ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Iraqi Kurd or Arab Male Authenticity Detection Based on Facial Feature

Bnar Abdulsalam Abdulrahman[1], Nama Ezzaalddin Mustafa[2]

[1]Department of Computer, School of Science, Komar University of Science and Technology, Sulaymaniyah, Iraq,
[2]Department of Computer Science, School of Science and Engineering, University of Kurdistan-Hewler, Erbil, Iraq

ABSTRACT

As an inherent human characteristic, ethnicity plays a fundamental and critical role in biometric identification. On the other hand, the human face is the core of man's identity, and facts such as age and race are often extrapolated automatically from the face. The objective is to utilize computer technologies to identify and categorize ethnic groups based on facial features. Convolutional neural networks (CNN), which can automatically identify underlying patterns from data, excel at learning image features and have shown state-of-the-art performance in several visual recognition challenges, such as ethnicity detection. Although the automated classification of traits such as age, gender, and ethnicity is a well-researched topic, Iraqi ethnic groupings have not yet been addressed. This study seeks to tackle the challenge of predicting the ethnicity of Iraqi male individuals based on their facial traits for the two largest ethnic groups, the Arabs, and the Kurds. Male Iraqi Kurds and Arabs were each represented by 260 image samples. The dataset underwent a diverse array of preprocessing and data enhancement techniques, including image resizing, isolation, gamma correction, and contrast stretching. Moreover, to augment the dataset and expand its diversity, various techniques such as brightness adjustment, rotation, horizontal flip, and grayscale augmentations were systematically applied, effectively increasing the overall number of images, and enriching the dataset for improved model performance. Face images of Kurds and Arabs were classified using the Faster region-based CNN (RCNN) approach of deep learning. Due to insufficient data in the dataset, we propose employing transfer learning to extract features using several pre-trained models. Specifically, we examined EfficientNetB4, ResNet-50, SqueezeNet, VGG16, and MobileNetV2, resulting in accuracies of 96.73%, 94.91%, 93.39%, 92.48%, and 90.32%, accompanied by corresponding precision values of 0.86, 0.81, 0.80, 0.70, and 0.69, respectively. It is essential to emphasize that the following inference speeds – VGG16 (4.5 ms), ResNet-50 (4.6 ms), SqueezeNet (3.8 ms), MobileNetV2 (3.7 ms), and EfficientNet-B4 (16 ms) – represent the computing times needed for each backbone. Moreover, to achieve a harmonious trade-off between precision and the time required for inference, we chose ResNet-50 as the foundational framework for our model aimed at classifying ethnicity. The study also acknowledges limitations such as the availability and diversity of the dataset. Nevertheless, despite these limitations, it provides valuable perspectives on the automated prediction of Iraqi male ethnicity through facial features, presenting potential applications in various domains. The findings contribute to the broader conversation surrounding biometric identification and ethnic categorization, underscoring the importance of ongoing research and heightened awareness of the inherent limitations associated with such studies.

Index Terms: Faster Region-based Convolutional Neural Network, Convolutional Neural Networks, Detection, Iraq, Ethnicity

Access this article online	
DOI: 10.21928/uhdjst.v8n1y2024.pp64-77	**E-ISSN:** 2521-4217 **P-ISSN:** 2521-4209

Copyright © 2024 Abdulrahman 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)

1. INTRODUCTION

Biometric recognition, particularly facial recognition, plays a pivotal role in various domains such as surveillance, advertising, human-computer interaction (HCI), and social media profiling. Ethnicity is a vital component in biometric

Corresponding author's e-mail: Bnar Abdulsalam Abdulrahman, Komar University of Science and Computer Department, School of Science, Komar University of Science and Technology, Sulaymaniyah, Iraq. E-mail:bnar.abdoulsalam@komar.edu.iq)

Received: 19-12-2023 Accepted: 18-02-2024 Published: 03-03-2024

recognition. Classifying individuals by race, nationality, and ethnicity substantially affects HCI, surveillance, and military contexts. Improving understanding of ethnicity has the potential to dramatically accelerate the development of inclusive and culturally sensitive technology in the field of HCI. Designing user interfaces that recognize and adapt to a wide range of ethnic origins makes technology more accessible to a greater range of users. This understanding extends to facial recognition systems, making them more accurate and sensitive to cultural differences. Precise classification of ethnic groups can also lead to advances in surveillance and security applications, such as border control, customs inspections, and public safety measures. In military situations, proper ethnicity identification is critical for information gathering and developing efficient communication strategies [1].

Continually shifting populations reshape the collective identities of regions, nations, and races. Even though it is complex, categorizing individuals by race and ethnicity has aided national censuses and national security in many nations. The classification process is also an important topic in social science, and it is helpful in studies of health care, education, and socioeconomic status [2]. It is economically significant in market research, especially in multiethnic nations.

Iraq is among the nations with the most significant number of ethnic groups. The current demographics of Iraq reveal a multiethnic, multicultural nation sharing the same landmass. According to World Bank statistics, Iraq's population is around 43 million and comprises several ethnic groups [3]. Arabs and Kurds are the majority, while Christians, Turkmen, Assyrians, and Yezidis are the minority. According to estimates, 75–80% of the population is Arab, while 15–20% is Kurdish [4]. Figs. 1 and 2 visually represent the main ethno-religious groups in Iraq and showcase individuals in their traditional attire, respectively.

In recent years, studying racial and ethnic groupings based on facial images has become a vital face recognition issue [2]. Faces provide a great deal of information, including identification, gender, age, ethnicity, expression, *et cetera*. Consequently, there has been a rise in interest in automated face ethnicity classification, and various methods have been created. In an uncontrolled context, it is not easy to properly and efficiently distinguish different races based merely on a human face. Face images must contain racial sensitivity for classification to be effective. These distinctive features may be categorized into chromatic/skin tone and local and global characteristics [5]. Due to the uniformity of skin color

Fig. 1. Main ethnoreligious groups in Iraq.

Fig. 2. Kurdish and Arabs in traditional dress.

between races and the variability of lighting conditions in everyday life, skin tone cannot be used to identify persons on its own. The addition of local or global descriptors may enhance the accuracy of categorization.

Humans have an innate aptitude for facial recognition, enhanced by cultural and cognitive processes, which permits the human visual system to comprehend familiar faces quickly and effectively. Genes, environment, culture, and other variables frequently impact facial characteristics [6].

Nevertheless, the genes of a particular ethnic group are rarely unique and may contain DNA pieces from various ethnic groups. As a result, there may be facial resemblance across various races. It is vital to utilize cognitive and perceptual intelligence while examining the facial traits of people of

different races. This means being able to see subtle variations influenced by genetics, environment, and cultural factors. In addition, the application of advanced technologies and intelligent algorithms contributes to the accurate analysis of facial features.

Considering the previously mentioned aspects, this research aims to construct a system for classifying males within Iraq's principal ethnic groups – namely, the Arabs and Kurds. Nonetheless, distinguishing between Kurds and Arabs based solely on ethnic traits poses significant challenges. Fig. 3 underscores this difficulty, revealing that individuals from these two ethnic groups often share similarities in facial features, including skeletal structure, eye color, nose size, and eye size. The primary distinguishing features, as highlighted in Fig. 4, include skin color and lip shape. However, it is important to note that there are instances where the differentiation based on skin tone and lip shape becomes challenging as well due to occasional similarities between the two groups. While skin color may initially appear to be a significant distinguishing factor, the presence of color variations within the same ethnic group complicates the categorization process based solely on this characteristic. Consequently, the manual extraction of features becomes a complex task.

To overcome these challenges, this research leverages deep learning techniques [7], specifically convolutional neural networks (CNN), to develop a model capable of distinguishing between Kurds and Arabs effectively.

Meanwhile, it is not easy to build a deep learning model without data, which is required for models to gain experience and the capacity to make predictions based on the provided data. Training datasets must be provided to the learning algorithm for deep learning models to learn how to do various tasks, followed by testing and validation datasets to ensure the model processes the data successfully. The quality of its training data hampers every deep learning technique [8]. Due to this and the fact that, as far as is known, there are no current datasets for ethnicity identification in Iraq, this study aimed to provide a suitable dataset to address this issue.

2. RELATED WORKS

Face ethnic identification has gained prominence during the last decade. Numerous research has been conducted to manually extract racial face features through a range of traditional machine learning approaches, and recently,

Fig. 3. Kurdish and Arab facial characteristics.

Fig. 4. Kurdish and Arab face characteristics differ from each other.

the efficient and speedy methods of the deep learning methodology have been used for end-to-end facial detection.

Starting with the traditional approach, one of these researchers, Lin *et al.* 2006 [9], introduced a method for analyzing face images, combining the Gabor filter to extract key facial features, AdaBoost learning to select a series of simple classifiers, and Support Vector Machine (SVM) classifier to recognize the face images. The system has been implemented on the FERET dataset, which focuses on human face images. This approach detects Mongoloid, Caucasian, and African ethnicity classifications. This study demonstrated remarkable accuracy in the identification of

gender, ethnicity, and age from facial images. Notably, it attained an accuracy of 91.58% for females and 90.40% for males, 89.05% for individuals of yellow ethnicity and 89.28% for those of white ethnicity, and 91.77% for younger age groups and 90.32% for older age groups. These results underscore the system's effectiveness in the realm of facial recognition.

Following that on the FERET dataset, Manesh *et al.* 2010 [10] offered a two-class ethnicity categorization for Asian and non-Asian images. The approach assesses the confidence of distinct face areas by applying a modified Golden ratio mask followed by an SVM classifier on facial characteristics such as the eyes, nose, and mouth rather than the whole image. Ensuring that facial patches are in a consistent position for comparison is deemed crucial in face processing. The algorithm they introduced yielded a notable accuracy of 94% on the dataset.

Furthermore, Xie *et al.* [11] proposed a technique for classifying ethnicity on massive face datasets using Kernel Class-dependent Feature Analysis (KCFA) and color-based facial traits. It targets the periorbital region as opposed to the whole face. The experiment used both public and self-collected datasets. Their proposed method exhibits remarkable accuracy, achieving an impressive rate of around 96% in classifying individuals into three distinct ethnic categories: Caucasian, African American, and Asian.

Moving further to the deep learning approach, Wang *et al.* 2016 [7] provide a unique solution to the issue of ethnicity classification using Deep Convolution Neural Networks (DCNN) to extract and categorize characteristics concurrently. Their proposed strategy undergoes rigorous testing across three scenarios: The classification of White and Black, Chinese and Non-Chinese, and Han, Uyghur, and Non-Chinese individuals. The evaluation is conducted using several widely recognized face image databases, including MORPH-II, CASIA-PEAL, and CASIA-WebFace. The outcome of this study illustrates the effectiveness of the suggested method compared to the previous approaches. Nevertheless, the trained model exhibits limited generalization capability when confronted with diverse factors, such as variations in illumination and head pose. Moreover, Srinivas *et al.* 2017 [12] investigated the fine-grained ethnic categorization of the Eastern Asia population. They introduce a new dataset Wild East Asian Face (WEAFD) Dataset that contains seven ethnic groups (Chinese, Filipino, Indonesian, Japanese, Korean, Malaysian, and Vietnamese). This dataset is well suited for performing age, gender, and detailed ethnicity classification tasks. They used CNN to obtain baseline data for the WEAFD. The findings show that determining a person's precise ethnicity is the hardest challenge, followed by determining their age and gender. Furthermore, Masood, *et al.* 2018 [13] utilized a CNN and Artificial Neural Network (ANN) in the FERET dataset to predict three ethnicities: Mongolian, Caucasian, and Negro. Both experiments successfully addressed the task of ethnicity classification. The CNN model outperformed the ANN approach, achieving an impressive accuracy of 98.6%, whereas the ANN approach yielded a comparatively lower accuracy of 82.4%. Recently, Belcar *et al.* 2022 [14] provided an overview of recent advancements in ethnicity classification with an emphasis on CNNs and suggested a novel approach for ethnicity classification utilizing just the central portion of the face and CNN. They used the UTKFace and FairFace datasets, for categorizing (White, Black, Latin, Indian, Middle Eastern, East Asian, and Southeast Asian). In their study, they ascertained that the area encompassing the nose and eyes holds the most significant concentration of visual data, playing a pivotal role in achieving successful ethnicity classification.

Furthermore, Chen *et al.* 2016 [15] used a K-nearest neighbor technique, an SVM classifier, a two-layer neural network, and a CNN to train a classifier to predict Chinese, Japanese, and Korean. Before being put into various learning techniques, the face images were first cropped and augmented. CNN had the highest accuracy rate, at 89.2%. However, the CNN displayed overfitting on new images, suggesting that more varied data and GPU utilization are required for improved outcomes in the future.

Besides this, some studies employed a hybrid strategy, Heng *et al.* 2018 [16] present a hybrid supervised learning strategy for performing ethnicity categorization that utilizes both the power of CNN and the abundant network-obtained information. A supervised hybrid SVM learning method is designed to train the combined feature vectors for ethnicity classification. A dataset including Bangladeshi, Chinese, and Indian ethnic groups is used to test the effectiveness of the suggested technique, which obtained an overall accuracy of 95.2%.

Likewise, Aina *et al.* 2022 [17] created a CNN model for identifying the ethnicity of Nigerians based on face images using transfer learning methods and VGG-16 architecture. The model's evaluation was conducted using a dataset comprising images of Nigerian ethnic groups, namely Yoruba, Hausa, and Igbo, achieving an impressive accuracy

of 92.86%. The research demonstrated that their novel model excels in tasks related to ethnicity classification, particularly when confronted with an exceptionally limited and imbalanced dataset.

3. METHODOLOGY

Several steps comprise the approach outlined in this study. Since no dataset existed for Iraq-specific identification of Arab and Kurdish ethnicity, it was required to create one. A collection of 130 images was assembled for each ethnic group, resulting in a combined dataset of 260 images. While collecting images, male Kurds and Arabs aged 18–70 were observed.

3.1. Pre-Processing

Due to variations in lighting and camera characteristics, the images require to be adjusted before further processing. To address this, the dataset underwent the following preprocessing methods:

- Image resizing is conducted since it alters the size of the images and, if necessary, stretches them to a more suitable pixel-by-pixel size. That will help optimize both training time and computational power. Furthermore, a consistent input size is required by several models, most notably CNNs, necessitating image resizing to ensure accurate processing of inputs.
- Item isolation, which is the ability to isolate an object inside a frame. In other words, implementing this process on the dataset amplifies the intensity of facial features while diminishing the intensity of non-facial elements. After this treatment, only the facial portion of the image is retained [18]. We adopt this approach because our study exclusively focuses on the facial region, which, in turn, plays a pivotal role in achieving enhanced accuracy in our findings.
- Gamma adjustment is employed to enhance visibility in high and low-light conditions by modifying brightness. A gamma value exceeding one results in darker images, while a value less than one produces brighter images. Following the determination of the average intensity level in the region, a new gamma value is selected for optimization [19]. Fig. 5 demonstrates the need to control unnecessary amplification in initially bright images and illuminate each image based on its actual brightness. We incorporate this approach into the dataset to guarantee perceptual uniformity and consistent picture quality throughout the entire set, given the critical role of constant brightness and contrast in object recognition tasks.

Fig. 5. Correction of images using gamma.

- Contrast stretching is an image enhancement technique that attempts to improve the contrast in an image by "stretching" the range of intensity values it contains to span a desired range of values [19]. Some of the images in our collection have insufficient or excessive contrast; thus, preprocessing the image contrast is advantageous. This is used because contrast preprocessing makes edges more visible by amplifying the contrasts between surrounding pixels; it also enhances normalization and line identification under various lighting situations and makes the data simpler for recognition models to understand. The image in Fig. 6 depicts the various preprocessing techniques applied to the dataset.

3.2. Data Augmentation

In addition, following preprocessing, data augmentation techniques are used to enhance the number of training samples and minimize overfitting. The following techniques have been applied to the dataset:

- The first augmentation technique applied is brightness to make the model more adaptable to lighting and camera settings, which alters the brightness of the images by randomly darkening and brightening the average intensity channel [20]. The intensity level is a representation of how light or dark each pixel is in the image.

Fig. 6. Preprocessing techniques employed.

- Furthermore, rotation augmentation is incorporated, which is a crucial upgrade since it modifies the angles at which objects appear in the training dataset; this will introduce variety to rotations and make the model more resistant to camera roll. This strategy may increase diversity to prevent a model from retaining training data and becoming camera orientation insensitive. Along with adding a horizontal flip to make the model face orientation insensitive [20].
- Finally, the grayscale augmentation approach is employed to include grayscale and RGB images in the training set [21]; hence, our model will not rely on skin color to differentiate Arab and Kurdish males since Kurds often have lighter skin than Arabs.

Fig. 7 illustrates the application of augmentation techniques to our dataset, leading to a growth in the training set size to 1,093 images. Adhering to best practices, data augmentation was exclusively performed on the training set, while the test set and validation set retained only the original, non-augmented data. In addition to integrating data augmentation methods during the preprocessing phase, we systematically evaluated their impact on the model's performance using thorough validation and testing protocols. Our aim was to confirm that these augmentation techniques not only enriched the variety within the training dataset but also translated into heightened generalization and resilience when assessing the model. The findings offered valuable insights into the model's generalization capabilities and the real-world effectiveness of the augmentation strategies in unfamiliar scenarios.

Fig. 7. Augmentation techniques applied.

3.3. Faster RCNN

This research investigates the accuracy and speed of inference provided by Faster RCNN, a neural network that achieves exceptional levels of both efficiency and accuracy for image classification applications.

This algorithm is one of the most well-known object detection architectures based on convolution neural networks (CNN). The architecture of the model comprises the network backbone, the region proposal network (RPN), and the fast RCNN algorithm, an older version of this method, as shown in Fig. 8

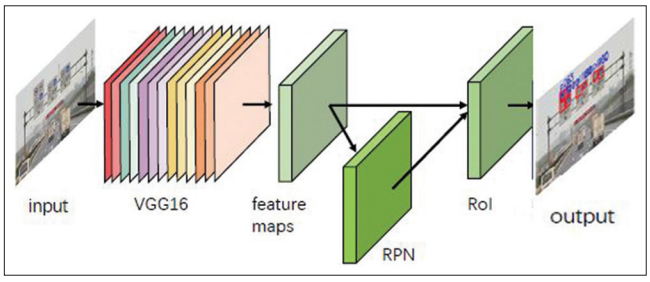

Fig. 8. Faster RCNN Algorithm [24].

- Backbone Network: The backbone is a feature extraction network with pre-trained deep neural network architectures. This network is usually pre-trained on extensive datasets, specifically designed for tasks such as image classification. The backbone operates by producing high-resolution feature maps, effectively capturing hierarchical features from the input image.

- Region Proposal Network (RPN): The RPN is responsible for suggesting potential regions within the image where objects could potentially exist. It predicts a classification score and bounding boxes which are dual predictions: Estimating the probability of an object's presence in a proposed region and determining the coordinates for a bounding box encapsulating that object. To achieve this, the network employs predetermined anchor boxes, pre-defined boxes with various scales, and aspect ratios on the images. These anchor boxes serve as the basis for generating region proposals, tiny portions of an image. The classification score determines whether all anchors contain an object. The bounding box regressions are used to identify the anchors' items accurately. In addition, to address the issue posed by the varying sizes of region proposals, a technique called Region of Interest Pooling (ROI pooling) is employed. ROI pooling ensures the generation of fixed-size feature maps, streamlining the processing for subsequent layers. This strategy overcomes the limitations linked to fully connected layers, thereby improving the model's capability to effectively handle regions of interest with different sizes.
- Fast RCNN: The last part of the architecture takes high-resolution feature maps from the network backbone and region proposals from the region proposal network. Moreover, by employing ROI pooling, the model acquires fixed-size feature maps for each region proposal. Subsequently, these fixed-size feature maps are inputted into fully connected layers to produce SoftMax scores for each class and revised bounding boxes for region proposals [22].

3.3.1. Backbone Networks

The selection of an appropriate network will directly impact the accuracy of the model for feature extraction. Numerous networks have been suggested and have become well-known pre-trained networks used for Deep Learning (DL) models in any AI endeavor. These networks are used for feature extraction or as backbones at the beginning of any DL model. A backbone is a well-established network that has been taught for several previous jobs and exhibits its efficacy. Commonly used backbones include VGG16, ResNet-50, SqueezeNet, MobileNetv2, and EfficientNet-B4. This study examines these currently common backbones to see which one provides a superior outcome when used with faster RCNN for ethnicity categorization.

3.3.1.1. VGG16

VGG16, one of the CNN algorithms, is one of the most advanced computer vision models. The 16 in VGG16 represents 16 weighted layers. There are 21 layers, consisting of 13 convolutional layers, five Max Pooling layers, and three dense layers, but only 16 weight layers (learnable parameters layer). VGG16 accepts 224 by 244 input tensors with three RGB channels. In place of a considerable number of hyper-parameters, VGG16 focuses on convolution layers of 3×3 filters with stride one and always uses the same padding and MaxPool layers of 2×2 filters with stride 2. The convolution and max pool layers are organized appropriately throughout the whole design. Conv-1 layer has 64 filters, Conv-2 contains 128 filters, Conv-3 contains 256 filters, and Conv 4 and Conv 5 have 512 filters. In addition, a stack of convolutional layers is followed by three fully connected (FC) layers: The first two have 4096 channels each, while the third does classification and has 1000 channels (one for each class). Soft-max layer is the last layer [23]. Fig. 9 illustrates the design of VGG16.

The rationale behind choosing VGG16 as the backbone stems from its simplicity and effectiveness in architecture and proven success in various computer vision tasks with its deep 16-layer structures, utilization of small 3x3 convolutional filter, and consistent filters, VGG16 emerges as a dependable choice for feature extraction tasks.

3.3.1.2. ResNet-50

ResNet-50 is a CNN with 50 layers (48 convolutional layers, one MaxPool layer, and one intermediate pool layer). ResNet50, like VGG16, accepts input tensors of sizes 224 and 244 with three RGB channels.

As seen in Fig. 10, ResNet's design includes convolution with a kernel size of 7×7 and 64 different kernels, each of which has a stride size of 2. Afterward, there is a maximum pooling size of 2 strides. In the second convolution, there is a $1 \times 1,64$ kernel, then a $3 \times 3,64$ kernel, and finally a $1 \times 1,256$ kernel; these three layers are repeated three times, resulting in a total of nine layers. Following that, there is a kernel of $1 \times 1,128$, then a kernel of $3 \times 3,128$, and lastly, a kernel of $1 \times 1,512$. This procedure is done four times for a total of twelve layers. In addition, there is a kernel of 1×1256 and two different kernels of $3 \times 3,256$ and $1 \times 1,1024$, and this pattern is repeated six times for a total of 18 layers. And then again, a $1 \times 1,512$ kernel followed by two additional instances of $3 \times 3,512$ and $1 \times 1,2048$, repeated three times to produce a total of nine layers. After that, there is an average pool, which concludes with a wholly linked layer having 1000 nodes and, in the end, a SoftMax function [23].

ResNet-50 is a widely adopted backbone due to its ability to manage deeper networks through residual connections

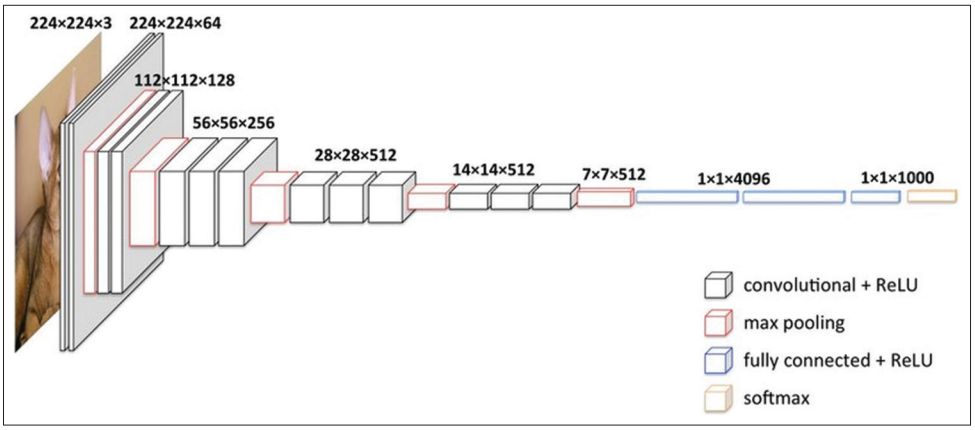

Fig. 9. The architecture of VGG-16 [28].

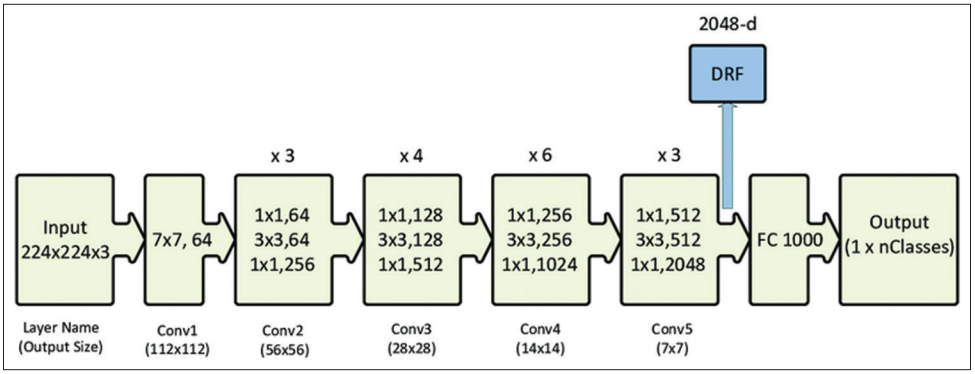

Fig. 10. The architecture of ResNet-50 [22].

and residual blocks. These features address the vanishing gradient problem, enabling the training of exceptionally deep networks and minimizing issues associated with vanishing gradients, making ResNet-50 a reliable option for object detection.

3.3.1.3. SqueezeNet1

SqueezeNet is a deep neural architecture that involves a "squeezed" network for training image classification models. The goal of inventing SqueezeNet was to construct a smaller neural network with fewer parameters (therefore requiring fewer calculations, less memory, and a shorter inference time) that can readily fit into memory devices and be communicated over a computer network with cutting-edge precision [24].

The primary goal of the work was to discover CNN architectures with minimal parameters and competitive accuracy. Therefore, they developed these strategies for improved implementation; they substituted 3 × 3 filters with 1 × 1 filters and used squeeze layers to reduce the number of input channels to 3 × 3 filters, so they deliberately reduced the number of parameters in a CNN while aiming to maintain accuracy. In addition, they down-sampled late in the network such that the activation maps of the convolution layers are large. Here, the theory is that more great activation maps (due to delayed down sampling) may lead to better classification accuracy, everything else remaining equal, and thus will optimize classification accuracy on a restricted parameter allocation.

This model has an image input size of 227 × 227. It consists of convolution layers to extract features, fire modules consisting of a squeeze convolution layer (1×1 filters alone) feeding into an expanded convolution layer with a mixture of 1 × 1 and 3 × 3 convolution filters. Thus, the number of inputs is restricted to 1 × 1, resulting in dimension reduction (limited number of channels). Finally, there are pooling layers for performing down-sampling [24]. Fig. 11 depicts the architectural design of SqueezeNet.

SqueezeNet is chosen for its lightweight design, making it ideal for limited computational resources. It achieves similar

performance as larger models with fewer parameters, and its use of 1 × 1 convolutions reduces parameter count while maintaining expressive capabilities.

3.3.1.4. MobileNet-v2

MobileNet-v2 is a CNN that has been purposefully developed to function optimally in resource-limited and mobile contexts. This model allows any input size more than 32 × 32, and the system's performance is improved for images with larger dimensions. In addition, it has 53 layers of convolution, each of which is either a 1 × 1 convolution or a 3 × 3 depth-wise convolution. This model uses two distinct blocks: A residual block with a stride of 1 and a block with a stride of 2 used for downsizing. In addition, there are three layers for each kind of block: The first layer is a 1 × 1 convolution using the ReLU6 algorithm, the second layer is the depth-wise convolution, and the third layer is another 1 × 1 convolution, but this time without any non-linearities [25]. Fig. 12 depicts an illustration of the architecture of MobileNet-v2.

MobileNetv2 is chosen for its optimal performance on mobile and edge devices, balancing model size, and accuracy. It optimizes computational resources, using inverted residuals and linear bottlenecks while maintaining accuracy.

3.3.1.5. EfficientNet-B4

Constructing neural networks that use convolutional layers require a specific amount of time and resources. When more resources are available, these networks are built later to achieve better degrees of precision. In which the intuition behind the networks is that if the input image is larger, the network needs more layers to increase the receptive field and more channels to capture more fine-grained patterns on the larger image. The network can get away with fewer layers and channels if the input image is smaller [26]. The traditional conventional approaches to model scaling are random; models are scaled in the direction of depth or width. This approach of randomly scaling models requires a substantial amount of manual tuning in addition to person-hours, and the result is frequently either insignificant or nonexistent performance increases. On the other hand, EfficientNet suggested scaling up CNN models to achieve more accuracy and efficiency in a manner far more virtuous using a compound coefficient that scales each dimension consistently using a specified set of scaling coefficients. The concept is to scale with a constant ratio to achieve a balanced relationship between the image's width, depth, and resolution [26].

The architecture of EfficientNet, as shown in Fig. 13, uses a mobile inverted bottleneck convolution, which is quite like

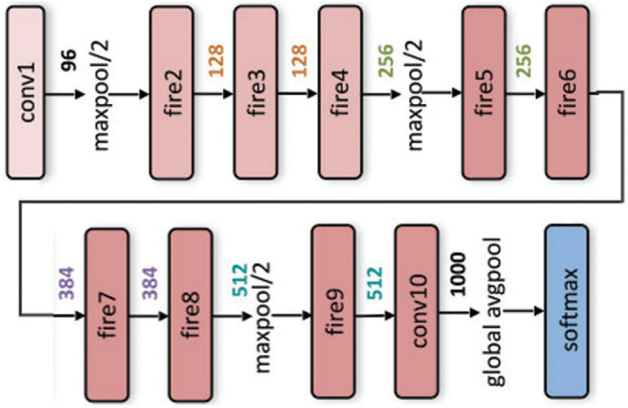

Fig. 11. The architecture of SqueezeNet [29].

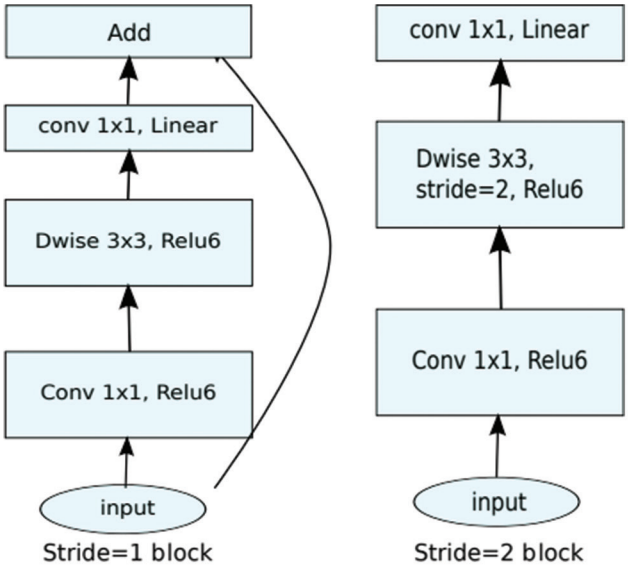

Fig. 12. The architecture of MobileNet-v2 [25].

MobileNet V2 but is much more complex due to the rise in FLOPS. To create the family of EfficientNet, this basic model is expanded to its full extent [27]. In addition, this model takes input images of shapes 224 × 224.

The decision to choose EfficientNet models is grounded in their recognized efficiency, achieved through simultaneous scaling of the model's depth, width, and resolution. Specifically, EfficientNet-B4 is selected for its adept balance between model size and performance. This model introduces a compound scaling method, which efficiently adjusts the model's dimensions, contributing to improved accuracy.

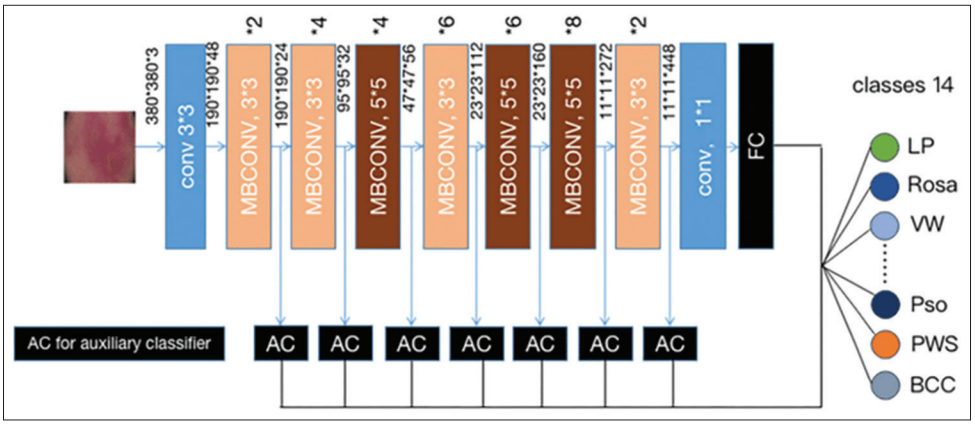

Fig. 13. The architecture of EfficientNet-B4 [30].

4. EXPERIMENTAL RESULTS

In this section, we will begin by describing the dataset partitioning utilized in the study, and then, we will present the findings for Faster RCNN. We have decided to use faster R-CNN as an object detector due to its widespread application and highly accurate object detection that has been reported for various applications. We trained the model using five different backbones, namely VGG16, ResNet-50, SqueezeNet, MobileNetv2, and EfficientNet-B4, to determine whether increased layer depth and model complexity had a positive impact on the detector's overall performance.

4.1. Splitting the Dataset

Since there are no pre-defined established criteria for separating a dataset, researchers often consider the size of the dataset as one of the splitting parameters. Splitting the dataset is required for unbiased evaluation of prediction performance and is often performed to prevent overfitting [25]. We have tried partitioning the dataset to avoid biased results or false impressions. The ratio for the training set is 0.7 to ensure that the model has sufficient data to learn. In addition, 0.2 is specified for the validation set to guarantee that it will lead to accurate model tuning. The remaining 0.1 is reserved for the test set to know how the model would perform on unseen data.

The dataset contains 260 images, each 130 for one of the Kurdish and Arab ethnicities. After splitting and applying data augmentation techniques, there are 1093 images in the training set, 52 in the validation, and 27 in the test set.

4.2. Training and Testing Phases

At first, several hyperparameters were examined in an effort to optimize the Faster R-CNN model's performance.

The model fine-tuning procedure uses a systematic approach with the dual goals of reducing validation losses and enhancing training accuracy. The initial stage in the process was selecting appropriate learning rates according to the architectures that are being used. For VGG16, ResNet-50, and EfficientNet-B4 backbones, the best performance was obtained at a learning rate of 0.001. On the other hand, networks such as SqueezeNet and MobileNetv2 performed better when their learning rate was set to 0.01. In addition, another crucial aspect of the fine-tuning procedure was the thoughtful selection of the batch size. After a thorough investigation of various batch sizes, it was found that 64 was the optimal batch size for the best results. This choice brought computational efficiency and model performance into a perfect balance.

Determining the maximum number of iterations was a crucial choice that determined how well the model worked. This value was first set at 60,000 iterations and was dynamically adjusted. The number of iterations was raised in situations when the validation accuracy indicated improvement. On the other hand, the number of iterations was reduced if overfitting was found. By ensuring that the model was trained for the ideal amount of time, this adaptive technique reduced the likelihood of both underfitting and overfitting.

We carried out iterative trials with hyperparameters, learning from the model's performance on both training and validation sets, with the goal of having the optimum fine-tuning procedure. For optimal performance, this iterative refinement aimed to match the Faster R-CNN model with the particulars of the task and dataset.

Regarding the results, the following points elucidate the key outcomes attained in terms of both accuracy and inference time on our dataset:

- EfficientNet-B4: Achieved the highest accuracy, but with the slowest inference time.

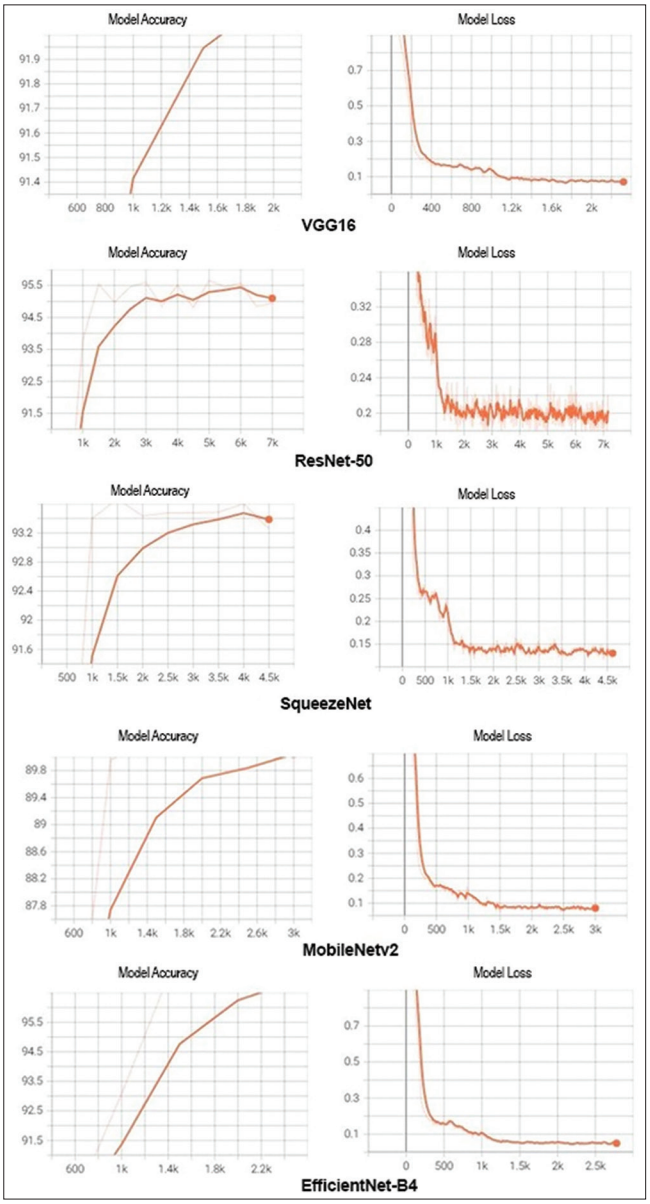

Fig. 14. Training and validation loss and accuracy metrics resulted from Faster RCNN with different backbones.

- MobileNetv2: Provided faster inference, but at the cost of the lowest accuracy.
- ResNet-50: Balanced accuracy, with a reasonable inference time.

The findings are shown in Table 1 when a separate backbone was used as a feature extraction network for the Faster RCNN. Fig. 14 displays the training loss and accuracy learning curves for each of the five backbones used in this research.

As seen from Table 1, our model attained more accuracy when using EfficientNet-B4, with an average mAP of 96.73%; however, its inference time is very sluggish compared to that of other networks since it takes 16 ms to identify a single image. Although Faster RCNN employing MobileNetv2 fared better, with an average inference speed of 3.7 ms, the accuracy is relatively poor compared to other backbones, which had an accuracy of 90.32%. On the other hand, when it comes to the problem of ethnicity categorization, it is preferable to have a model that combines high accuracy with a short inference time. Because of this, the model which makes use of ResNet-50 will be used to address the issue of ethnicity classification in Iraq. This model has a maximum accuracy percentage of 94.91% and an operating time of 4.6 ms for classifying a single image.

The decision to adopt ResNet-50 as the definitive model for ethnicity categorization in our study was reached after a thorough evaluation considering practical factors for real-world deployment, such as:

- This study emphasizes the inherent trade-offs in ethnicity categorization between achieving high accuracy and guaranteeing fast inference times. A fine balance between inference time and accuracy is important in situations when prompt decision-making is required, especially for safety checks, and similar applications.
- Likewise, the importance of minimal inference time cannot be emphasized in real-time deployment scenarios when quick judgments are required. ResNet-50 manages to keep a realistic balance by offering remarkable accuracy without sacrificing comparatively quick inference times.

TABLE 1: Results of faster RCNN with different backbones

Backbone	Arab (%)	Kurd (%)	Total Accuracy (%)	Prec.	Rec.	F_1	Inf. Time (ms)	Training Time
VGG16	91.95	93	92.48	0.70	0.90	0.79	4.5 ms	2 h 5 min 10 s
ResNet-50	92.95	96.88	94.91	0.81	0.90	0.84	4.6 ms	1 h 31 min 5 s
SqueezeNet	90.01	96.77	93.39	0.80	0.91	0.85	3.8 ms	40 min 25 s
MobileNetv2	86.2	94.45	90.32	0.69	0.89	0.78	3.7 ms	1 h 51 min 30 s
EfficientNet-B4	95.61	97.84	96.73	0.86	0.92	0.89	16 ms	3 h 2 min 31 s

- It underscores the challenges associated with inaccurately categorizing male Arab and Kurdish individuals due to facial similarities. Nevertheless, it also demonstrates the commendable performance of ResNet-50 in effectively addressing intricate ethnicity classification issues, rendering it a viable choice for practical applications.

A thorough assessment of several factors, such as accuracy, inference time, interpretability of the model, robustness, and ease of deployment, helped determine the choice of ResNet-50. Because of its shown effectiveness in image classification tasks and the substantial support it receives in the literature, its adoption is strongly validated. The consistent results obtained on our dataset further validate the appropriateness of choosing this backbone architecture.

Figure 15 depicts some ethnicity classification results with a high degree of accuracy, as when the skin tone and facial characteristics are distinct, the model will perform nicely and accurately detect and categorize them.

However, it is crucial to recognize the limitations of our model when evaluating its efficacy, especially in scenarios where Arab males share physical attributes with Kurds. For instance, Arabs with fair skin and Kurds with dark skin, along with other facial features, pose a significant challenge to the model. This similarity may lead to a decrease in accuracy or, in certain cases, result in misclassification, where individuals are mistakenly classified as belonging to the opposing ethnicity. A visual representation of various instances of this complex scenario can be found in Fig. 16.

Fig.15. Detecting Arabs and Kurds accurately by the model.

Fig. 16. Examples of the wrong classification by the model

The model's vulnerability to physical characteristics, especially those tied to appearance, raises concerns about the accuracy of ethnicity categorization. This highlights the necessity for a nuanced strategy that surpasses simplistic traits. Furthermore, there is a crucial demand for a more extensive dataset, recognizing the inherent richness and diversity within Arab and Kurdish ethnic groups. There's apprehension that the current model might not fully grasp the entire spectrum of their diversity, underscoring the urgency for a more comprehensive dataset. Augmenting the dataset to accurately reflect the varied traits within each ethnic group holds the promise of substantially enhancing the model's performance. In addition, the quality and format of the training data significantly impact how well the model performs. Any bias in the data may lead to incorrect classifications, particularly when dealing with similar-looking groups. A detailed analysis of the representativeness and quality of the training data is required to address these constraints.

On the other hand, a significant challenge lies in the model's inability to factor in historical and cultural variables, particularly those related to ethnicity. A more holistic approach is essential, one that considers language patterns and cultural context. Integrating these elements during the model's training phase holds the potential to improve accuracy and reduce misclassifications. Emphasizing the importance of incorporating language nuances and cultural context becomes pivotal in addressing this limitation and elevating the overall performance of the model.

5. CONCLUSIONS

Our study has introduced a robust CNN-based approach for predicting the Kurdish/Arab ethnicity of males in Iraq based on facial images. Our work has important implications across diverse fields such as HCI, face recognition, biometric-based recognition, surveillance, and military.

In the absence of an Iraqi ethnicity dataset, we generated a comprehensive dataset. Our advanced CNN model, faster RCNN, leverages different feature extraction models, including VGG16, ResNet-50, SqueezeNet, MobileNetv2, and EfficientNet-B4.

Regarding accuracy assessment, Faster RCNN with EfficientNet reached the highest accuracy by achieving an average mAP of 96.73%. Remarkably, MobileNetv2 showcased the swiftest operational speed, at a rapid 3.7 ms. Thus, considering the balance between accuracy and inference time, we selected the algorithm with ResNet-50 backbone as the optimal model for ethnicity classification attaining a noteworthy accuracy of 94.91% and a processing time of 4.6 ms.

Future research should focus on creating a more comprehensive dataset that includes a diverse range of gender identities. Although this study focuses on the largest ethnic groups in Iraq, the Kurds, and Arabs, there is potential to broaden its coverage to include all ethnicities in the country. Recognizing potential data biases, particularly in the complicated context of Iraq, is critical for broadening the study to minimize any negative effects on real-world applications. The significance of this research relies on ensuring that the produced technology is not only useful but also adaptable beyond a binary ethnicity categorization. This versatility is necessary for dealing with the region's broad and nuanced ethnic environment. Furthermore, future research should investigate the larger social and ethical implications of facial recognition technology for detecting ethnicity in Iraq. This necessitates a comprehensive and responsible approach to the development and deployment of such systems, as well as an in-depth knowledge of their social impact.

REFERENCES

[1] M. Smith and S. Miller. "The ethical application of biometric facial recognition technology". *Ai and Society*, vol. 37, pp. 167-175, 2022.

[2] N. Narang and T. Bourlai. "*Gender and Ethnicity Classification Using Deep Learning in Heterogeneous Face Recognition. In: International Conference on Biometrics (ICB)*". IEEE, Piscataway, pp. 1-8, 2016.

[3] TWB. "*The World Bank, World Bank Open Data*", 2023. Available from: https://data.worldbank.org

[4] EUAA. "*Religious and Ethnic Minorities, and Stateless Persons*". European Union Agency for Asylum, Grand Harbour, 2021. Available from: https://euaa.europa.eu/country-guidance-iraq-2021/215-religious-and-ethnic-minorities-and-stateless-persons [Last accessed on 2023 Jan 02]

[5] M. Jmal, W. S. Mseddi, R. Attia and A. Youssef. "*Classification of Human Skin Color and its Application to Face Recognition. In: MMEDIA 2014: The Sixth International Conference on Advances in Multimedia*". *IARIA*, 2014.

[6] S. Richmond, L. J. Howe, S. Lewis, E. Stergiakouli and A. Zhurov. "Facial genetics: A brief overview". *Frontiers in Genetics*, vol. 9, p. 462, 2018.

[7] W. Wang, F. He, and Q. Zhao. "*Facial Ethnicity Classification with Deep Convolutional Neural Networks. In: Chinese Conference on Biometric Recognition*". Springer, Berlin, pp. 176-185, 2016.

[8] C. Janiesch, P. Zschech and K. Heinrich. "Machine learning and deep learning". *Electronic Markets*, vol. 31, no. 3, pp. 685-695, 2021.

[9] H. Lin, H. Lu and L. Zhang. "*A New Automatic Recognition System of Gender, Age and Ethnicity. In: Congress on Intelligent Control*

and Automation". vol. 2, no. 3, pp. 9988-9991, 2006.

[10] F. S. Manesh, M. Ghahramani and Y. P. Tan. "*Facial Part Displacement Effect on Template-based Gender and Ethnicity Classification. In: Proceedings of 11th International Conference on Control Automation Robotics and Vision (ICARCV)*". Singapore, pp. 1644-1649, 2010.

[11] Y. Xie, K. Luu and M. Savvides. "*A Robust Approach to Facial Ethnicity Classification on Large Scale Face Databases. In: IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems*". IEEE, Piscataway, pp. 143-149, 2012.

[12] N. Srinivas, H. Atwal, D. C. Rose, G. Mahalingam, K. Ricanek and D. S. Bolme. "*Age, Gender, and Fine-grained Ethnicity Prediction Using Convolutional Neural Networks for the EAST Asian Face Dataset. In: 2017 12th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2017)"*. pp. 953-960, 2017.

[13] S. Masood, S. Gupta, A. Wajid, S. Gupta and M. Ahmed. Prediction of human ethnicity from facial images using neural networks. In: "*Data Engineering and Intelligent Computing*". Springer, Berlin, pp. 217-226, 2018.

[14] D. Belcar, P. Grd and I. Tomičić. "Automatic ethnicity classification from middle part of the face using convolutional neural networks". *Informatics*, vol. 9, no. 1, p. 18, 2022.

[15] H. Chen, Y. Deng and S. Zhang. "*Where am I from?-East Asian Ethnicity Classification from Facial Recognition*". Project study in Stanford University, San Francisco, 2016.

[16] Z. Heng, M. Dipu and K. H. Yap. "*Hybrid Supervised Deep Learning for Ethnicity Classification Using Face Images. IEEE.2018, International Symposium on Circuits and Systems* (*ISCAS*)". Florence, Italy, pp. 1-5, 2018.

[17] S. Aina, M. O. Adeniji, A. R. Lawal and A. I. Oluwaranti. "Development of a convolutional neural network-based ethnicity classification model from facial images". *International Journal of Innovative Science and Research Technology*, vol. 7, no. 4, pp. 1216-1221, 2022.

[18] Roboflow. "*New Feature: Isolate Objects"*, 2021. Available from: https://blog.roboflow.com/isolate-objects [Last accessed on 2025 Jan 25].

[19] S. Rahman, M. M. Rahman, M. Abdullah-Al-Wadud, G. D. Al-Quaderi and M. Shoyaib. "An adaptive gamma correction for image enhancement". *EURASIP Journal on Image and Video Processing*, vol. 2016, no. 1, p. 35, 2016.

[20] OpenGenus. "*Data Augmentation Techniques, Computing Expertise and Legacy*", 2019. Available from: https://iq.opengenus.org/data-augmentation [Last accessed on 2023 Jan 25].

[21] J. Wang and S. Lee. "Data augmentation methods applying grayscale images for convolutional neural networks in machine vision". *Applied Sciences*, vol. 11, no. 15, p. 6721, 2021.

[22] A. Mahmood, A. G. Ospina, M. Bennamoun, S. An, F. Sohel, F. Boussaid, R. Hovey, R. B. Fisher and G. A. Kendrick. "Automatic hierarchical classification of kelps using deep residual features". *Sensors*, vol. 20, no. 2, p. 447, 2020.

[23] D. Theckedath and R. R. Sedamkar. "Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks". *SN Computer Science*, vol. 1, pp. 1-7, 2020.

[24] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally and K. Keutzer. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and and <0.5MB model size". [arXiv preprint] arXiv:1602.07360, 2016.

[25] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. C. Chen. "*Mobilenetv2: Inverted Residuals and Linear Bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*". pp. 4510-4520, 2018.

[26] M. Tan and O. Le. "*Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks. In: International Conference on Machine Learning"*. pp. 6105-6114, 2019.

[27] A. Shahid. "*EfficientNet: Scaling of Convolutional Neural Networks Done Right Medium"*, 2020. Available from: https://towardsdatascience.com/efficientnet-scaling-of-convolutional-neural-networks-done-right-3fde32aef8ff [Last accessed on 2023 Feb 03].

[28] B. Shi, R. Hou, M. A. Mazurowski, L. J. Grimm, Y. Ren, J. R. Marks, L. M. King, C. C. Maley, E. S. Hwang and J. Y. Lo. "*Learning Better Deep Features for the Prediction of Occult Invasive Disease in Ductal Carcinoma in Situ through Transfer Learnin. In: Proceedings of the SPIE Medical Imaging 2018: Computer-Aided Diagnosis"*. vol. 10575, pp. 620-625, 2018.

[29] T. H. B. Nguyen, E. Park, E., X. Cui, V. H. Nguyen and H. Kim. "fPADnet: Small and efficient convolutional neural network for presentation attack detection". *Sensors*, vol. 18, no. 8, p. 2532, 2018.

[30] C. Y. Zhu, Y. K. Wang, H. P. Chen, K. L. Gao, C. Shu, J. C. Wang, L. F. Yan, Y. G. Yang, F. Y. Xie and J. Liu. "A deep learning-based framework for diagnosing multiple skin diseases in a clinical environment". *Frontiers in Medicine*, vol. 8, p. 626369, 2021.