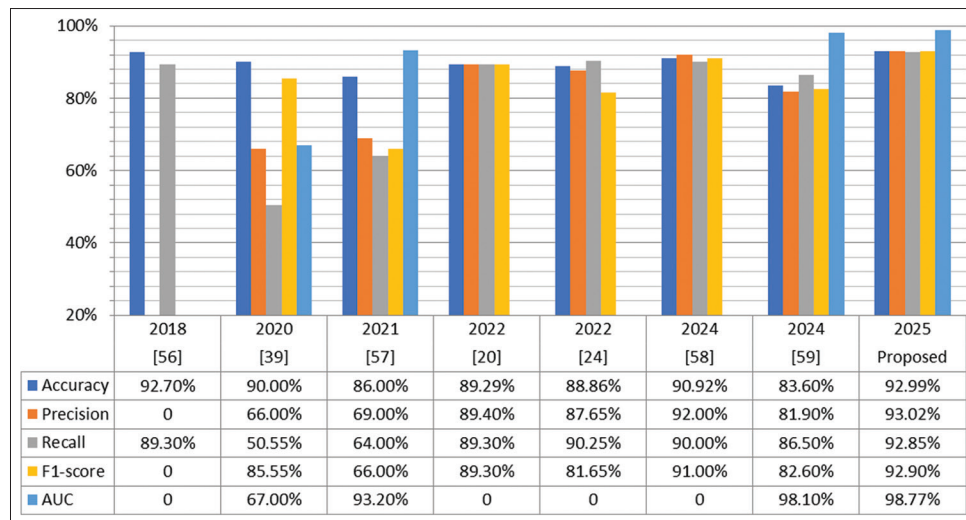


Fig. 10. Comparison between the proposed approach and current methods.

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 and 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 (Allot)*, 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 (TENSYP)*, pp. 1333-1338, 2020.
- [16] M. S. Khan, N. Tafshir, K. N. Alam, A. R. Dhruva, 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. Iliyasa, R. Bestak and F. Shi, Eds. *New Trends in Computational Vision and Bio-inspired Computing: Selected Works Presented at the ICCVIC 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.