RESEARCH ARTICLE

DeepDiabetic: An Identification System of Diabetic Eye Diseases Using Deep Neural Networks

ARWA ALBELAIHI[1] AND DINA M. IBRAHIM[1,2]

[1]Department of Information Technology, College of Computer, Qassim University, Buraydah 51452, Saudi Arabia
[2]Computers and Control Engineering Department, Faculty of Engineering, Tanta University, Tanta 31733, Egypt

Corresponding author: Dina M. Ibrahim (d.hussein@qu.edu.sa; dina.mahmoud@f-eng.tanta.edu.eg)

ABSTRACT Deep Learning (DL) plays a successful and influential role in medical imaging diagnosis, image detection, and image classification. Diabetes is a significant public health concern, and diabetic eye disease will be the leading cause of vision loss across the globe. This research proposed a multi-classification deep learning model for diagnosing and identifying four different diabetic eye diseases: Diabetic Retinopathy (DR), Diabetic Macular Edema (DME), glaucoma, and cataract which we called the DeepDiabetic framework. The proposed models were assessed using 1228 images from six different available datasets (DIARETDB0, DIARETDB1, Messidor, HEI-MED, Ocular, and Retina). In addition to the original dataset, we measured the performance of the deep learning models according to two different geometric augmentation methods called online augmented and offline augmented. The present work considers five architectures' performances: EfficientNetB0, VGG16, ResNet152V2, ResNet152V2 + Gated Recurrent Unit (GRU), and ResNet152V2 + Bidirectional GRU (Bi-GRU). A comprehensive analysis and evaluation of these deep learning architectures is provided using public fundus datasets with four classes (i.e., DR, DME, Glaucoma, and Cataract). To the best of our knowledge, no other deep learning models for choosing between these models for these specific diseases are found in the literature. The results of the experiments showed that the EfficientNetB0 model outperforms the other four proposed models. The EfficientNetB0 model achieved 0.9876 in accuracy, 0.9876 recall, 0.9876 precision, and 0.9977 AUC based on fundus images. Our EfficientNetB0 model achieves 98.76% accuracy, while the previous studies only achieved 88.33%, 89.54%, 97.23%, and 80.33% accuracy, respectively. When compared to the previous studies as Fast-RCNN, RCNN-LSTM, and InceptionResNet, our EfficientNetB0 model achieves much higher accuracy, recall, precision, and AUC. According to the outcomes, our proposed models, especially the EfficientNetB0 model, are significantly more accurate than the state-of-the-art models.

INDEX TERMS Deep learning, diabetic eye disease, multi-class classification, offline augmented, online augmented, retinal fundus images.

I. INTRODUCTION

Deep learning (DL) is being used in many domains to discover new approaches to urgent issues and is demonstrating remarkably good performance in categorization tests. Artificial intelligence (AI) tools and techniques are appropriate in medicine. AI is one of the most powerfully transformative technologies in the 21st century. This transformation occurred through the use of powerful machine learning (ML) tools and techniques such as deep convolutional networks, generative adversarial networks (GANs), deep reinforcement learning (DRL), convolutional neural networks (CNNs), recurrent neural network (RNN), and artificial neural networks (ANNs). Lately, deep learning (DL)

The associate editor coordinating the review of this manuscript and approving it for publication was Ikramullah Lali.

has been outperforming traditional artificial intelligence (AI) in crucial tasks like speech recognition, image characterization, and natural language generation.

DL has been proven to be a successful and influential method in many fields of medical imaging diagnostics in image detection and classification. DL can be used for detecting and classifying eye diseases, including diabetic eye disease, by employing fundus images to analyze and diagnose eye diseases. Diabetic retinopathy (DR), diabetic macular edema (DME), glaucoma, and cataracts all fall under the umbrella of diabetic eye disease. Patients between the ages of 20 and 74 can suffer severe vision loss or perhaps go low vision as a result of diabetic eye disease. Preventing vision loss is impossible without early identification of diabetic eye illness. Ninety percent of diabetics can avoid diabetic eye damage if they are detected [1].

This research aims to enhance DL detection models for diabetic eye diseases. Fundus images of diabetic eye disease should be collected for use as input into DL models. The images are then processed using several image preprocessing techniques. Extracting features and learning analysis rules is done automatically using pre-processed images. The authors in [2] reviewed the interest in the detection and classification of respiratory diseases employing ML including the most significant articles published between 2018 and the end of 2021. Because of these findings, the researchers may better organize their work and contribute more effectively to the field. So, we found Ibrahim et al. [3] employing TL on a combination of chest X-ray and CT scan images for multiclass classification using the VGG19-CNN, ResNet152V2, ResNet152V2 + gated recurrent unit (GRU), and ResNet152 V2 + bidirectional GRU (Bi-GRU) architectures. The developments of their research were very rich and motivated us to apply GRU and Bi-GRU in the field of multiclassification of diabetic eye diseases, especially as they are models that have not been used before in this field. Furthermore, Eko et al. [4] discovered the combination of EfficientNetB7 models and some augmentation processes obtain increased classification accuracy for diabetic eye diseases.

In this work, models such as VGG16, EfficientNetB0, and ResNet152V2 are chosen as a CNN pre-trained model. In addition, ResNet152V2 is even integrated with two RNN models: GRU and Bi-GRU. Due to limitations of the availability of datasets, we applied two different geometric augmentation methods to measure the performance of the deep learning models besides the original dataset. These methods are the online augmented and offline augmented geometric augmentation methods.

Diabetes patients are one of the world's most populous disease categories today. diabetic fundus diseases, the leading cause of blindness, can cause visual loss. DR, Glaucoma, cataracts, and other fundus diseases now affect visual function [2]. When fundus disease reaches a late stage, it greatly impairs the patient's vision and cannot be specifically treated. Because even highly skilled eye specialists often misidentify

eye lesions, reliably diagnosing diabetic retinopathy using retinal fundus images is difficult. Supporting a technique that aids in disease detection is beneficial since early identification of diseases helps avoid blindness.

Due to its great feature-learning ability, CNN has achieved exceptional success in the field of fundus images. The literature has presented several deep-learning architectures that have achieved excellent results in detecting certain diabetic eye diseases. To the best of our knowledge, at the beginning of our research, the classification model of the four diseases of diabetes (DR, DME, Glaucoma, and Cataract) had been rarely touched. But recently, by the end of 2021, we found Rubina Sarki et al. [5] presented a classification framework for those four diseases. However, they only applied one CNN model in their work, and their dataset showed class imbalance. They attained only 81.33% accuracy, which is still considered low compared to recent research in DL classifications.

According to the previous studies, cataract disease did not have sufficient studies investigating cataract classification and prediction [6]. They were also independent studies to classify the cataract disease separately. Reference [7] recommended DME disease detection as the gap in research direction since detecting DME is highly likely to mean that the retina is developing DR, which supports researchers in understanding the causes of retina-based diseases better. The further DL model concerns training minimal data and datasets with a class imbalance between different diseases. The accuracy of the findings produced may suffer if the training set is too small. Utilizing improvement strategies via traditional data augmentation is looked at as an appropriate choice.

The main contributions of this study are:
- Propose the DeepDiabetic Framework, a multi-classification deep learning model for detecting and recognizing the four most common complications of diabetes affecting the eyes: diabetic retinopathy (DR), diabetic macular edema (DME), glaucoma, and cataract.
- We employed both online and offline geometric augmentation methods to assess the accuracy of the deep learning models, in addition to the original dataset.
- EfficientNetB0, VGG16, ResNet152V2, ResNet152V2 + Gated Recurrent Unit (GRU), and ResNet152V2 + Bidirectional GRU (Bi-GRU) are the five architectures whose performances are taken into consideration in this work. Using public fundus datasets with four classes, a thorough examination, and evaluation of these deep learning architectures (i.e., DR, DME, Glaucoma, and Cataract). As far as we are aware, there aren't any other GRU models used for classifying these models for these particular diseases in the literature.
- Comparing the performance measures of our proposed work with different models reported in the prior studies for diagnosing and classifying diabetic eye disorders.

The remainder of the paper is outlined as follows: Section II presents a summary of the literature review,

Section III describes the dataset of the study, and image augmentation explains the experimental setup conducted to create the framework, including the deep learning proposed classification models. Section IV presents the results and discussions while Section V illustrates the comparative analysis and discussion. Finally, Section VI concludes the paper.

II. LITERATURE REVIEW

In this section, we introduce some related studies on diabetic eye diseases using deep neural networks. We discovered that building an efficient neural network classifier requires careful consideration of both the network architecture and the data input. The literature has much research about the use of deep learning as a classification model for diabetic eye disease using fundus images, as in Refs. [14], [15], [25], and [26].

The principle of Transfer Learning (TL) approach is commonly adopted in diabetic eye disease, including the authors in [29], [30], [31], [32], [33], and [34]. Instead of generating parameters at random, the TL initializes them using information from prior learning. The initial layers intuitively learn to extract core features similar to edges, textures, etc. The top layers, similar to blood vessels and exudates, are more particular to the task. In situations where there is insufficient data to train a neural network from scratch, the TL is employed effectively in [29], [32], [33], and [34].

Pan et al. [29] Compared three CNN models: DenseNet, ResNet50, and VGG16, among four types of lesions of DR. The experimental outcome shows the effective model is DenseNet to identify and differentiate retinal lesions automatically in multi-label classified images. Nevertheless, the procedure does not accurately determine microaneurysms since they are easily misclassified in the pervasive presence of fluorescein.

Moreover, Samanta et al. [30] created TL architecture CNN-based on color fundus photography that executes relatively well in the identification of DR (No DR, Mild DR, Moderate DR, and Proliferative DR) from hard exudates, blood vessels, and texture on limited dataset. They utilized their model with several frameworks, including Inceptionv1, Inceptionv2, Inceptionv3, Xception, VGG16, ResNet-50, DenseNet, and AlexNet.

Zhang et al. [31] presented a framework for actively detecting the existence and severity of DR. By Utilizing various TL architectures: ResNet50, InceptionV3, DenseNets, Xception, and InceptionResNetV2. Despite the proposed framework attaining a sensitivity of 97.5% and a specificity of 97.7%, their model must be assessed using a more extensive and detailed dataset.

Additionally, the Lookahead optimizer and CNN model were utilized for the image classification of cataract disease in [32] to increase accuracy and minimize processing time. Through the CNN-AlexNet architecture and the Lookahead optimizer on Stochastic gradient descent and Adam, the model successfully identified the label of the images. As a result, CNN-AlexNet enhances optimizer Stochastic gradient descent by 2.5% and increases accuracy by 20%.

Sarki et al. [34] introduced a deep learning architecture with 13 CNN models merged with image processing techniques to endorse early detection of diabetic eye diseases (DR, DME, glaucoma). The early classification of diabetic eye disease had several limitations that were realized. Later, they produced an automatic classification scheme that examined multi-class and mild multi-class diabetic eye disease [33]. Different performance improvement techniques, including fine-tuning, optimizing, and contrast enhancing, were employed by applying the VGG16 and InceptionV3. In addition, the VGG16 model attained accurate results of 88.3% for multi-class classification and 85.95% for mild multi-class classification.

To retrain a new image, such as a set of medical images, there are several TL implementations available, largely in DL. However, this architecture performs less effectively when classifying medical images. Because such TL frameworks are made for purposes like animals, flowers, etc., Pan et al [29]'s utilization of VGG16 for DR detection using fundus images achieved around 79.6% specificity. Hence, TL methods are not suitable for real-time medical images.

Recent research has integrated ML and DL classifiers, including random forests (RF), support vector machines (SVM), naive Bayes, and decision trees, to detect diabetic eye diseases [35], [36], [37]. Grassmann et al. [35] trained several CNNs for the severity scale for age-related macular degeneration separately, and based on the outcome of the single CNNs, RF algorithms were trained to build a model ensemble. A framework with several ML techniques, such as Decision Tree, RF, naive Bayes, and artificial neural network algorithms, was developed by Malik et al. [36]. The researchers discovered that tree-based techniques outperformed artificial neural networks. There are various types of studies that combine DL with ML using SVM, as SVM is a rapid and accurate classification method that works well with limited evaluation data. Theera-Umpon et al. [21] used SVM, hierarchical adaptive neuro-fuzzy inference system, multilayer perceptron network, and CNNs to develop a model to detect hard exudates with a limited dataset. They discovered that CNN performs well; however, it is not the greatest classifier for hard exudate. Furthermore, a combination of effective classifiers and image-processing techniques can achieve satisfactory results in hard extraction. A recent paper illustrated the combination of DL and ML models, CNN based on the Resnet152 model, and several ML classifiers, including SVM, K-Nearest Neighbors, Naive Bayes, Multi-layer Perceptron, Decision Tree, and Random Forest, to detect human eye infections of Glaucoma disease.

Various DL diabetes-based eye disease techniques have been built in several studies, which employed RCNN for localization and segmentation [27], [42], [43]. Although the technique is robust for glaucoma detection, it is computationally complex. The effectiveness is affected by

growing the network's hierarchy, which affects the loss of the discriminated collection of features.

Recently, researchers have developed new network architectures as an alternative to TL, as in [38], [39], [40], and [41]. The authors have created their DL frameworks to detect diabetic eye diseases automatically with the indicated classifier, the number of layers, and a model used, and illustrate the findings obtained.

In [38], two models were presented, the first to predict the DR and the second to classify the five DR stages. The Siamese-like CNN structure employed in the Zeng et al. framework was trained using TL by the weight-sharing layer principle based on Inception V3. The model, unfortunately, needed paired fundus photos to work; therefore, datasets without paired fundus images may not benefit as much from it.

The authors of research [39] presented a DL approach to predict the expected DR class and assign scores to individual pixels to demonstrate their importance in each input sample. Then the assigned score is to make the final categorization decision. Their model was provided, which increased its sensitivity and specificity values by more than 90%. However, by taking the proper measures, the evaluation performance of the learning process can be improved.

A new approach to detect hypertensive retinopathy was developed by Abbas et al. [40] and consists of multiple multilayer architectures with trained and dense feature layers combined within a CNN algorithm. Combining dense feature transform and dense, hyper learning to improve classification accuracy. They came to a substantial finding, obtaining an accuracy of 95% and AUC values of 0.96. However, a limited dataset was used to produce a highly accurate DL method. Araujo et al. established a DL model to aid in decision-making by providing attention maps that can be understood medically and an approximation of how uncertain that prediction is, leaving it up to the ophthalmologist to determine how much of the result could be trusted [41]. To infer an image grade associated, they developed a Gaussian-sampling technique built-in Multiple Instance Learning structures, including a description graph and prediction uncertainty. Consequently, low-quality images are frequently linked to greater uncertainty, indicating that unsuitable images for processing result in less precise predictions.

In [8], the authors presented a framework that integrated the retrieved features to leverage the benefits of the chosen models, VGG and InceptionV3. The classification stage was proceeded by the use of the entropy concept. Their models are appropriate for classifying the highlights such as veins, liquid dribbling, exudates, hemorrhages, and miniature-scale aneurysms.

Recently, Rubina Sarki et al. [5] introduced the CNN model for the multi-classification of DR, DME, Cataract, Glaucoma, and healthy classes. The datasets were gathered from four different open-source Messidor, Messidor-2, DRISHTI-GS, and

Retina datasets. They reached significant results, accuracy was 81.33%, the sensitivity was 100%, and the specificity was 100%. RMSprop optimizer was used to enhance their model; which optimizer technique offered in [28] was based on the adaptive learning rate. However, they only applied one CNN model in their work, and their dataset showed class imbalance. They achieved only 81.33% accuracy, which is still considered low compared to recent research in DL classifications.

A recent study utilized CNN based on several TL models. In [20] applied VGG16 based on CNN architecture, to classify normal and abnormal eye diseases. Similarly, [24] employed CNN with residual networks to produce the results of ocular damage levels. An enhanced activation function that automatically reduces loss and processing time in the various CNN models was proposed in [23] for diagnosing DR using fundus images. With the Kaggle dataset, the ResNet-152 model has the highest accuracy (99.41%).

On the other hand, some researchers focused on improving the hybrid DL features to attempt to solve class imbalance issues in an available dataset as [16] used TL to extract funds image features with ResNet-18 and GoogleNet models. Likewise, [17] used a hybrid DL classifier to enhance the image quality by taking advantage of the new architecture, which is designed based on ResNet50.

The main mechanism for identifying ophthalmology diseases is fundus imaging. Finding a fund's image dataset that covers diabetic eye disease is considered to be the most challenging stage during the research. Recently, an Ocular Disease Intelligent Recognition (ODIR) dataset containing eight classes of eye diseases included three diabetic eye diseases (DR, Glaucoma, and Cataract) [45]. Studies have emerged using the ODIR dataset to provide a multi-classification of eye diseases [12], [13], [18]. The R-CNN+LSTM architecture was employed in [12] to extract deep features, using a residual approach, and the LSTM model was added to the RCNN+LSTM model, the classification accuracy improving by 4.28% and 1.61%, respectively. Moreover, In [13], the author utilized the ODIR dataset for four classification models based on TL: the VGG-16 model that reached an accuracy of 97.23%; the Resnet-34 model obtained an accuracy of 90.85%; the MobileNetV2 model achieved an accuracy of 94.32%, and the EfficientNet model attained an accuracy of 93.82%. The suggested architecture in [18] enables CNN to learn discriminative features with an attention module without adding additional cost. They solved the issue of class balancing in the extremely imbalanced ODIR dataset with several common labels for a patient's left and right fundus picture pairs.

In this research, we focus on evaluating various image augmentation techniques as preprocessing steps carried out before training the DL models. Images pre-processed with different filters before being entered into the training model produced significant results, as experimental results

TABLE 1. Overview of studies using deep learning approaches with their working methods and performance metrics for Diabetic eye disease images.

Study	Methods	Result	Comments
[29]	DenseNet, ResNet50, and VGG16 models	DR lesion identification and the process is efficient in terms of computing	The method does not accurately identify microaneurysms are easily misclassified in the pervasive influence of fluorescein.
[30]	Inceptionv1, Inceptionv2, Inceptionv3, Xception, VGG16, ResNet-50, DenseNet, and AlexNet.	The validation set scored a value of 0.8836 for Cohen Kappa of while the training set scored 0.9809.	TL and fine-tuning on the pre-trained DenseNet was achieving high classification results.
[31]	ResNet50, InceptionV3, Inception, ResNetV2, Xception and DenseNets	Attained a sensitivity value of 97.5%, together with a specificity value of 97.7%.	The model requires to assessment with a more complex and larger dataset.
[32]	CNN AlexNet architecture with the Lookahead optimizer on Stochastic gradient descent and Adam.	Increases accuracy Adam by about 20% and improves optimizer Stochastic gradient descent by about 2.5%.	Lookahead optimizer improves the accuracy.
[33]	VGG16 and InceptionV3 and other different performance enhancement methods were used, i.e., fine-tune, optimization, and contrast enhancement.	The highest accuracy of 88.3% was achieved on the VGG16 model to multi-class classification also 85.95% to mild classification	The model classified four different diseases from retinal fundus images. However, the result is considered to be low accuracy, not only not up to the mark.
[37]	Using supervised learning techniques including SVM, MLP network, hierarchical ANFIS, and CNNs	Achieve an AUC of 0.998. The AUCs are more than 0.95 for all four classifiers.	The combination of image processing techniques & suitable classifiers in hard exudate can perform very well.
[35]	Train multiple CNNs independently, and RF algorithms trained to build a model ensemble.	Achieve, 94.3% of the healthy fundus photographs were correctly classified.	CNN performs well but not the best suitable the classifier in hard exudate.
[36]	Decision Tree, RF, Naive Bayes and artificial Neural Network algorithms	Found that methods based on tree performed better than the artificial neural network	RF algorithm had a marginally longer execution time than the decision tree algorithm.
[38]	Siamese-like CNN structure trained with TL by the concept of the weigh-sharing layer based on Inception V3	Reached a kappa score of 0.829.	The model does not perform well for those datasets where paired fundus images are not available.
[39]	Allocate scores to distinct pixels to display their relevance in each image sample then employed the assigned score to take the decision	Introduced DL system gained more than 90% of sensitivity and specificity values.	The evaluation performance of the learning the procedure can be enhanced through suitable measures.
[40]	Comprising different multilayer architectures with various trained features layer and dense feature transform layer integrated within a CNN algorithm	Obtained accuracy of 95% and 0.96% of AUC.	Dataset used is small to generate a highly accurate DL algorithm.
[8]	Employed feature fusion and selection procedures. The usage of a CNN for DR grading was then explored, including the VGGNe and InceptionV3 models.	Obtained accuracy of 96.4%	Investigated for DR grading based on international Clinical DR severity scale.
[5]	Improved the CNN model with RMSprop optimizer for multi-classification of DR, DME, Cataract, Glaucoma, and healthy classes.	Reached significant results, accuracy of 81.33%, sensitivity and specificity of 100%.	Applied one CNN model, and dataset showed class imbalance. Reached only 81.33% accuracy.
[41]	They created a Gaussian-sampling strategy built-in Several Instance Learning structures to deduce an image grade associated.	There were come with a Cohen Kappa score between 0.71 and 0.84 in five different datasets.	Low-quality images are commonly associated with higher uncertainties, recording that images not fitting for analysis lead to less accurate predictions

demonstrated in [22]. Table 1 shows an overview of studies using deep learning approaches with their working methods, performance metrics results, and comments on these studies for diabetic eye disease images. Table 2 illustrates an overview of studies using deep learning approaches with their working methods and performance metrics for diabetic eye disease images that classify at least three diabetic diseases from the four diseases focused on in this study.

III. MATERIALS AND METHODS

The purpose of this study is to compare the ability of several deep-learning models to detect these four diabetes-related eye diseases. Six different open-source databases will be utilized to generate our dataset. After that, the gathered images go through a series of preprocessing steps. Following the completion of the dataset, five distinct models are used to determine how each disease should be classified. The pre-trained CNN models VGG16, EfficientNetB0, and ResNet152V2 are selected. The RNN models gated recurrent unit (GRU) and bidirectional gated recurrent unit(Bi-GRU) are also coupled with ResNet152V2. Next, we have the VGG16, EfficientNetB0, ResNet152V2, GRU+ResNet152V2, and Bi-GRU+ResNet152V2 models for deep learning. The models for diabetic eye diseases were then assessed using a variety of criteria. Figure 1 depicts a summary of our proposed DeepDiabetic framework methodology.

TABLE 2. Overview of studies using deep learning approaches with their working methods and performance metrics for Diabetic eye disease images that classify at least three diabetic diseases.

Study	Method	Dataset	Disease	Performance
[33]	VGG16	Messidor, Messidor-2, DRISHTI-GS, and Retina	DR, DME, and Glaucoma	Acc.=88.3% Pre.=98% F1=91.03% Rec.=85%
[27]	Fast RCNN	DIARETDB1, Messidor, ORIGA, DR-HAGIS, and HRF	DR, DME, and Glaucoma	Acc.=95.2% Pre.=95% F1=96% Rec.=96.2% AUC=96.3%
[12]	RCNN+LSTM	ODIR	DR, Cataract, Glaucoma and others	Acc.=89.54% F1=89.97% AUC=97%
[13]	VGG16	ODIR	DR, Cataract, Glaucoma and others	Acc.=97.23% Pre.=96.73% F1=95.22% Rec.=93.76%
	ResNet34			Acc.=90.85% Pre.=93.7% F1=92.65% Rec.=93.17%;
	MobileNetV2			Acc.=94.32% Pre.=93.33% F1=89.67% Rec.=91.46%
	EfficientNet			Pre.=92.73% F1.=96.25% Rec.=93.74% AUC=93.82%
[18]	InceptionResNet	ODIR	DR, Cataract, Glaucoma and others	Acc.=94.38% Pre.=94.28% F1=94.58% Rec.=96.08% AUC=96.08%

Acc.= Accuracy, Pre.= Precision, F1=F1_score, Rec.=Recall

TABLE 3. The details of the collected dataset for the four DR, Cataract, Glaucoma, and DME classes.

Disease	Dataset	images	Total	Sample
DR	DIARETDB0	100	219	
	DIARETDB1	119		
	Messidor	151		
DME	HEI-MED	169	320	
	ODIR	312		
Cataract	Retina	100	412	
	ODIR	178		
Glaucoma	Retina	99	277	

A. DATASETS FOR THE STUDY
1) DATA COLLECTION
For our experiments, many sources of fundus images were accessed and collated. This collection includes DR, Cataract, Glaucoma, and DME. First, for DR, we selected datasets from DIARETDB0 and DIARETDB1 [44] with about 219 images. Secondly, we used Messidor [46] and HEI-MED [47] for DME with 151 images and 169 images, respectively. The third dataset source is the ODIR dataset [45] which contains multiple classes of eye diseases. We used 312 images for Cataracts and 178 images for Glaucoma. The last selected dataset for Cataracts and Glaucoma also is a Retina dataset [48] of 100 images for Cataracts and 99 images for Glaucoma. The outcomes of the collected datasets are 412, 320, 280, and 219 images for Cataract, DME, Glaucoma, and DR, respectively, as shown in Table 3. There are a total of 1228 images obtained from the datasets collected, and these images have been randomly divided into training and validation sets. Figure 2 shows samples of fundus images for DR, Cataract, Glaucoma, and DME.

2) DATASET PRE-PROCESSING
Pre-processing is typically utilized to get the dataset ready to fulfill the prerequisites of the deep learning model. The first technique is to split the data between the four training classes: DR, Cataract, Glaucoma, and DME. Then, the
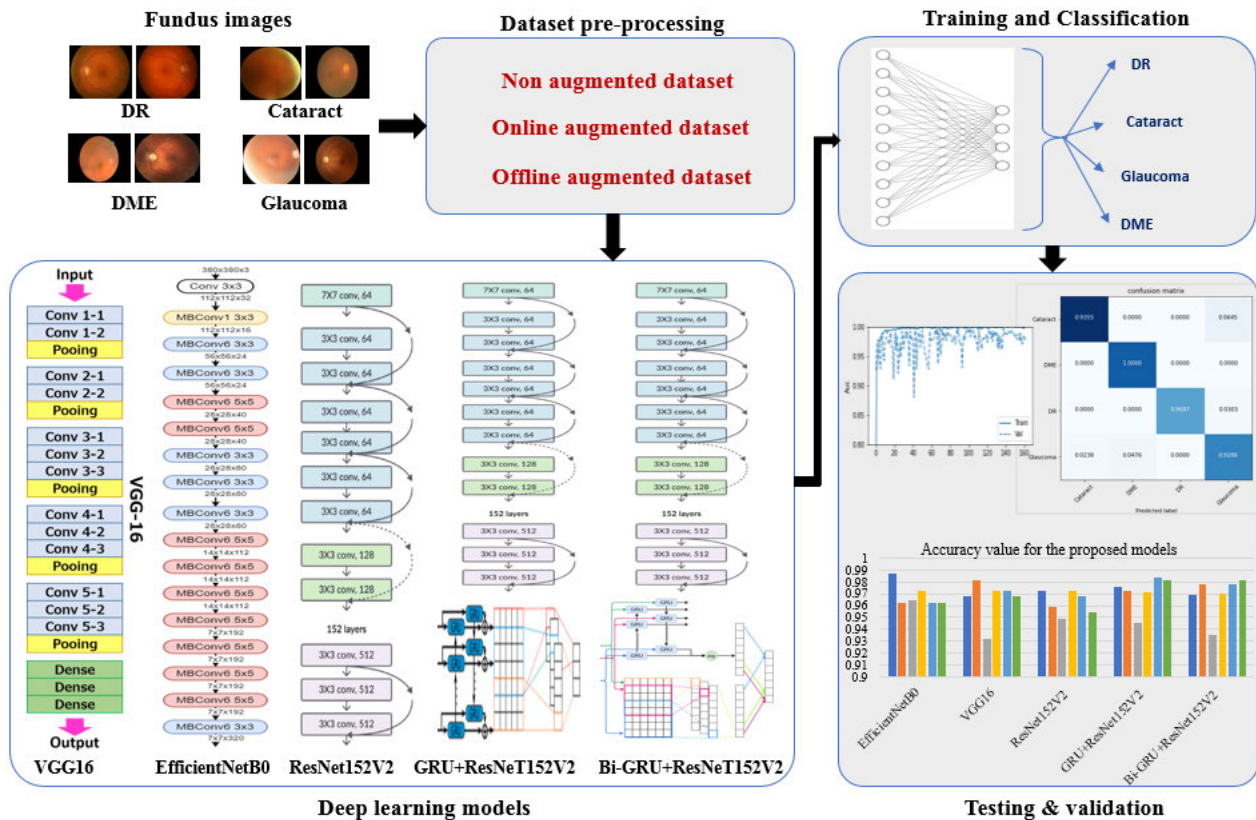
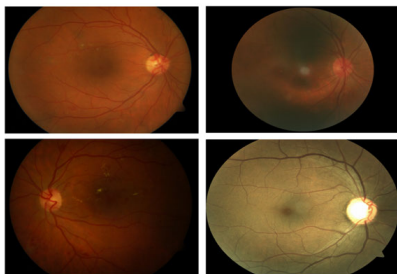**FIGURE 1.** The block diagram of our proposed DeepDiabetic framework.

FIGURE 2. Samples from the original dataset of the fundus images.

images will be applied as input for the pre-processing data stage; in our model, different pre-processing steps in input images, including resizing, normalizing, and converting the images into arrays for employing them as input for the following stage of the model. Then, we will randomly split our dataset (1228 images) for training the deep learning model to ensure the variety of the images into two parts, 70% (about 858 images) and 30% (370 images) for training and validation, respectively.

Pre-processing is widely used to prepare a dataset to meet the requirements of a deep learning model. In our model, different pre-processing steps in input images are applied. The first process is to split the dataset between the four classes: DR, Cataract, Glaucoma, and DME. TensorFlow

does not support some image formats such as (JPEG, PNG, GIF, BMP, and TIFF). The extension of those image files is indicative only and does not execute anything on the image's content. This research replaces all the image extensions to the JPG format. Then, these images will be applied as input for the pre-processing data stage. They were resizing images to 224×224. after reading and having a batch size equal to 32. Since the pipeline processes batches of images that all have the same size, this must be provided. They were providing true value for shuffling whether to shuffle the data. They were adding optional random seeds for shuffling and transformations equal to 100. Then, normalizing and converting the images into arrays for employing them as input for the following stage of the model. The label_mode is categorial are a float 32 tensor of shape (batch_size, num_classes), representing a one-hot encoding of the class index. Lastly, we randomly split our dataset for training the deep learning model into two parts: 70% (858 images) and 30% (370 images) for training and validation, respectively, to guarantee that the images are diverse.

3) DATASET AUGMENTATION
Due to the limitations of the available dataset of diabetic eye disease, this research used data augmentation techniques that increased the sample size of each class with affine-altered

TABLE 4. A summary of the augmentation types with their feature value.

Augmentation type.	Feature value.
Rotation	35
Width shift	0.1
Height shift	0.1
Shear	0.1
Brightness	(0.25, 1.25)
Horizontal flip	True
Vertical flip	True
Fill mode	Nearest
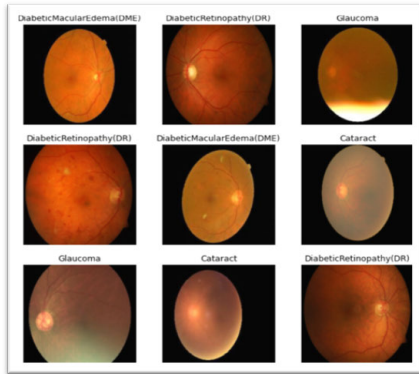
FIGURE 3. Samples of fundus images after augmentation.

TABLE 5. The proposed EfficientNetB0 architecture.

Layer (type)	Output Shape	Parameters
efficientnetb0 (Functional)	(None, 7, 7, 1280)	14714688
reshape (Reshape)	(None, 7, 7, 1280)	0
average_pooling2d		
(Global-Average-Pooling2d)	(None, 2, 2, 1280)	0
dropout_1 (Dropout)	(None, 1280)	0
dense_1 (Dense)	(None, 4)	5124
Total parameters: 4,054,69		
Trainable parameters: 4,012,672		
Non-trainable parameters:42,023		

TABLE 6. The proposed VGG16 architecture.

Layer (type)	Output Shape	Parameters
vgg16 (Functional)	(None, 7, 7, 512)	4049571
reshape (Reshape)	(None, 7, 7, 512)	0
average_pooling2d		
(Global-Average-Pooling2d)	(None, 2, 2, 512)	0
dropout_1 (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 4)	2052
Total parameters: 14,716,740		
Trainable parameters: 14,716,740		
Non-trainable parameters:0		

images and provided the models with more learnable features. Image data augmentation can be achieved in two different ways [9]: (1) Online augmentation - it is also referred to as on-the-fly augmentation - which is processed by performing transformations on the mini-batches that would be fed to the model during training. (2) offline augmentation, which is processed by performing the transformations on the images and saving the results in memory or on the desk. Thereby, increasing the size of the dataset by a factor equal to the number of transformations performed. In our work, we will use three methods based on three kinds of datasets:

- Method 1: Non-augmented dataset
  This method uses the original collected dataset mentioned in the Data Collection III-A1 which consists of 1228 images without any augmentation.
- Method 2: Online augmented dataset
  This method applies the augmentation during the model training. This means that in each epoch the model is fed by a randomly selected batch of the original dataset and then the transformations are executed online. Moreover, for each epoch, the images fed to the model are different based on the transformations applied.
- Method 3: Offline-augmented dataset
  This method applies the augmentation to the original dataset before it is used in the model. As explained before, the original dataset is randomly split into two sets: the training set and the validation set. Here, we perform the augmentation only on the training set of images, which is about 858 images. We used six different transformations on each image (in addition

to the original image) to result in a total of 6006 images.

The data augmentation will be executed utilizing the preprocessing layers of Keras; the specified values for augmentation will generate new variations of the original images without affecting their essential characteristics. Therefore, this research uses subtle transformations that prevent such a problem. Table 4 illustrates the augmentation types along with their feature value that will be used in image data augmentation. Figure 3 shows samples of fundus images after augmentation.

B. DEEP LEARNING PROPOSED METHODS

In this paper, several supervised deep-learning methods are employed to develop the proposed classification models. They aim to examine their performance in detecting the four considered diabetic eye diseases and ending the best of them. Three pre-trained models consisting of a CNN are VGG16, EfficientNetB0, and ResNet152V2. Recurrent neural network (RNN) combined gated recurrent unit (GRU) and bidirectional gated recurrent unit (Bi-GRU) models are also available. In the following sections, we provide further information on each of the five models we've developed.

1) EFFICIENTNETB0 DEEP MODEL

The EfficientNetB0 has been selected to achieve accurate classifications and cost-effectiveness. A reshape layer, two dimensions of global average pooling, a dropout layer, and a dense layer with a Softmax activation function were used to classify the data into their appropriate categories. The proposed architecture of the EfficientNetB0 model is shown in Figure 4, and further information regarding the architecture of the model can be found in Table 5. The total
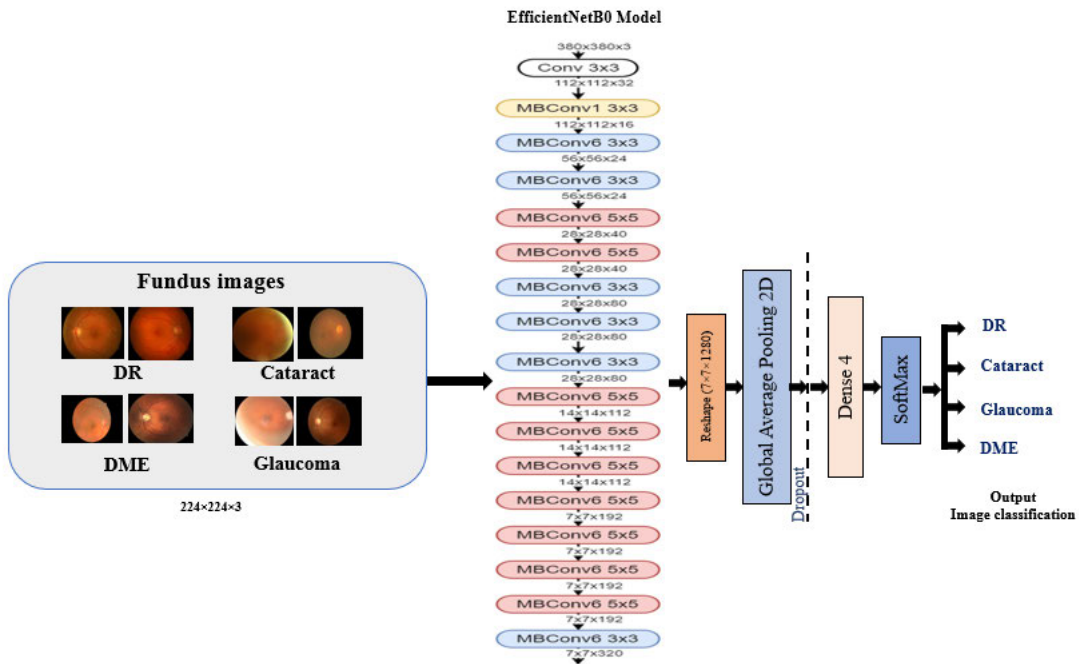
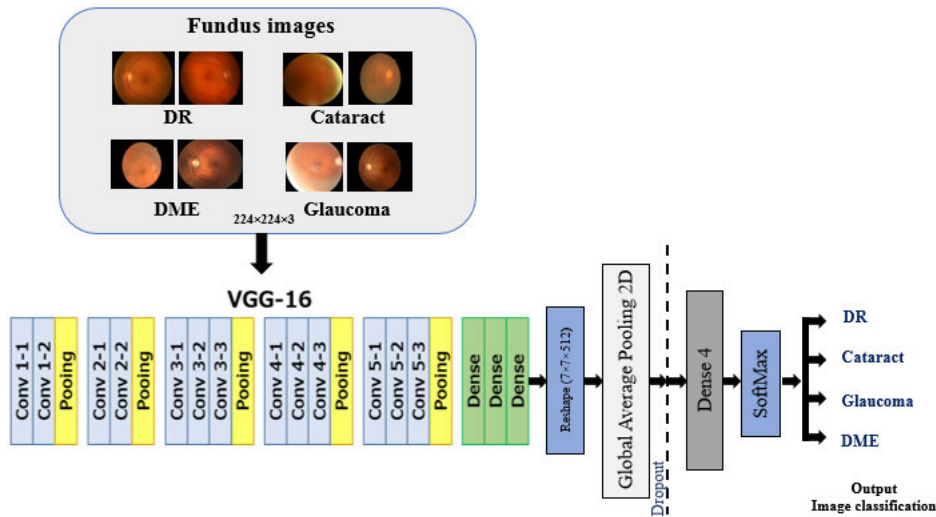**FIGURE 4.** EfficientNetB0 proposed model architecture.

FIGURE 5. VGG16 proposed model architecture.

number of parameters for the EfficientNetB0 is 4,054,695, and these parameters are divided into two categories: the trainable parameters, which have a value of 4,012,672, and the non-trainable parameters, which have a value of 42,023, respectively.

2) VGG16 DEEP MODEL
VGG16 was chosen because it is the most efficient model for image classification and has the simplest feature extraction. Figure 5 illustrates the VGG16 proposed model architecture.

Followed by a layer that does reshaping, two dimensions of global average pooling, a layer that does dropouts, and a layer that does dense layering. Table 6 provides further information regarding the design of the model, which has a total of 14,716,740 trainable parameters.

3) RESNET152V2 DEEP MODEL
For this study, we used the ResNet152V2 model since it provides both high precision and processing speed. The architecture of the model comprises the ResNet152V2 model,
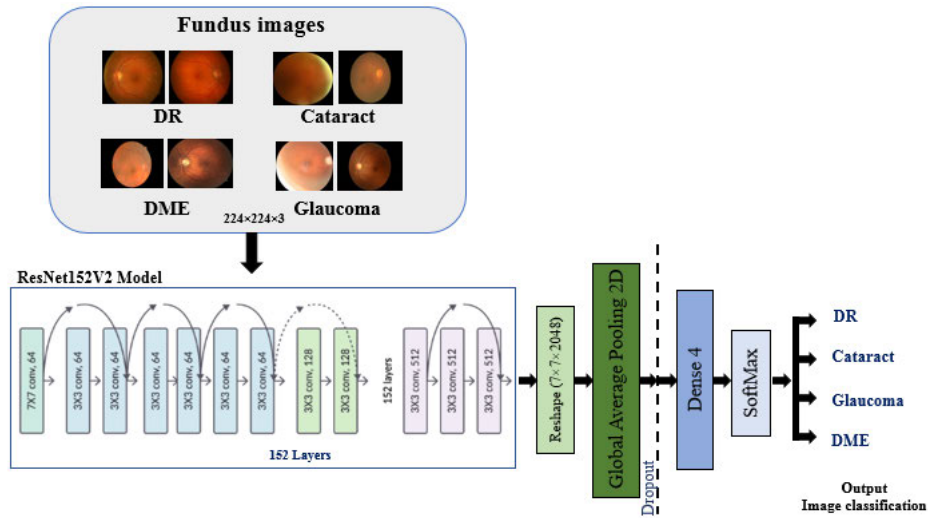
FIGURE 6. ResNet152V2 proposed model architecture.

TABLE 7. The proposed ResNet152V2 architecture.

Layer (type)	Output Shape	Parameters
Resnet152v2	(None, 7, 7, 2048)	58331648
reshape (Reshape)	(None, 7, 7, 2048)	0
average_pooling2d (Global-Average-Pooling2d)	(None, 2, 2, 512)	0
dropout_1 (Dropout)	(None, 2048)	0
dense_1 (Dense)	(None, 4)	8196
Total params: 58,339,844		
Trainable params: 58,196,100		
Non-trainable params: 143,744		

TABLE 8. The proposed GRU+ResNet152V2 architecture.

Layer (type)	Output Shape	Parameters
Resnet152v2	(None, 7, 7, 2048)	58331648
reshape (Reshape)	(None, 7, 7, 2048)	0
Time_distributed_7	(None, 7, 14336)	0
gru_6 (GRU)	(None, 256)	11208192
dense (Dense)	(None, 128)	32896
dropout_4 (Dropout)	(None, 128)	0
dense_7 (Dense)	(None, 4)	516
Total params: 69,573,353		
Trainable params: 69,429,508		
Non-trainable params: 143,744		

which is then followed by a reshape layer, a 2D of global average pooling, a dropout layer, and a dense layer with a Softmax activation function to categorize the image into its appropriate class. Table 7 and Figure 6 both contain more information regarding the architecture of the model. The total number of parameters for the ResNet152V2 is 58,339,844, and these parameters are split into two categories: trainable parameters and non-trainable parameters, with the former category containing 58,196,100 parameters and the latter category containing 143,744 parameters.

TABLE 9. Proposed Bi-GRU+ResNet152V2 architecture.

Layer (type)	Output Shape	Parameters
Resnet152v2	(None, 7, 7, 2048)	58331648
reshape (Reshape)	(None, 7, 7, 2048)	0
Time_distributed_7	(None, 7, 14336)	0
bidirectional GRU	(None, 512)	22416384
dense (Dense)	(None, 128)	65664
dropout (Dropout)	(None, 128)	0
dense (Dense)	(None, 4)	516
Total params: 80,814,212		
Trainable params: 80,670,468		
Non-trainable params: 143,744		

4) RESNET152V2 AND GRU DEEP MODEL

As illustrated in Figure 7, The sequential model, which comprises ResNet152V2 followed by GRU, is used to achieve the study's objective. The main advantage of GRU is that it permits the long-term retention of information that is not essential to the forecast without getting rid of it. The model consists of the ResNet152V2 layer, followed by a reshape layer, a GRU layer with 256 units, a flattened layer, a dense layer with 128 neurons, a dropout layer, and a dense layer, all of which use the Softmax activation function to classify the image as belonging to one of our four disease categories. Table 8 provides more information regarding the architecture. The ResNet152V2 has a total of 69,573,252 parameters, which are divided into two categories: trainable parameters make up 69,429,508 of the parameters, and non-trainable parameters make up 143,744 of the parameters.

5) RESNET152V2 AND BI-GRU DEEP MODEL

The sequential model of ResNet152V2, followed by bi-GRU, is the last proposed deep learning model in this research. The model architecture is described in detail in Figure 8. This model uses a SoftMax activation function to classify the image into one of our four disease categories. It starts
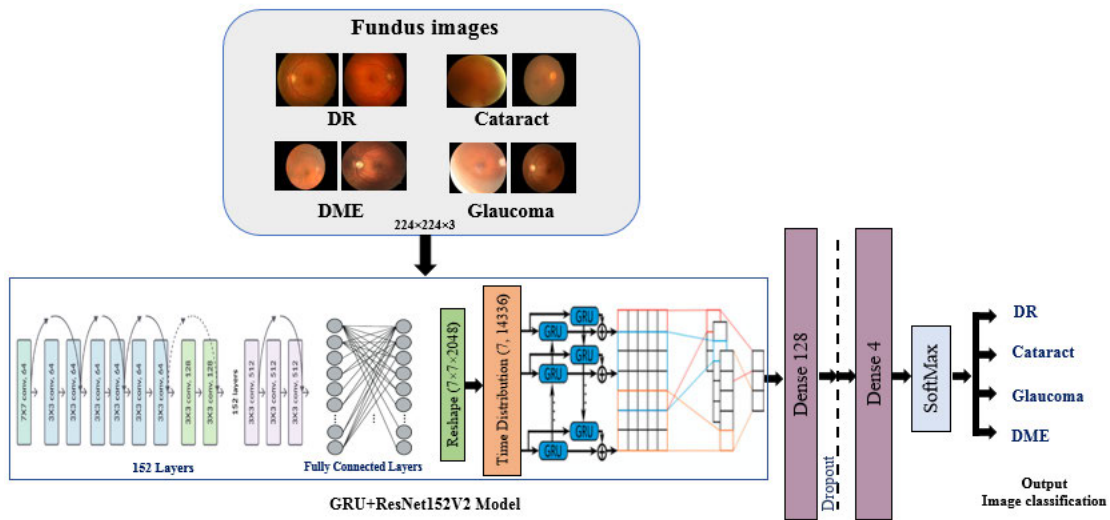
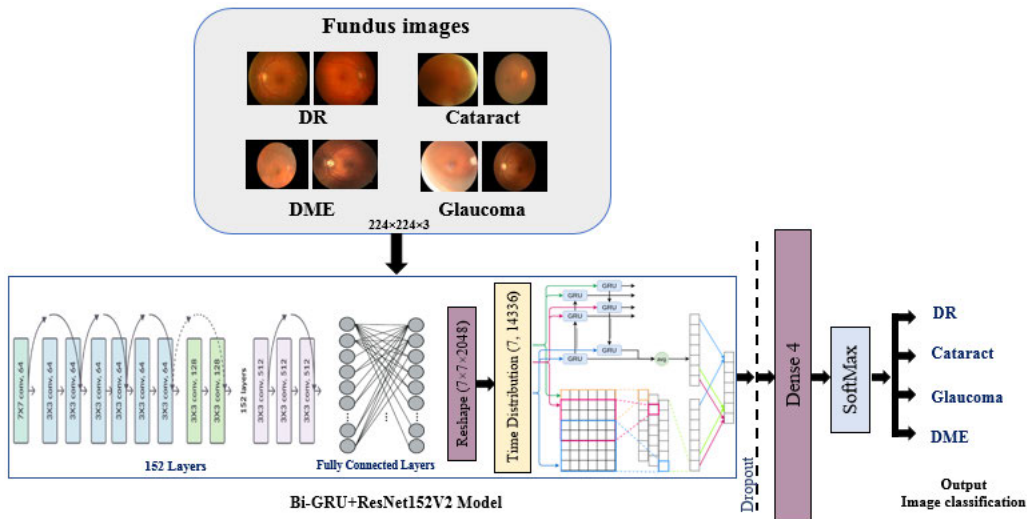**FIGURE 7.** GRU+ResNet152V2 proposed model architecture.

FIGURE 8. Bi-GRU+ResNet152V2 proposed model architecture.

with a ResNet152V2 and continues with a reshape layer, a bi-GRU layer with 512 units, a dense layer, a dropout layer, and another dense layer. The structure of the model is presented in Table 9. The overall number of parameters is 80,814,212, and there are two categories of parameters: trainable parameters make up 80,670,468 of the totals, while non-trainable parameters make up 143,744 of the totals.

IV. RESULTS
A. TRAINING PARAMETERS
Using the same parameters and criteria, each CNN model was trained and assessed independently in this research project. We needed to increase the overall amount of data and epochs to adequately train deep learning models. We trained the models on different epochs and batch size values to

accomplish the best results. Our models were trained and validated with the use of an optimizer and appropriate fit functions. During training, each model completed approximately 300 epochs, and the batch size was set to 32. The results were acquired by including the performance metric equations into the outcomes of the resultant validation data, and the reported results represent the maximum validation values that were attained. The Adam optimizer [10] was utilized to get the best results for our proposed models. Table 10 displays the learning rate (LR) values and optimizers used for all models.

B. EARLY STOPPING (CALLBACK)
An important concern with training neural networks is selecting the number of training epochs. Too many epochs

TABLE 10. The training parameters of the Models: learning rate value and optimizer.

Model	Optimizer	Learning Rate
EfficientNetB0	Adam	0.00001
VGG16		0.00001
ResNet152V2		0.00001
GRU+ResNet152V2		0.00001
Bi-GRU+ResNet152V2		0.00001

can cause the training dataset to overfit, while too few can result in an underfit model. There are many techniques on how to set the number of epochs like early stopping and model checkpoint. We use the early stopping strategy in our work, which involves setting a very large number of epochs and turning off the training when the improvement over the subsequent epochs is not satisfactory and meets our expectations.

Early stopping is a technique for specifying an arbitrarily large number of training epochs and stopping training once the model performance on a hold-out validation dataset stops improving. Keras has a callback named EarlyStopping that allows you to stop training early. Callbacks allow the execution of the code automatically and interact with the training model process. At the end of each epoch, the loss function of the validation dataset chosen to be optimized for models is calculated. The first sign of no further improvement is not always the optimum time to quit training. This is because the model may reach a point of no improvement or perhaps worsen slightly before improving significantly. We set the "patience" argument to 60 in this experiment to delay the number of epochs we would like to see no improvement.

The execution of our experiment is divided into three main cases: non-augmented dataset, online augmented dataset, and offline augmented dataset. For each case, we will experiment with two different execution parameters: First assigned to 300 epochs and second with a callback. So, this will be conducted through the five deep learning models: VGG16, EfficientNetB0, ResNet152V2, GRU+ResNet152V2, and Bi-GRU+ResNet152V2.

C. PERFORMANCE METRICS
The performance of diabetic eye disease classification models was evaluated using several metrics: accuracy, loss, recall, precision, specificity, and AUC. For each model, a confusion matrix is then introduced. Accuracy, in Eq. (1), is the number of instances that can be accurately predicted based on the overall number of instances.

$$Accuracy(ACC) = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} \quad (1)$$

where TP and TN stand for the true positive and true negative parameters, respectively, and where FP and FN stand for the false positive and false negative values, respectively. Recall, given in Eq. (2), is the number of samples that were positive in addition to the number of samples that were anticipated to

TABLE 11. Confusion matrix explanations.

			Predicted		
		Cataract	DME	DR	Glaucoma
Actual	Cataract	Pnn	Ppn	Pcn	Pln
	DME	Pnp	Ppp	Pcp	Plp
	DR	Pnc	Ppc	Pcc	Plc
	Glaucoma	Pnl	Ppl	Pcl	Pll

Pcc: DR class were correctly classified as DR
Ppc: DR class were incorrectly classified as DME
Pnc: DR class were incorrectly classified as Cataract
Plc: DR class were incorrectly classified as Glaucoma
Pcp: DME class were incorrectly classified as DR
Ppp: DME class were correctly classified as DME
Pnp: DME class were incorrectly classified as Cataract
Plp: DME class were incorrectly classified as Glaucoma
Pcn: Cataract class were incorrectly classified as DR
Ppn: Cataract class were incorrectly classified as DME
Pnn: Cataract class were correctly classified as Cataract
Pln: Cataract class were incorrectly classified as Glaucoma
Pcl: Glaucoma class were incorrectly classified as DR
Ppl: Glaucoma class were incorrectly classified as DME
Pnl: Glaucoma class were incorrectly classified as Cataract
Pll: Glaucoma class were correctly classified as Glaucoma

be positive out of the total number of samples.

$$Recall(Sensitivity) = \frac{Tp}{Tp + Fn} \quad (2)$$

Eq. (3) shows the precision, also called the positive predictive value [11], which is the number of samples that were both actually and predicted to be positive out of the total number of samples predicted to be positive.

$$Precision = \frac{Tp}{Tp + Fp} \quad (3)$$

Moreover, recent studies support the use of confusion matrix analysis in model validation [11] since it is robust in categorizing data relationships and any distribution. It provides extra information on illustrating the classification models. To analyze our models using a confusion matrix, we have to understand how it is structured and we need to define all variables and parameters that can be extracted from Table 11.

D. MULTI-CLASSIFICATION DEEP LEARNING MODEL RESULTS
To determine the important findings, we conducted a large-scale experiment and collected several outputs that matched the performance indicators. As a result, we presented the findings for each model experiment in tables. To determine the optimal derivative value, we experimented with six different parameter approaches. Columns on the tables illustrate the performance of the classifier models in the following metrics: validation loss, validation accuracy, validation precision, validation recall, and validation AUC. The bold font is used to select the highest value in each column. The rows in each table illustrate the type of each experiment. The rows in the first column are divided to present the data method used in execution into three main categories: non-augmented, online augmented, and offline

TABLE 12. True positives, true negatives, false positives, and false negatives variables definitions.

| True Positives:	True Negatives:
TP (Cataract): Pnn	TN (Cataract): Ppc + Ppl + Pcl + Plc + Plp +Pcp + Pcc + Pll + Ppp
TP (DME): Ppp	TN (DME): Pnl + Pnc + Pln + Pcn + Pnn +Pcl + Plc + Pcc + Pll
TP (DR): Pcc	TN (DR): Pll + Ppl + Pnl + Pnp + Ppp + Plp + Pln + Ppn +Pnn
TP (Glaucoma): Pll	TN (Glaucoma): Pnn + Ppn + Pcn + Pnp + Ppp + Pcp + Pnc + Ppc + Pcc
False Positives:	**False Negatives:**
FP (Cataract): Pnl + Pnc + Pnp	FN (Cataract): Pln + Pcn + Ppn
FP (DME): Ppl + Ppc + Ppn	FN (DME): Plp +Pcp + Pnp
FP (DR): Pcn + Pcp + Pcl	FN (DR): Pnc + Ppc + Plc
FP (Glaucoma): Plc + Plp + Pln	FN (Glaucoma): Pcl +Ppl + Pnl
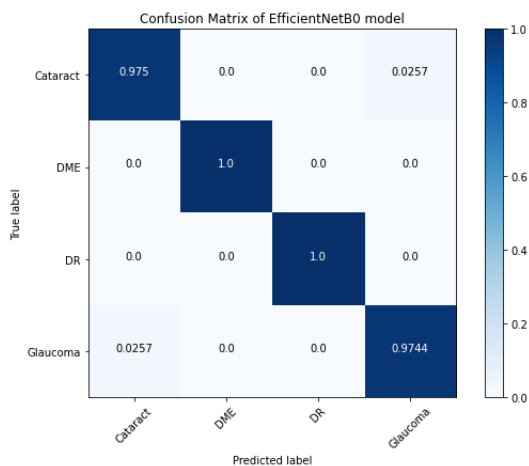
FIGURE 9. Confusion matrix of the proposal EfficientNetB0 model.

FIGURE 10. Loss, AUC, precision, recall, and accuracy between the training and validation phases, with number of epochs for the EfficientNetB0 model.

augmented. Furthermore, each of those categories is divided into two rows the first row for the results of the early stopped method, equal to 60 of the patients of loss improvement, and the second row for the results of 300 epochs.

The most popular technique for evaluating classification errors is the confusion matrix. Table 12 explains the confusion matrix and we used it to generate the confusion matrix for each of the proposed models. For both the training and validation stages we report the loss of AUC, precision, recall, and accuracy between the training and validation phases in two ways: one is by using several 300 epochs, and the second with using a callback function for all the proposal models.

1) EFFICIENTNETB0 DEEP MODEL

The performance measures of the EfficientNetB0 model using the three different data augmentation methods are illustrated in Table 13. In addition to the non-augmented images, online and offline augmented images are produced using three different ways of image data augmentation. Additionally, both with and without callbacks are taken into account for each method. the bolded values showed the higher values. The non-augmented dataset achieved the highest result, 0.9876, in accuracy, precision, and recall performance matrices. The best value of loss matrices of the EfficientNetB0 model was obtained from the online
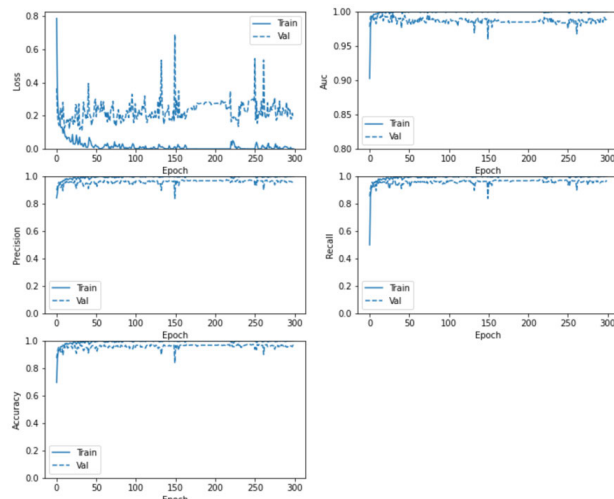
augmented data method with the callback function, 0.1044, which stopped training execution after 172 epochs. Consequently, offline augmented data seems to have lower results for both methods of the epochs. The confusion matrix in Figure 9 shows that the EfficientNetB0 model successfully classifies the four patient statuses (DR, DME, Glaucoma, and Cataract) with the highest balance to the DMR and DME images, then Cataract (0.9750), followed by Glaucoma (0.9744). This result confirms that the classification is executed correctly for the four classes. Furthermore, Figure 10 displays the loss, AUC, precision, recall, and accuracy between the training and validation phases with the number of epochs for the EfficientNetB0 model with online augmented.
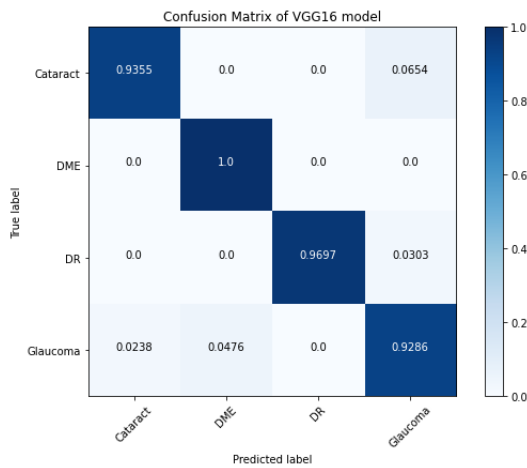
2) VGG16 DEEP MODEL

Performance metrics for the VGG16 model employing the three distinct data augmentation techniques are shown in Table 14. As the EfficientNetB0 model, the non-augmented dataset with the callback achieved the highest result, 0,9811, in accuracy, precision, and recall performance matrices. The best value of loss matrices of the VGG16 model was obtained from the online augmented data method with the callback function (0.0889). Accordingly, VGG16

TABLE 13. Performance measures of EfficientNetB0 model using different methods: online and offline augmented images, as well as non-augmented images. Also, each method with 300 epochs and no callback is taken into account.

Methods		Accuracy	Loss	Precision	Recall	AUC
NON-AUGMENTED	Callback	0.9622	0.1198	0.9622	0.9622	0.9973
	300 epochs	**0.9876**	0.1435	**0.9876**	**0.9876**	**0.9977**
ONLINE AUGMENTED	Callback	0.9730	**0.1044**	0.9730	0.9730	0.9943
	300 epochs	0.9649	0.1931	0.9649	0.9649	0.9869
OFFLINE AUGMENTED	Callback	0.9622	0.1632	0.9622	0.9622	0.9909
	300 epochs	0.9622	0.1632	0.9622	0.9622	0.9909

TABLE 14. Performance measures of VGG16 model using different methods: Online and offline augmented images, as well as non-augmented images. In addition, for each method with 300 epochs and without callback are considered.

Methods		Loss	Accuracy	Precision	Recall	AUC
NON-AUGMENTED	Callback	0.0904	**0.9811**	**0.9811**	**0.9811**	0.9940
	300 epochs	0.3356	0.9676	0.9676	0.9676	0.9854
ONLINE AUGMENTED	Callback	**0.0889**	0.9730	0.9730	0.9730	0.9973
	300 epochs	0.3713	0.9324	0.9324	0.9324	0.9779
OFFLINE AUGMENTED	Callback	0.0918	0.9676	0.9728	0.9649	**0.9984**
	300 epochs	0.5073	0.9730	0.9730	0.9730	0.9838

FIGURE 11. Confusion matrix of the proposal VGG16 model.

FIGURE 12. Loss, AUC, precision, recall, and accuracy between the training and validation phases, with number of epochs for the VGG16 model.

offline augmented data seems to have lower results in loss, accuracy, precision, and recall performance matrices, despite it acquiring the highest AUC value with 0.9984. The confusion matrix for the VGG16 model can be seen in Figure 11. This matrix shows that the model can categorize all four patient states (DR, DME, Glaucoma, and Cataract). We find that the DME images have the highest ratio, then DR (0.9697), followed by Cataract (0.9355), and finally Glaucoma (0.9286). These results are less impressive than those produced by the EfficientNetB0 model, which had higher DR values. Furthermore, Figure 12 illustrates the loss, AUC, precision, recall, and accuracy of the model with an early stopped function when comparing the training phase to the validation phase. The graphs show that the model stopped training after 160 epochs, which means there is no improvement after epoch 100.

3) RESNET152V2 DEEP MODEL
The performance measures of the ResNet152V2 model using the three different data augmentation methods are illustrated

in Table 15. The non-augmented and online augmented method of the ResNet152V2 model obtained similar results, 0.9730 for accuracy, precision, and recall performance matrices. The difference between them is achieving the same result faster using the online enhanced method. In contrast, the values of loss and AUC, 0.1143 and 0.9978, respectively, achieved the best value for the ResNet152V2 model in the augmented offline method. The best value of loss and AUC matrices of the ResNet152V2 model was obtained from the offline augmented data method with the callback function. Likewise, in Figure 14, the values of loss, AUC, precision, recall, and accuracy between the training and validation phases of the ResNet152V2 model, with the number of epochs, are displayed. Consistently, the ResNet152V2 model confusion matrix is presented in Figure 13. The Figure indicates the successful classification of the four patient statuses (DR, DME, Glaucoma, and Cataract), starting with the DR

TABLE 15. Performance measures of ResNet152V2 model using different methods: Online and offline augmented images, and non-augmented images. Additionally, every procedure with 300 epochs and no callback is considered.

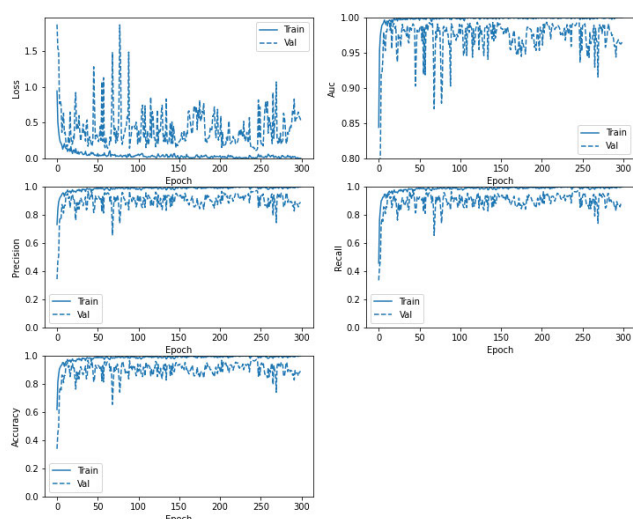| Methods		Loss	Accuracy	Precision	Recall	AUC
NON-AUGMENTED	Callback	0.1266	0.9595	0.9595	0.9595	0.9944
	300 epochs	0.5621	**0.9730**	**0.9730**	**0.9730**	0.9854
ONLINE AUGMENTED	Callback	0.1246	**0.9730**	**0.9730**	**0.9730**	0.9960
	300 epochs	0.1786	0.9486	0.9485	0.9459	0.9934
OFFLINE AUGMENTED	Callback	**0.1143**	0.9541	0.9590	0.9486	**0.9978**
	300 epochs	0.4019	0.9676	0.9676	0.9676	0.9822
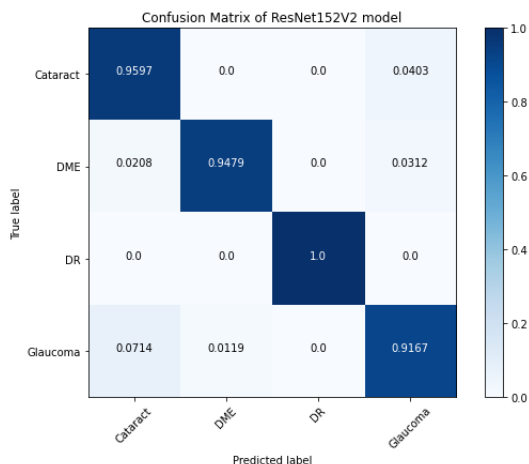
FIGURE 13. Confusion matrix of the proposal ResNet152V2 model.
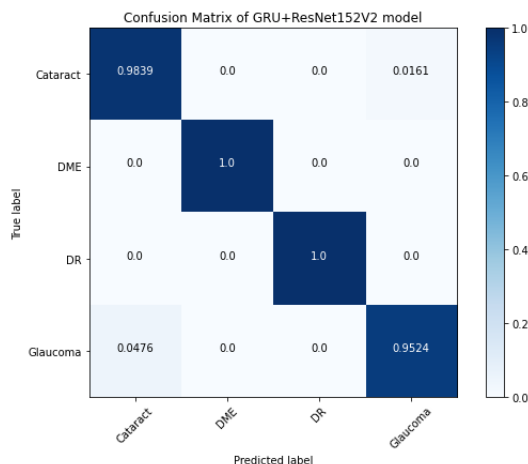
FIGURE 15. Confusion matrix of the proposal GRU+ResNet152V2 model.
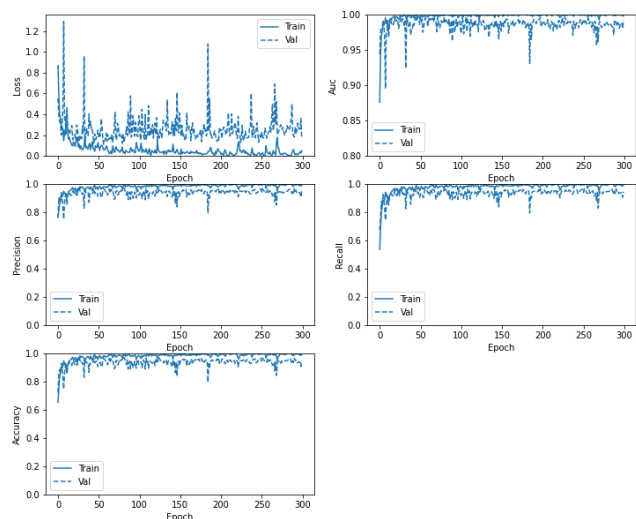
FIGURE 14. Loss, AUC, precision, recall, and accuracy between the training and validation phases, with number of epochs for the ResNet152V2 model.

FIGURE 16. Loss, AUC, precision, recall, and accuracy between the training and validation phases, with number of epochs for the GRU+ResNet152V2 model.

images, which have the highest ratio, then Cataract, followed by DME, and lastly Glaucoma.

4) RESNET152V2 AND GRU DEEP MODEL
The GRU+ResNet152V2 model showed that the performance of ResNet152V2 was better when combined with GRU, which is evident from the higher values shown

in Table 16. The offline augmented data of the GRU+ ResNet152V2 model had the best values in the performance matrix, except the AUC value reached the highest value with the online augmented value. Furthermore, the GRU+ResNet152V2 with offline augmented method reached 0,9838 in accuracy, precision, and recall performance matrices. In addition, GRU+ResNet152V2 with the offline

TABLE 16. Performance measures of GRU+ResNet152V2 model using different methods: Online and offline augmented images, and non-augmented images. In addition, each method with 300 epochs and no callback is considered.

Method		Loss	Accuracy	Precision	Recall	AUC
NON-AUGMENTED	Callback	0.3172	0.9721	0.9721	0.9721	0.9874
	300 epochs	0.2846	0.9757	0.9757	0.9757	0.9855
ONLINE AUGMENTED	Callback	0.1221	0.9714	0.9539	0.9514	**0.9972**
	300 epochs	0.2478	0.9459	0.9508	0.9405	0.9863
OFFLINE AUGMENTED	Callback	**0.0679**	0.9811	0.9811	0.9811	0.9961
	300 epochs	0.1623	**0.9838**	**0.9838**	**0.9838**	0.9908
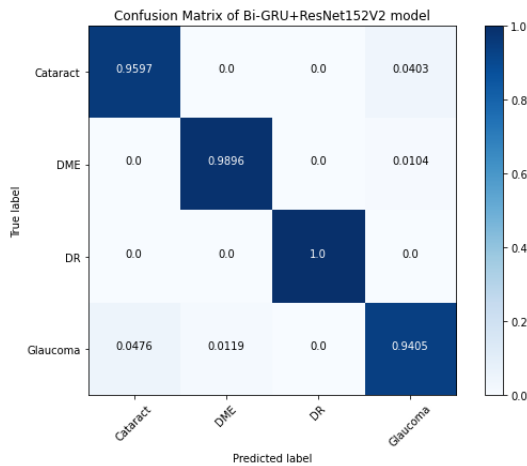
FIGURE 17. Confusion matrix of the proposal Bi-GRU+ResNet152V2 model.

FIGURE 18. Loss, AUC, precision, recall, and accuracy between the training and validation phases, with number of epochs for the Bi-GRU+ResNet152V2 model.

augmented method scores 0.0679, which is better than the VGG16 and ResNet152V2 models. While the highest AUC value of GRU+ResNet152V2 was with the online augmented method, it obtained 0.9972. Similarly, in Figure 15, the confusion matrix of the GRU+ResNet152V2 model shows the classification of the four patient statuses ((DR, DME, Glaucoma, and Cataract), with the highest ratio to the DME and DR images, then Cataract (0.9839), and finally Glaucoma (0.9542). Figure 16 depicts the epoch-by-epoch variation in loss, AUC, precision, recall, and accuracy for the ResNet152+GRU model throughout training and validation.

5) RESNET152V2 AND BI-GRU DEEP MODEL
The performance measures of the Bi-GRU+ResNet152V2 model using the three different data augmentation methods are described in Table 17. Offline augmentation gives the best method for the Bi-GRU+ResNet152V2 model in the performance matrix as the GRU+ResNet152V2 model. The offline augmented data mothed achieved the highest result, in all performance matrices, with the callback function, which stopped training execution after 107 epochs. This method achieved 0.0578, 0.9811, 0.9811, 0.9811, and 0.9980 For loss, accuracy, precision, recall, and AUC, respectively. The loss value (0.0578) achieved by the Bi-GRU+ResNet152V2 model was lower than that of the GRU+ResNet152V2 model. The Bi-GRU+ResNet152V2 model training results showed that the callback approach outperformed the other

augmentation strategies. Similarly, in Figure 18, loss, AUC, precision, recall, and accuracy between the training and validation phases of the Bi-GRU+ResNet152V2 model, with the number of epochs, are displayed.

Also, the Bi-GRU+ResNet152V2 model confusion matrix is presented in Figure 17. In the Figure, we can see the four diagnosed conditions (DR, DME, Glaucoma, and Cataract). The DR images had the highest ratio, followed by those of DR, DME, Cataract, and Glaucoma, in that order. Comparatively, these results are inferior to those achieved by the GRU+ResNet152V2 model but superior to those obtained by the ResNet152V2 model alone.

V. COMPARATIVE ANALYSIS AND DISCUSSION
This paper proposes five deep-learning models for diabetic eye disease detection (EfficientNetB0, VGG16, ResNet152V2, GRU+ResNet152V2, and Bi-GRU+ResNet152V2). These models are used to classify the most common diabetic eye diseases: DR, DME, Cataract, and Glaucoma. The various models that were suggested were evaluated based on their accuracy, recall, precision, loss, and AUC. The obtained results show that the EfficientNetB0 model gave the best classification performance (0.9876 in accuracy), followed by GRU+ResNet152V2, with 0.9838 in accuracy. Conversely, ResNet152V2 is the lowest of the

TABLE 17. Performance measures of Bi-GRU+ResNet152V2 model using different methods: Online and offline augmented images, and non-augmented images. Additionally, every procedure with 300 epochs and no callback is considered.

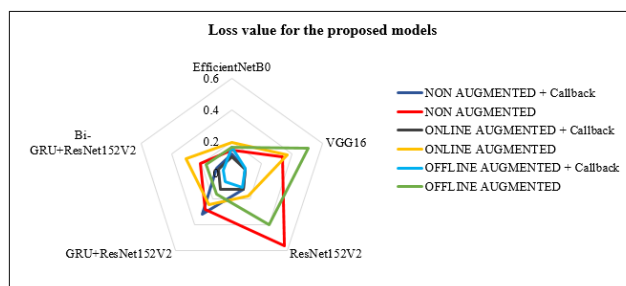| Method		Loss	Accuracy	Precision	Recall	AUC
NON-AUGMENTED	Callback	0.1009	0.9784	0.9784	0.9784	0.9872
	300 epochs	0.2062	0.9690	0.9720	0.9690	0.9872
ONLINE AUGMENTED	Callback	0.0879	0.9703	0.9703	0.9703	0.9975
	300 epochs	0.3009	0.9351	0.9351	0.9351	0.9777
OFFLINE AUGMENTED	Callback 60	**0.0578**	**0.9811**	**0.9811**	**0.9811**	**0.9980**
	300 epochs	0.1720	0.9784	0.9784	0.9784	0.9890
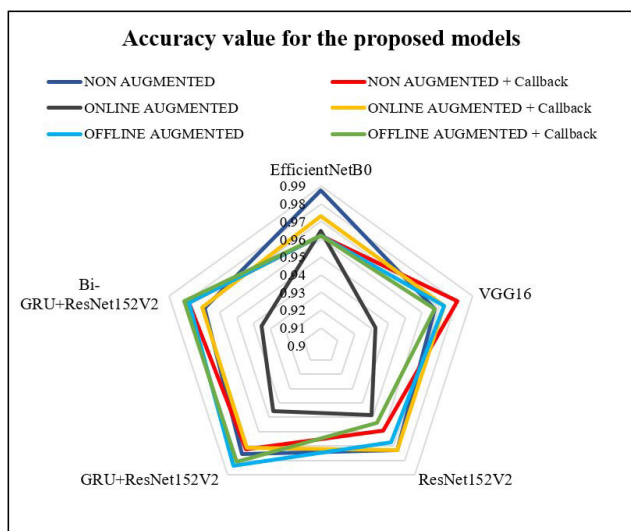
FIGURE 19. Accuracy measures for the proposed models (EfficientNetB0, VGG16, ResNet152V2, GRU+ResNet152V2, and Bi-GRU+ResNet152V2). Three different methods of image data augmentation are achieved: non-augmented images, and online and offline augmented images. Additionally, every procedure with 300 epochs and no callback is considered.
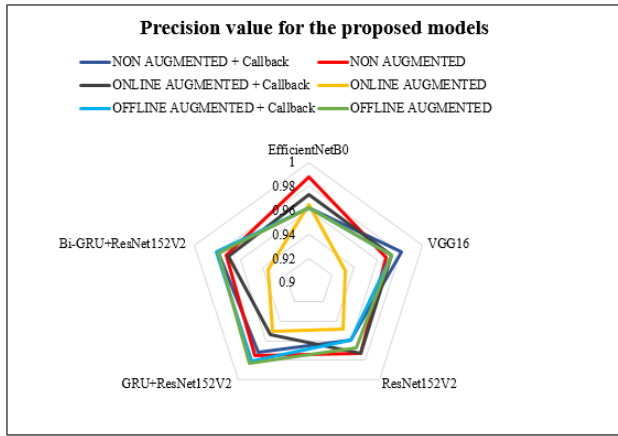

FIGURE 20. Loss measures for the proposed models (EfficientNetB0, VGG16, ResNet152V2, GRU+ResNet152V2, and Bi-GRU+ResNet152V2). Three different methods of image data augmentation are achieved: non-augmented images, and online and offline augmented images. Additionally, every procedure with 300 epochs and no callback is considered.

other architectures since this model has obtained the lowest accuracy, as illustrated in Figure 19. Furthermore, the VGG16 and the Bi-GRU+ResNet152V2 models reached the same value of 0.9811 in accuracy, precision, and recall performance matrices.

In this experiment, three different methods of image data augmentation are achieved in online and offline augmented besides the non-augmented images. In addition, for each method with and without a callback is considered. We discovered that offline augmented obtained better values with sequential models (GRU+ResNet152V2 and Bi-GRU+ResNet152V2) because this augmentation method increases the dataset size of the training set to six times the original size. At the same time, EfficientNetB0 and VGG16 models assumed the best values of the evaluation matrixes in the non-augmented data.

The loss value for GRU+ResNet152V2 was better, however, the AUC performance value for VGG16 was superior. Figure 20 showed the Bi-GRU+ResNet152V2 reached the most useful value of the loss metrics by contacting 0.0578. The Bi-GRU+ResNet152V2 with offline-augmented reached the lowest, which means the best value of the loss metrics by contacting 0.0578.
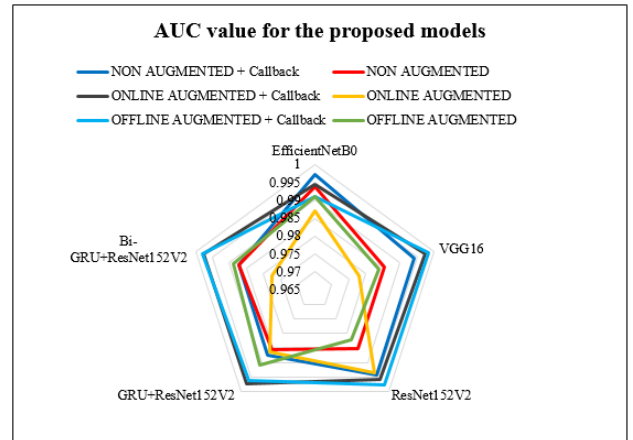
After assigning the callback method for the whole experiment implementation, 300 training epochs stopped training once the model performance on the loss matric validation dataset stopped improving for 60 more epochs. We found that the callback method performs more valuable to the augmented online technique for all five models. At the same time, we didn't notice the benefit of using the callback method with non-augmented and offline-augmented. In addition, the results of the Bi-GRU+ResNet152V2 model showed that callback performed more useful for all types of augmentation.

Figures 21 and 22 display the measurement of the precision and the recall for the following models: EfficientNetB0, VGG16, ResNet152V2, GRU+ResNet152V2, and Bi-GRU+ResNet152V2. The results of precision and recall are considerably equal in all the methods. Slight differences between precision and recall result in the offline augmented value with callback in all the models. Furthermore, the last figure, Figure 23, indicates all the details values of the AUC measurement for all these models. The best AUC measurements were conducted during the callbacks with offline augmented for three models, VGG16, ResNet152V2, and Bi-GRU+ResNet152V2.

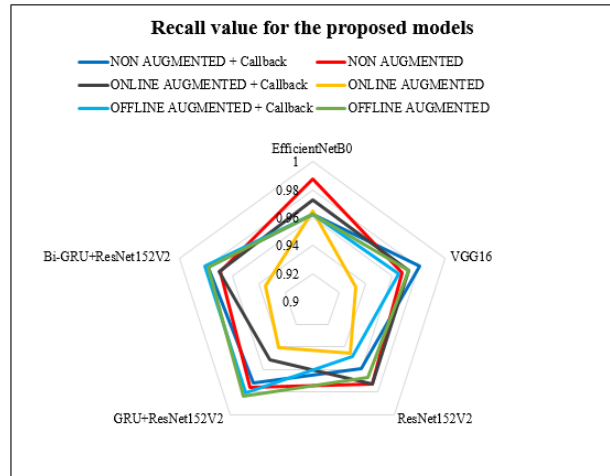According to the results of the accuracy, our deep learning models achieved higher than the previous studies. As a consequence of this, we recommend that the EfficientNetB0 model, which is based on fundus images, be utilized in the process of diagnosing the health status of diabetic patients concerning DR, DME, cataracts, and glaucoma. We have introduced some deep models, and their results, in the hopes

FIGURE 21. Precision measures for the proposed models (EfficientNetB0, VGG16, ResNet152V2, GRU+ResNet152V2, and Bi-GRU+ResNet152V2). Three different methods of image data augmentation are achieved: non-augmented images, and online and offline augmented images. Additionally, every procedure with 300 epochs and no callback is considered.

FIGURE 22. Recall measures for the proposed models (EfficientNetB0, VGG16, ResNet152V2, GRU+ResNet152V2, and Bi-GRU+ResNet152V2). Three different methods of image data augmentation are achieved: non-augmented images, and online and offline augmented images. Additionally, every procedure with 300 epochs and no callback is considered.

that these can be used as a starting point for constructing a system that can diagnose eye diseases based on images of the fundus.

As discussed in the literature review section, few methods have focused on classifying between three or four diabetic eye diseases. As illustrated in table 2, DR, DME, GC, and Cataract were the four complications of diabetes, although their classification scheme was rarely modified. By the end of 2021, we found Rubina Sarki et al. [5] presented a classification framework for those four diseases. However, they only applied one CNN model in their work, and their dataset showed class imbalance. They attained only 81.33% accuracy, which is still considered low compared to recent research in DL classifications. The other studies on the same table classified between three diabetic eye diseases

FIGURE 23. AUC measures for the proposed models (EfficientNetB0, VGG16, ResNet152V2, GRU+ResNet152V2, and Bi-GRU+ResNet152V2). Three different methods of image data augmentation are achieved: non-augmented images, and online and offline augmented images. Additionally, every procedure with 300 epochs and no callback is considered.

TABLE 18. Comparison of the performance evaluation metrics of the different models presented in previous studies and our proposed EfficientNetB0 model.

Models	Acc.	Pre.	F1	Rec.	AUC
Vgg16 [33]	88.3	98	91.03	85	-
Fast-RCNN [27]	95.2	95	96	96.2	96.3
RCNN-LSTM [12]	89.54	-	89.97	-	97
Vgg16 [13]	97.23	96.73	95.22	93.76	-
ResNet34 [13]	90.85	93.7	92.65	93.17	-
MobileNetV2 [13]	94.32	93.33	89.67	91.46	-
EfficientNet [13]	93.82	92.73	96.25	93.74	-
InceptionResNet [18]	-	94.38	94.28	94.58	96.08
EfficientNetB0 (ours)	98.76	98.86	98.76	98.66	99.77
VGG16 (ours)	98.11	98.01	98.11	98.21	99.4
ResNet152V2 (ours)	97.3	97.4	97.3	97.5	99.6
GRU+ResNet152V2 (ours)	98.38	98.28	98.38	98.48	99.08
Bi-GRU+ResNet152V2 (ours)	98.11	98.01	98.11	98.21	99.8

Acc.= Accuracy, Pre.= Precision, F1=F1_score, Rec.=Recall

Table 18 shows a comparison of the performance evaluation metrics of the different DL models presented in previous studies and our proposed five models (EfficientNetB0, VGG16, ResNet152V2, GRU+ResNet152V2, and Bi-GRU+ResNet152V2), as also represented in Figure 24. In this paper, we evaluated the proposed model in terms of accuracy, precision, f1-score, recall, and AUC. The accuracy reached by our EfficientNetB0 model is 98.76% in comparison with Vgg16 [33], Fast-RCNN [27], RCNN-LSTM [12], Vgg16 [13], ResNet34 [13], MobileNetV2 [13], and EfficientNet [13] which have 88.3%, 95.2%, 89.54%, 97.23%, 90.85%, 94.32%, and 93.82%, respectively. While the AUC achieved by our EfficientNetB0, Vgg16, ResNet152V2, GRU+ResNet152V2, and Bi-GRU+ResNet152V2 models are 99.77%, 99.4%, 99.6%, 99.08,%, and 99.8%, respectively, compared with Fast-RCNN [27], RCNN-LSTM [12], and InceptionResNet [18] which got 95.8%, 97%, and 96.08%, respectively. Similarly, the table shows the differences between the previous studies and ours in terms of precision, f1-score, and recall evaluation metrics. The results
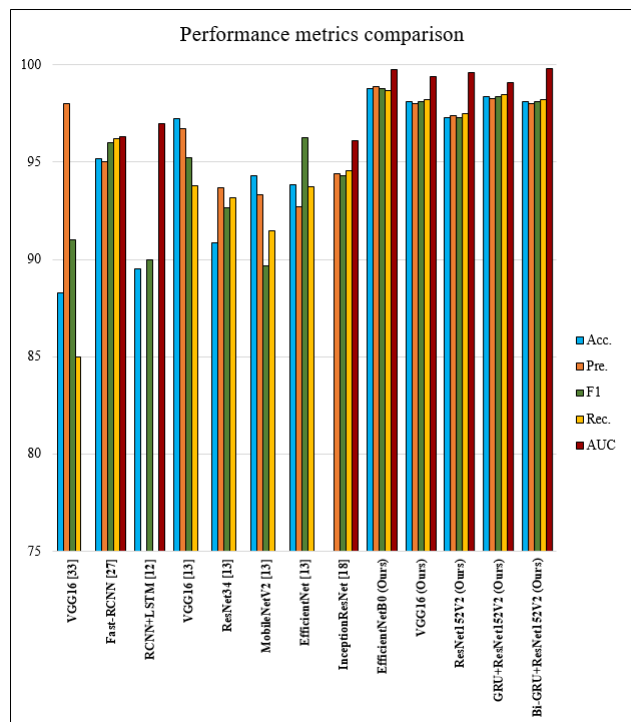
FIGURE 24. Comparison of the performance evaluation metrics: The accuracy, precision, F1-score, Recall, and AUC values between the different models presented in the previous studies [12], [13], [27], [33], [18], and our proposed five models.

show that our proposed EfficientNetB0 model outperforms all the existing models by a significant value, which demonstrates the effectiveness of our predictive model.

VI. CONCLUSION AND FUTURE WORK

In this research, a multi-classification deep learning model, called the DeepDiabetic framework was designed and evaluated for detecting DR, DME, Glaucoma, and Cataract from fundus images. It is important to correctly diagnose these diseases early to determine the proper treatment. However, it is challenging to accurately identify via fundus images, and even highly experienced ophthalmologists are prone to misdiagnose eye lesions.

To the best of our knowledge, there are no other GRU models in the literature that choose between DR, DME, cataract, and glaucoma. Five model architectures were offered in this research: EfficientNetB0, VGG16, ResNet152V2, GRU+ResNet152V2, and Bi-GRU+ResNet152V2. Most of the previous research focused on the classification and development of a single fundus disease independently; despite this research, the primary focus was to provide a multi-classification of the four diabetic eye diseases, which presents a big obstacle. This research employed a combination of datasets from various sources to examine the technique's robustness and adaptability in dealing with real-world cases.

The accuracy, precision, f1-score, recall, and AUC of the proposed models were measured in this paper. Our

EfficientNetB0 model achieves 98.76% accuracy, while Vgg16 [33], RCNN-LSTM [12], Vgg16 [13], and CNN [5] only achieve 88.33%, 89.54%, 97.23%, and 80.33% accuracy, respectively. When compared to Fast-RCNN [27], RCNN-LSTM [12], and InceptionResNet [18], our EfficientNetB0 model achieves a far higher AUC of 99.77%. The accuracy reached by our EfficientNetB0 model is 98.76% in comparison with Vgg16 [33], Fast-RCNN [27], RCNN-LSTM [12], Vgg16 [13], ResNet34 [13], MobileNetV2 [13], and EfficientNet [13] which have 88.3%, 95.2%, 89.54%, 97.23%, 90.85%, 94.32%, and 93.82%, respectively. As you can see, the AUC for our EfficientNetB0, Vgg16, ResNet152V2, GRU+ResNet152V2, and Bi-GRU+ResNet152V2 models is 99.77%, 99.47%, 99.6%, 99.8%, and 99.8%, respectively. For Fast-RCNN [27], it was 95.64%, for RCNN-LSTM [12], and for InceptionResNet [18], it was 96.8%. Similarly, the table shows the differences between the previous studies and ours in terms of precision, f1-score, and recall evaluation metrics. According to the outcomes, our suggested EfficientNetB0 model is significantly more accurate than the state-of-the-art models.

To the best of our knowledge, our five multi-classification models of the four diabetic diseases (DR, DME, Glaucoma, and Cataract) achieved the highest accuracy result in state-of-the-art. The EfficientNetB0 model achieved 98.76% accuracy, 98.76% recall, 98.76% precision, and 99.73% AUC based on fundus images. The EfficientNetB0 model did better than the other five proposed models in a large set of experiments and results using fundus images from a number of different sources.

In future work, understanding various features of deep neural networks and visualization is also a significant research field to increase clinical acceptance of deep learning models. We'll have to collect huge training datasets with tens of thousands of abnormal cases from other hospitals using various types of cameras in the future. So, more features will be added to help enhance accuracy and generalization.

More work can be done to make the suggested model work better by adding more images to the datasets that are used, training epochs, and using different GAN architectures and other deep learning techniques for classification and enhancement.

REFERENCES

[1] P. Vashist, N. Gupta, S. Singh, and R. Saxena, "Role of early screening for diabetic retinopathy in patients with diabetes mellitus: An overview," *Indian J. Community Med.*, vol. 36, no. 4, p. 247, 2011.

[2] B. Aljaddouh and D. Malathi, "Trends of using machine learning for detection and classification of respiratory diseases: Investigation and analysis," *Mater. Today, Proc.*, vol. 62, pp. 4651–4658, Jan. 2022.

[3] D. M. Ibrahim, N. M. Elshennawy, and A. M. Sarhan, "Deep-chest: Multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases," *Comput. Biol. Med.*, vol. 132, May 2021, Art. no. 104348.

[4] A. E. Minarno, M. H. C. Mandiri, Y. Azhar, F. Bimantoro, H. A. Nugroho, and Z. Ibrahim, "Classification of diabetic retinopathy disease using convolutional neural network," *Int. J. Informat. Visualizat.*, vol. 6, no. 1, pp. 12–18, 2022.

[5] R. Sarki, K. Ahmed, H. Wang, Y. Zhang, and K. Wang, "Convolutional neural network for multi-class classification of diabetic eye disease," *EAI Endorsed Trans. Scalable Inf. Syst.*, vol. 9, no. 4, p. e15, 2022.

[6] R. Sarki, K. Ahmed, H. Wang, and Y. Zhang, "Automatic detection of diabetic eye disease through deep learning using fundus images: A survey," *IEEE Access*, vol. 8, pp. 151133–151149, 2020.

[7] 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.

[8] F. Zia, I. Irum, N. N. Qadri, Y. Nam, K. Khurshid, M. Ali, I. Ashraf, and M. A. Khan, "A multilevel deep feature selection framework for diabetic retinopathy image classification," *Comput., Mater. Continua*, vol. 70, no. 2, pp. 2261–2276, 2022.

[9] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019.

[10] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[11] U. Özkaya, S. Öztürk, and M. Barstugan, "Coronavirus (COVID-19) classification using deep features fusion and ranking technique," in *Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach*. Berlin, Germany: Springer, 2020, pp. 281–295.

[12] F. Demir and B. Tasci, "An effective and robust approach based on R-CNN+LSTM model and NCAR feature selection for ophthalmological disease detection from fundus images," *J. Personalized Med.*, vol. 11, no. 12, p. 1276, Dec. 2021.

[13] N. M. Dipu, S. Alam Shohan, and K. M. A. Salam, "Ocular disease detection using advanced neural network based classification algorithms," *ASIAN J. Converg. Technol.*, vol. 7, no. 2, pp. 91–99, Aug. 2021.

[14] J.-H. Han, "Artificial intelligence in eye disease: Recent developments, applications, and surveys," *Diagnostics*, vol. 12, no. 8, p. 1927, Aug. 2022.

[15] N. Durga, "A systematic review on diabetic retinopathy and common eye diseases detection through deep learning techniques," *J. Positive School Psychol.*, vol. 6, no. 4, pp. 1905–1919, 2022.

[16] M. M. Butt, D. N. F. 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, Jul. 2022.

[17] S. H. Abbood, H. N. A. Hamed, M. S. M. Rahim, A. Rehman, T. Saba, and S. A. Bahaj, "Hybrid retinal image enhancement algorithm for diabetic retinopathy diagnostic using deep learning model," *IEEE Access*, vol. 10, pp. 73079–73086, 2022.

[18] A. Bhati, N. Gour, P. Khanna, and A. Ojha, "Discriminative kernel convolution network for multi-label ophthalmic disease detection on imbalanced fundus image dataset," 2022, *arXiv:2207.07918*.

[19] 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, p. 1966, Jun. 2022.

[20] A. Arora, S. Gupta, S. Singh, and J. Dubey, "Eye disease detection using transfer learning on VGG16," in *Proc. 3rd Int. Conf. Comput., Commun., Cyber-Security*, 2023, pp. 527–536.

[21] N. Badah, A. Algefes, A. AlArjani, and R. Mokni, "Automatic eye disease detection using machine learning and deep learning models," in *Pervasive Computing and Social Networking*. Berlin, Germany: Springer, 2023, pp. 773–787.

[22] K. N. S. Sree, D. V. Sree, G. H. Lakshmi, and S. Ramesh, "Diabetic retinopathy detection using deep learning," in *Proc. 2nd Int. Conf. Electron. Sustain. Commun. Syst. (ICESC)*, Aug. 2021, pp. 1670–1674.

[23] U. Bhimavarapu and G. Battineni, "Deep learning for the detection and classification of diabetic retinopathy with an improved activation function," *Healthcare*, vol. 11, no. 1, p. 97, Dec. 2022.

[24] J. Hyma, M. R. Murty, S. R. Mishra, and Y. Anuradha, "Classification of diabetic retinopathy using deep neural networks," in *Intelligent System Design*. Berlin, Germany: Springer, 2023, pp. 475–482.

[25] S. Aykat and S. Senan, "Deep learning in retinal diseases diagnosis: A review," in *Machine Learning and AI Techniques in Interactive Medical Image Analysis*. 2023, pp. 1–34, doi: 10.4018/978-1-6684-4671-3.ch001.

[26] B. Sowmyashree, K. R. Mahesh, and H. K. Chethan, "Approaches for detection of diabetic retinopathy: A review," in *Computer Networks and Inventive Communication Technologies* (Lecture Notes on Data Engineering and Communications Technologies), vol. 141, S. Smys, P. Lafata, R. Palanisamy, and K. A. Kamel, Eds. Singapore: Springer, doi: 10.1007/978-981-19-3035-5_16.

[27] T. Nazir, A. Irtaza, A. Javed, H. Malik, D. Hussain, and R. A. Naqvi, "Retinal image analysis for diabetes-based eye disease detection using deep learning," *Appl. Sci.*, vol. 10, no. 18, p. 6185, Sep. 2020.

[28] G. Hinton, N. Srivastava, and K. Swersky, "Neural networks for machine learning lecture 6a overview of mini-batch gradient descent," *Cited*, vol. 14, no. 8, p. 2, 2012.

[29] X. Pan, K. Jin, J. Cao, Z. Liu, J. Wu, K. You, Y. Lu, Y. Xu, Z. Su, J. Jiang, K. Yao, and J. Ye, "Multi-label classification of retinal lesions in diabetic retinopathy for automatic analysis of fundus fluorescein angiography based on deep learning," *Graefe's Arch. Clin. Experim. Ophthalmol.*, vol. 258, no. 4, pp. 779–785, Apr. 2020.

[30] A. Samanta, A. Saha, S. C. Satapathy, S. L. Fernandes, and Y.-D. Zhang, "Automated detection of diabetic retinopathy using convolutional neural networks on a small dataset," *Pattern Recognit. Lett.*, vol. 135, pp. 293–298, Jul. 2020.

[31] W. Zhang, J. Zhong, S. Yang, Z. Gao, J. Hu, Y. Chen, and Z. Yi, "Automated identification and grading system of diabetic retinopathy using deep neural networks," *Knowl.-Based Syst.*, vol. 175, pp. 12–25, Jul. 2019.

[32] 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 *Proc. Int. Conf. Sci. Appl. Sci. (ICSAS)*, 2020, Art. no. 020034.

[33] R. Sarki, K. Ahmed, H. Wang, and Y. Zhang, "Automated detection of mild and multi-class diabetic eye diseases using deep learning," *Health Inf. Sci. Syst.*, vol. 8, no. 1, pp. 1–9, Dec. 2020.

[34] R. Sarki, K. Ahmed, H. Wang, S. Michalska, and Y. Zhang, "Early detection of diabetic eye disease from fundus images with deep learning," in *Proc. Australas. Database Conf.*, 2020, pp. 234–241.

[35] F. Grassmann, J. Mengelkamp, C. Brandl, S. Harsch, M. E. Zimmermann, B. Linkohr, A. Peters, I. M. Heid, C. Palm, and B. H. F. Weber, "A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography," *Ophthalmology*, vol. 125, no. 9, pp. 1410–1420, Sep. 2018.

[36] S. Malik, N. Kanwal, M. N. Asghar, M. A. A. Sadiq, I. Karamat, and M. Fleury, "Data driven approach for eye disease classification with machine learning," *Appl. Sci.*, vol. 9, no. 14, p. 2789, Jul. 2019.

[37] N. Theera-Umpon, I. Poonkasem, S. Auephanwiriyakul, and D. Patikulsila, "Hard exudate detection in retinal fundus images using supervised learning," *Neural Comput. Appl.*, vol. 32, no. 17, pp. 13079–13096, Sep. 2020.

[38] X. Zeng, H. Chen, Y. Luo, and W. Ye, "Automated diabetic retinopathy detection based on binocular siamese-like convolutional neural network," *IEEE Access*, vol. 7, pp. 30744–30753, 2019.

[39] J. de la Torre, A. Valls, and D. Puig, "A deep learning interpretable classifier for diabetic retinopathy disease grading," *Neurocomputing*, vol. 396, pp. 465–476, Jul. 2020.

[40] Q. Abbas and M. E. A. Ibrahim, "DenseHyper: An automatic recognition system for detection of hypertensive retinopathy using dense features transform and deep-residual learning," *Multimedia Tools Appl.*, vol. 79, nos. 41–42, pp. 31595–31623, Nov. 2020.

[41] T. Araújo, G. Aresta, L. Mendonça, S. Penas, C. Maia, Â. Carneiro, A. M. Mendonça, and A. Campilho, "DR|GRADUATE: Uncertainty-aware deep learning-based diabetic retinopathy grading in eye fundus images," *Med. Image Anal.*, vol. 63, Jul. 2020, Art. no. 101715.

[42] M. N. Bajwa, M. I. Malik, S. A. Siddiqui, A. Dengel, F. Shafait, W. Neumeier, and S. Ahmed, "Two-stage framework for optic disc localization and glaucoma classification in retinal fundus images using deep learning," *BMC Med. Informat. Decis. Making*, vol. 19, no. 1, pp. 1–16, Dec. 2019.

[43] Y. Jiang, L. Duan, J. Cheng, Z. Gu, H. Xia, H. Fu, C. Li, and J. Liu, "JointRCNN: A region-based convolutional neural network for optic disc and cup segmentation," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 2, pp. 335–343, Feb. 2020.

[44] R. V. H. Kälviäinen and H. Uusitalo, "DIARETDB1 diabetic retinopathy database and evaluation protocol," in *Medical Image Understanding and Analysis*, vol. 2007. Citeseer, 2007, p. 61.

[45] Dataset. (2020). *Ocular Disease Intelligent Recognition (ODIR)*. Accessed: Jun. 12, 2023. [Online]. Available: http://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k

[46] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, B. Charton, and J.-C. Klein, "Feedback on a publicly distributed image database: The Messidor database," *Image Anal. Stereology*, vol. 33, no. 3, p. 231, Aug. 2014.

[47] L. Giancardo, F. Meriaudeau, T. P. Karnowski, Y. Li, S. Garg, K. W. Tobin, and E. Chaum, "Exudate-based diabetic macular edema detection in fundus images using publicly available datasets," *Med. Image Anal.*, vol. 16, no. 1, pp. 216–226, Jan. 2012.

[48] (2020). *Retina Dataset*. Accessed: Jun. 12, 2023. [Online]. Available: http://www.kaggle.com/datasets/jr2ngb/cataractdataset

ARWA ALBELAIHI received the B.S. and M.S. degrees from the Information Technology Department. She is currently pursuing the Ph.D. degree in data science with the University of Leeds, U.K. She is also a Demonstrator with the Department of Information Technology, College of Computer, Qassim University, Saudi Arabia. Her research interests include reinforcement learning, artificial intelligence, machine learning, and deep neural networks.

DINA M. IBRAHIM was born in United Arab Emirates. She received the B.Sc., M.Sc., and Ph.D. degrees from the Computers and Control Engineering Department, Faculty of Engineering, Tanta University, Egypt, in 2002, 2008, and 2014, respectively. She has been an Associate Professor with the Department of Information Technology, College of Computers, Qassim University, Buraydah, Saudi Arabia, since September 2015. In addition, she is also an Assistant Professor with the Computers and Control Engineering Department, Faculty of Engineering, Tanta University. She was a Consultant Engineer, then a Database Administrator, and the Vice Manager of the Management Information Systems (MIS) Project, Tanta University, from 2008 to 2014. She has published more than 60 articles in various refereed international journals and conferences. Her research interests include networking, wireless communications, machine learning, and the Internet of Things. She has been serving as a Reviewer for Springer journals, such as *Wireless Networks* (WINE), and the *Journal of Mobile Communication, Computation, and Information*, since 2015; and the MDPI journals, such as IEEE Access and *International Journal of Supply and Operations Management* (IJSOM), in 2021. She has been serving as the Co-Chair for the International Technical Committee for the Middle East Region of the ICCMIT Conference, since 2021.

● ● ●