Indonesian Journal of Electrical Engineering and Computer Science
Vol. 37, No. 3, March 2025, pp. 1830∼1844
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v37.i3.pp1830-1844

RNN_Diabetic framework for identifying diabetic eye diseases

Arwa Albelaihi[1,2], Dina M. Ibrahim[1,3]
[1]Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia
[2]School of Computing, University of Leeds, Leeds, United Kingdom
[3]Department of Computers and Control Engineering, Faculty of Engineering, Tanta University, Tanta, Egypt

Article Info

Article history:

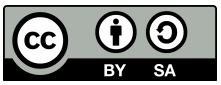
Received Jan 9, 2024
Revised Oct 3, 2024
Accepted Oct 7, 2024

Keywords:

Bi-GRU
Deep learning
Diabetic eye disease
GRU
Multi-class classification
Recurrent neural networks
Retinal fundus images

ABSTRACT

Many areas of image identification and classification for medical imaging diagnostics have greatly benefited from deep learning (DL). Diabetic retinopathy (DR) will become the most common cause of blindness worldwide, making diabetes a major threat to public health. This research proposes an automated identification system using deep recurrent neural networks (RNNs) to identify and classify four categories of diabetic eye diseases: DR, cataract, glaucoma, and diabetic macular edema (DME). We use three different model architectures based on RNN and their types, we called our proposed system RNN_Diabetic framework. These models are combined with one of the commonly used architectures that support sufficient accuracy and speed for the model which is residual network (ResNet)152V2. The three model architectures are RNN+ResNet152V2, gated recurrent unit (GRU)+ResNet152V2, and bidirectional GRU (Bi-GRU)+ResNet152V2. The proposed models were assessed as collected datasets: DIARETDB0, DIARETDB1, messidor, HEI-MED, ocular, and retina. A full analysis and evaluation of these three deep RNN architectures are presented. The experiments showed that the Bi-GRU+ResNet152V2 model worked better than the other two proposed models. In addition, we compare these three proposed models with the previous studies and find that the proposed Bi-GRU+ResNet152V2 model achieves the highest results with accuracy equal to 99.8%, 98.1% sensitivity, 98.6% specificity, 99.8% precision, 99.8% F1_score, and 99.8% areas under the curve (AUC).



Corresponding Author:

Dina M. Ibrahim
Department of Information Technology, College of Computer, Qassim University
Buraydah, Saudi Arabia
Email: d.hussein@qu.edu.sa

1. INTRODUCTION

Deep learning (DL) is utilized in many areas to find new ways to solve important problems, and it does well on tests that sort things into groups. Applications of artificial intelligence (AI) tools and techniques are appropriate in the medical field. One of the 21st century's most potently revolutionary technologies is AI. Strong machine learning (ML) tools and methods like deep convolutional networks, generative adversarial networks (GANs), convolutional neural networks (CNNs), recurrent neural networks (RNNs), and artificial neural networks (ANNs) were used to bring about this change. DL has outperformed AI in speech recognition, picture characterisation, and natural language creation.

In several areas of medical imaging diagnostics, including image recognition and classification, DL has proven to be an effective and significant technique. By using fundus images to assess and diagnose eye disorders, DL can be used to detect and classify eye diseases, including diabetic eye disease. Under the general category of diabetic eye illness are diabetic retinopathy (DR), cataract, glaucoma, and diabetic macular edema (DME) [1], [2]. Diabetic eye disease can cause serious vision loss or even cause blindness in patients between the ages of 20 and 74. Without early detection of diabetic eye disease, vision loss cannot be prevented. If identified early, 90% of diabetics can prevent diabetic eye injury.

The goal of this research is to improve diabetic eye disease (DL) detection models. To be employed, fundus images of diabetic eye diseases must be gathered in DL models. After that, several image preparation methods are applied to the images. Utilizing pre-processed images, features are automatically extracted, and analysis rules are learned and evaluated in the literature with a focus on using ML for respiratory disease detection and classification. Including the most important articles released from 2018 till the end of 2021. These findings help researchers plan their work and provide a useful contribution. Therefore, we discovered that Ibrahim *et al.* [3] used transfer learning (TL) with the VGG19-CNN, ResNet152V2, ResNet152V2+gated recurrent unit (GRU), and ResNet152 V2+bidirectional GRU (Bi-GRU) architectures for multiclass classification on a combination of chest X-ray and computed tomography (CT) scan images. Their research yielded a wealth of advancements that encouraged us to use GRU and Bi-GRU for the multiclassification of diabetic ocular disorders, particularly since these models are novel to this field. Additionally, Minarno *et al.* [4] found that using some augmentation procedures along with EfficientNetB7 models increases the classification accuracy for diabetic eye disorders.

Patients with diabetes are among the most common illness groups in the world today. Visual loss can result from diabetic fundus disorders, which are the primary cause of blindness. Visual function is currently affected by fundus illnesses such as DR, glaucoma, cataracts, and others [5]–[10]. CNN's outstanding feature learning capabilities have allowed it to excel in the realm of fundus imaging. Numerous deep-learning architectures that have demonstrated exceptional performance in identifying certain diabetic eye conditions have been documented in the literature. To the best of our knowledge, the categorization paradigm of the four diabetes-related disorders (DR, DME, glaucoma, and cataract) had not been significantly altered before the start of our study. However, by the end of 2021, Sarki *et al.* [11] had provided a framework for those four disorders' classification. They only used one CNN model in their research, though, and the dataset they used revealed class disparities. Their accuracy of 81.33% is still regarded as low when compared to the latest studies on DL classifications.

Previous studies on cataract disease did not sufficiently examine cataract prediction and classification [12]. To categorize the cataract disease individually, there was also independent research. DME disease identification was suggested as a research gap in [13] because it is extremely likely to indicate that the retina is developing DR. This information helps researchers better understand the causes of retina-based disorders. The second DL model focuses on training datasets with unequal distributions of classes across various diseases and limited data. The findings of the training set may not be accurate enough to be deemed satisfactory if it is small. It is taken into consideration as a potential remedy to apply traditional data augmentation techniques as enhancing methods. This study investigated the effects of combining RNNs with residual networks (ResNet) in identifying diabetic eye diseases. While earlier studies have explored the impact of only ResNet, they have not explicitly addressed its influence on increasing the performance of identifying diabetic eye diseases. The main contributions of this study are:

- Presenting the RNN_diabetic framework, a multi-classification DL model designed to detect and classify the four main eye disorders associated with diabetes: glaucoma, cataract, DME, and DR.
- Illustrating the influence of combining RNNs with ResNet in identifying diabetic eye diseases by using ResNet152V2, GRU, and Bi-GRU.
- Analyzing these DL architectures are provided using four classes of public fundus datasets (DR, DME, glaucoma, and cataract).
- Evaluating the suggested work by comparing its performance metrics to those of various models given in prior studies for detecting and classifying diabetic eye diseases.

The paper's remaining outline: section 2 summarizes the literature review, whereas section 3 details the study's materials and techniques. Section 4 presents the comparative analysis and discussion. In section 5, the paper concludes.

2. LITERATURE REVIEW

In this section, we present a few related deep neural network studies on diabetic eye problems. We found that careful evaluation of the architecture and the data input is necessary for creating an effective neural network classifier. Many studies have been conducted in the literature, as shown in [14]–[17], regarding using DL for classifying diabetic eye illness utilizing fundus images. More people with diabetic eye disease are presented in [18]–[22] adopt the TL approach. Rather than using random generation, the TL initializes the parameters based on prior learning. The first layers pick up on extracting essential elements like edges, textures, and intuitively. The top layers, analogous to blood vessels and exudates, are more particular to the task. The TL is useful in cases where there is not enough data to train a neural network from the beginning [23]–[29].

Three CNN models DenseNet, ResNet50, and VGG16 were compared by Pan *et al.* [30] on four different categories of DR lesions: microaneurysms, laser scars, leakages, and non-perfusion regions (NP). 4,067 fundus fluorescein angiography images from 435 eyes, 218 left eyes and 217 right eyes were collected into databases by them. In terms of computation, DR lesion identification and the procedure were determined to be efficient. According to experimental results, DenseNet is a useful model for automatically recognizing and classifying retinal lesions in multi-label-classified fundus fluorescein angiography (FFA) images. Nevertheless, because microaneurysms are often misclassified due to the widespread presence of fluorescein, the procedure is not very accurate in identifying them. Samanta *et al.* [31] introduced a TL-based CNN architecture for color fundus photography, which performs relatively well in identifying DR from hard exudates, blood vessels, and texture on a much smaller dataset. The dataset was trained using their model on four classes (no DR, mild DR, moderate DR, and proliferative DR) and reached a Cohen's Kappa score of 0.8836 on the validation set and 0.9809 on the training set. Their model uses several architectures, such as Inceptionv1, Inceptionv2, Inceptionv3, Xception, VGG16, ResNet-50, DenseNet, and AlexNet.

Zhang *et al.* [32] presented a DeepDR framework for DR detection. DeepDR actively detects the existence and severity of DR from fundus images by ensemble learning and TL, using several transfer and ensemble learning methods. Additionally, the authors introduced a new dataset for DR images called macula-centered retinal fundus images (13,767 images of 1,872 patients). The proposed network achieved 97.5% sensitivity with a specificity of 97.7%. However, their model needs to be assessed with a more complex and larger dataset. Using a small dataset, Badah *et al.* [23] developed a model to detect hard exudates using support vector machine (SVM), a multilayer perceptron network. They found that CNN is not the best classifier for hard exudate, but it does perform well. Moreover, good outcomes in challenging extraction can be obtained by combining efficient classifiers with image-processing methods. A recent study demonstrated the application of CNN based on the ResNet152 model in conjunction with other ML classifiers such as SVM, K-nearest neighbors (KNNs), Naive Bayes, multi-layer perceptron, decision tree, and random forest to identify infections causing glaucoma in patients.

Furthermore, Dipu *et al.* [10], The ODIR dataset was classified using four TL-based models. The accuracy of the models was 97.23% for VGG-16, 90.85% for ResNet-34, 94.32% for MobileNetV2, and 93.82% for EfficientNet. CNN can train discriminative features with an attention module for free, thanks to the design proposed in [20]. They found multiple common labels for a patient's left and right fundus image pairings, which allowed them to address the class balancing problem in the highly unbalanced ODIR dataset.

Bhimavarapu and Battineni [25], an enhanced activation function was suggested for using fundus images to diagnose DR. This function automatically lowers loss and processing time in the different CNN models. To categorize DR images, four datasets DRIVE, CHASE, DIARETDB0, and Kaggle were taken into consideration. The current activation functions used were scaled exponential linear unit (SELU), rectified linear unit (ReLU), Sigmoid, and exponential linear unit (ELU). On the Kaggle dataset, ResNet-152 has the greatest accuracy (99.41%). Butt *et al.* [18] suggests utilizing TL to extract fund picture features using ResNet-18 and GoogleNet models to correct class imbalance in a public dataset. Similarly, [19], [26] enhances picture quality with a hybrid DL classifier, utilizing ResNet50-based architecture. Table 1 shows a summary of studies that used dDL methods, including how they worked and how well they classified images of the fundus that show diabetic diseases using various models of ResNet.

3. MATERIALS AND METHODS

Our mission is to evaluate multiple DL models for detecting these four diabetes-related ocular problems. Our dataset will be compiled using data sourced from six separate open-source databases. Once images

have been collected, they undergo several preprocessing steps. Once the dataset has been completed, the pre-trained ResNet152V2 model is used to classify each disease using three different models. There are also RNN models of the GRU and the Bi-GRU that are linked to ResNet152V2. Hence, we named the system we suggested RNN_Diabetic framework. Figure 1 illustrates a summary of our proposed RNN_Diabetic framework methodology.

Table 1. Summary of classification-based methods using different models of ResNet and their corresponding measures in previous studies

Study	Dataset	DL model	Performance	
[30]	Kaggle	ResNet50-I	Acc=81.40%	Spc=87.9%
			Sen=59.0%	AUC=85.2%
		ResNet50-II	Acc=90.97%	Spc=92.7%
			Sen=97.6%	AUC=92.31%
		ResNet50-III	Acc=95.85%	Spc=98.8%
			Sen=70.4%	AUC=96.04%
		ResNet50-IV	Acc=91.15%	Spc=95.4%
			Sen=69.7%	AUC=93.51%
[25]	DIRATEDB0	ResNet-50	Acc=93.54%	Pre=99.43%
			Sen=95.27%	F1=98.42%
			Spe=98.32%	AUC=89%
		ResNet-152	Acc=96.64%	Pre=99.53%
			Sen=97.96%	F1=99.15%
			Spe=98.79%	AUC=97%
	Kaggle	ResNet-50	Acc=94.64%	Pre=95.74%
			Sen=94.24%	F1=97.72%
			Spe=96.86%	AUC=93%
		ResNet-152	Acc=99.41%	Pre=99.89%
			Sen=98.28%	F1=99.93%
			Spe=99.94%	AUC=98%
	DRIVE	ResNet-50	Acc=92.44%	Pre=94.83%
			Sen=93.72%	F1=95.88%
			Spe=95.27%	AUC=93%
		ResNet-152	Acc=97.84%	Pre=99.68%
			Sen=98.45%	F1=99.57%
			Spe=99.26%	AUC=94%
	CHASE	ResNet-50	Acc=93.83%	Pre=95.73%
			Sen=93.22%	F1=94.68%
			Spe=96.95%	AUC=96%
		ResNet-152	Acc=99.05%	Pre=99.94%
			Sen=98.45%	F1=99.89%
			Spe=99.59%	AUC=97%
[10]	ODIR	ResNet34	Acc=90.85%	Pre=93.7%
			Sen=93.17%	F1=92.65%
[20]	ODIR	InceptionResNet	Acc=97.38%	F1=94.58%
			Sen=96.08%	AUC=96.08%
			Pre=94.28%	

3.1. Stage 1: datasets for the study

3.1.1. Data collection

Numerous sources of fundus images were accessed and compiled for our experiments. Included in this group are DME, DR, cataract, and glaucoma. Initially, we chose 219 image datasets from DIARETDB0 and DIARETDB1 [33] for DR. Second, for DME with 151 images and 169 images, respectively, we used Messidor [34] and HEI-MED [35]. The ODIR dataset [36] is the third source of the dataset and includes information on several classifications of eye diseases. For cataracts, we used 312 images, and for glaucoma, we used 178 images. The final dataset used for both glaucoma and cataracts is a retina dataset consisting of 99 and 100 pictures, respectively, according to [37]–[49]. 412, 320, 277, and 219 images for cataract, DME, glaucoma, and DR, respectively, are the results of the datasets that were gathered. The datasets that were gathered yielded a total of 1,228 images, which were then randomly split into training and validation sets. The collected datasets yielded 412, 320, 277, and 219 pictures for cataract, DME, glaucoma, and DR, respectively as shown in Table 2. From the datasets, 1,228 photos were randomly separated into training and validation sets.

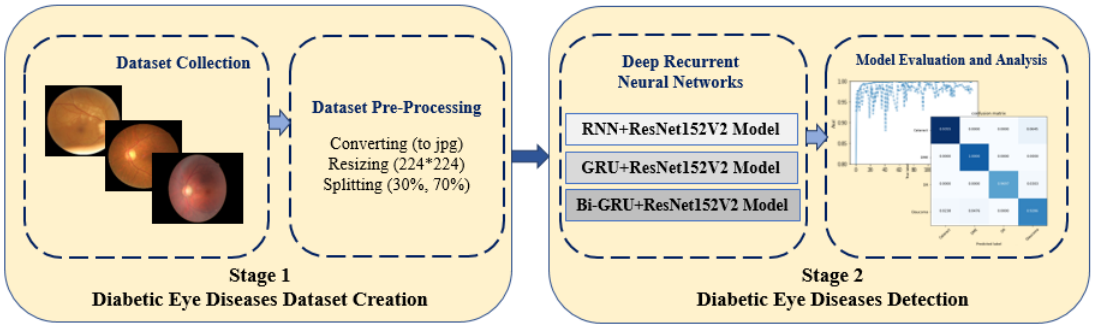


Figure 1. The methodology of our proposed RNN_Diabetic framework

Table 2. Dataset details for DR, cataract, glaucoma, and DME classes

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

3.1.2. Dataset pre-processing

Preparing a dataset to satisfy the demands of a DL model is commonly achieved through pre-processing. Distinct pre-processing stages in the input images of our model. The dataset is divided into four classifications: DR, cataract, glaucoma, and DME, as the initial step in the process. JPEG, PNG, GIF, BMP, and TIFF are found among the image formats that TensorFlow does not support. The files' extensions are merely for indication; the contents of the images are not altered in any way. All image extensions are replaced with the JPG format in this research. These images will then be used as input in the step of pre-processing data. After reading, they were downsizing the images to 224×224, with a batch size of 32. This needs to be supplied because the pipeline handles batches of images that are all the same size. They were offering genuine value when it came to data shuffle decisions.

To get transformations equal to 100 and shuffling, they added optional random seeds. After that, the images should be normalized and turned into arrays so that, they may be used as input for the model's next step. With the label mode categorial, a float 32 tensor of shape (batch_size, num_classes) encodes the class index one-hot. To ensure our dataset for DL model training include different photos. Finally, we randomly divided it into 70% and 30% sections for training and validation. Sample fundus pictures for DR, cataract, glaucoma, and DME are shown in Figure 2.

3.2. Stage 2: deep recurrent neural networks

Several supervised DL algorithms are used to create classification models in this research. To detect and eliminate the four diabetic eye illnesses, they will test their abilities. Three models are used: RNN+ ResNet152V2, GRU+ResNet152V2, and Bi-GRU+ResNet152V2. We explain each of the three models in the following subsections.

3.2.1. Stage 2_model 1: RNN+ResNet152V2 deep model

The ResNet152V2 model was selected for this research as the model supports sufficient accuracy and speed. The model has a RNN using the ResNet152V2 model, a reshape layer, 2D global average pooling, a dropout layer, and a dense layer with a Softmax activation function to classify the image. Figure 3 provides more details on the model's architecture.

3.2.2. Stage 2_model 2: ResNet152V2 and GRU deep model

As shown in Figure 4, this model consists of GRU followed by ResNet152V2. The primary benefit of using GRU is that it allows for the maintenance of details that are irrelevant to the prediction for an extended

period without eliminating them. ResNet152V2, a reshape layer, a GRU layer after layer with 256 units, a collapsed layer, a layer that is dense with 128 neurons, a layer with dropouts, and a dense layer use the function of Softmax activation to classify the image into one of our four disease classifications.

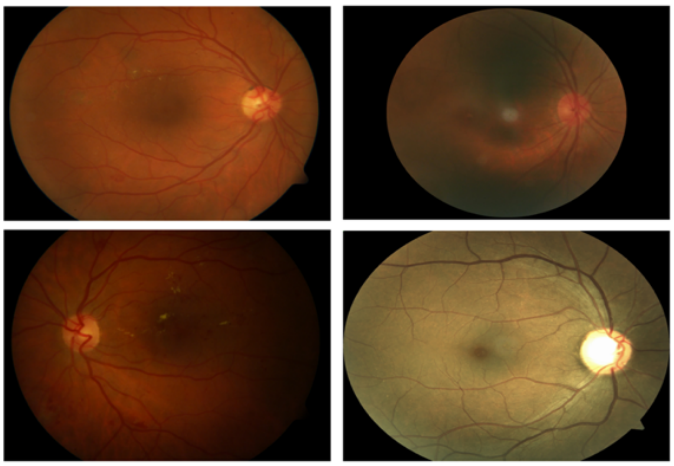


Figure 2. Fundus image samples from the original dataset

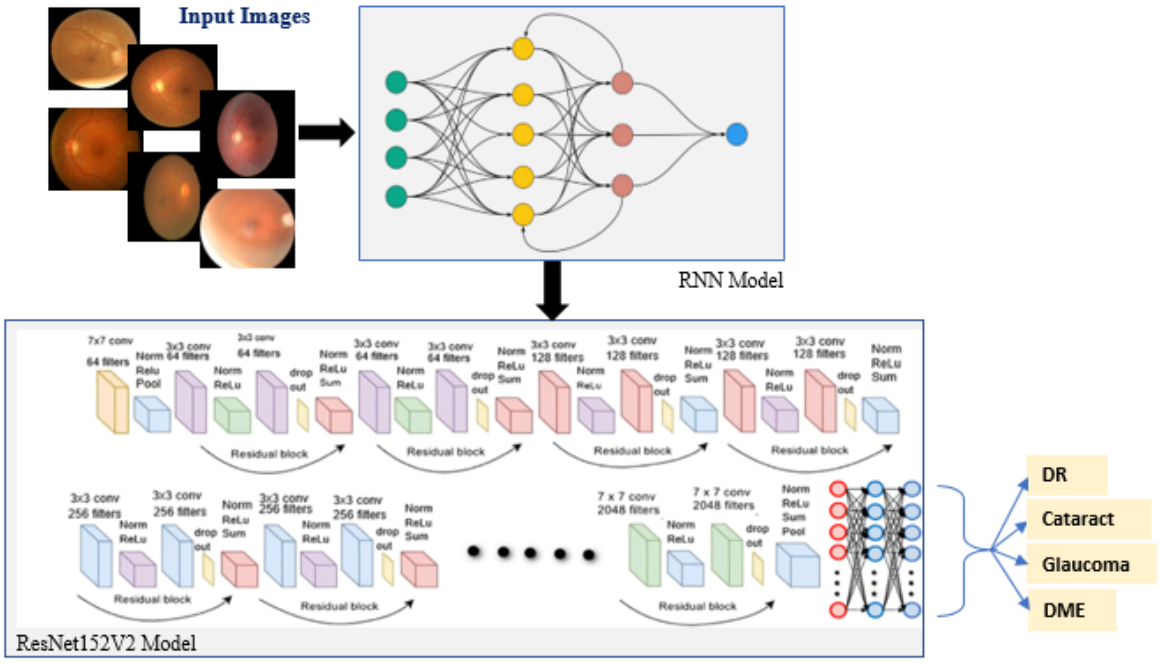


Figure 3. Stage 2_model 1: RNN+ResNet152V2 model architecture

3.2.3. Stage 2_model 3: ResNet152V2 and Bi-GRU deep model

The sequential model of bi-GRU, followed by ResNet152V2, is the last proposed DL model in our proposed RNN_Diabetic framework. The model architecture is described in detail in Figure 5. The model consists of a ResNet152V2, followed by a reshape layer, a bi-GRU layer with 512 units, a dense layer, a dropout layer, and another dense layer, and it utilizes a Softmax activation function to categorize the image as belonging to one of our four categories of disorders.

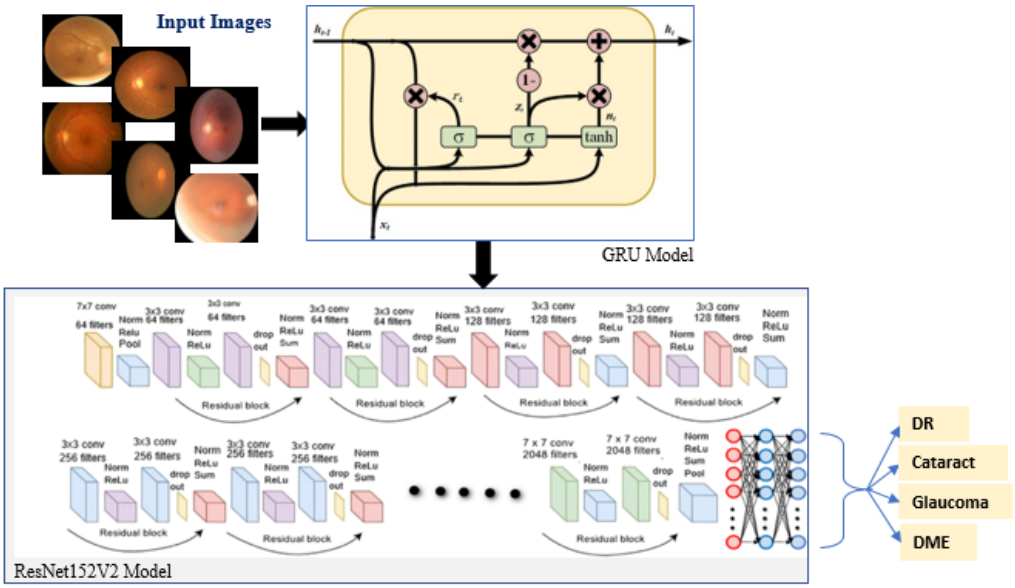


Figure 4. Stage 2_model 2: GRU+ResNet152V2 model architecture

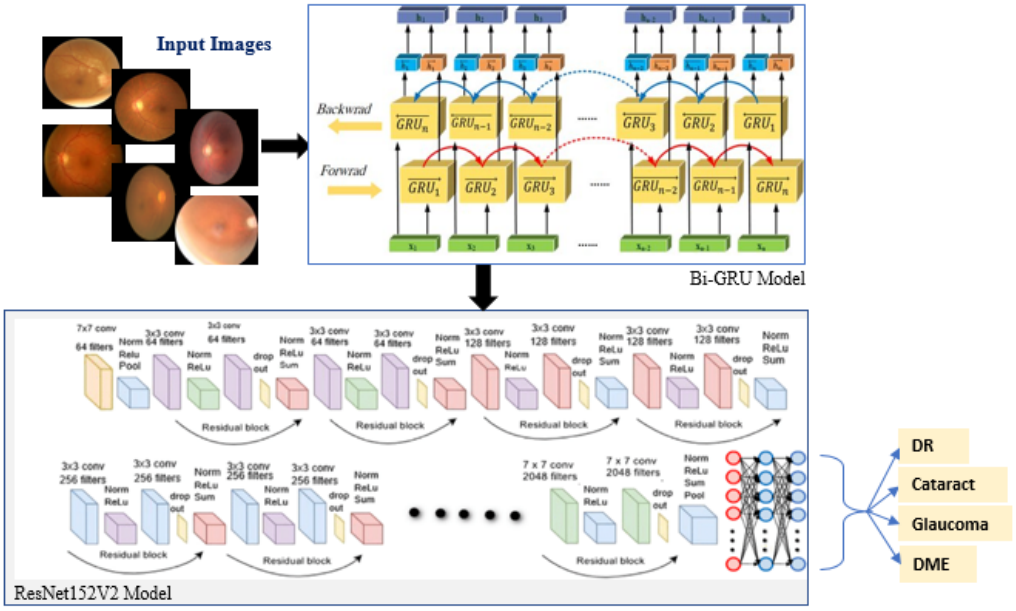


Figure 5. Stage 2_model 3: Bi-GRU+ResNet152V2 model architecture

4. RESULTS AND DISCUSSIONS

4.1. Training parameters

The models in this study were trained and assessed independently but with identical parameters and metrics. We needed to increase the overall amount of data and epochs to adequately train DL models. To achieve the best results, we trained the models on various epochs and batch size settings. Using an optimizer and suitable fit functions, our models were trained and verified. Every model ran through about 300 epochs of training, with a batch size of 32. The presented findings are the highest validation values that could be obtained. The results were obtained by including the performance metric equations in the outputs of the resulting validation data. For our suggested models, the (Adam) [7] optimizer was used to obtain the optimal outcomes. Table 3 displays the learning rate (LR) values and optimizers used for all models.

Table 3. The training parameters of the models: learning rate value and optimizer

Model	Optimizer	Learning rate
RNN+ResNet152V2	Adam	0.00001
GRU+ResNet152V2		0.00001
Bi-GRU+ResNet152V2		0.00001

4.2. Early stopping (callback)

Neural network training requires epoch selection. Too many or too few epochs might overfit or underfit the training dataset. Set epochs with early stopping, model checkpoint, and other approaches. Our early stopping method involves setting many epochs and halting training when the improvement is not enough and matches our expectations. Early stopping sets an arbitrarily high number of training epochs and stops training when model performance on a hold-out validation dataset stops improving. The early stopping callback from Keras lets you terminate training early. Automatic callbacks run code and interact with the training model. The validation dataset loss function for model optimization is calculated after each epoch. First evidence of no improvement is not always ideal to stop training. This is because the model may stall or worsen before improving rapidly. To postpone epochs without improvement, we set "patience" to 60 in this experiment.

4.3. Performance metrics

The RNN_Diabetic framework was assessed for diabetic eye disease categorization utilizing measures such as accuracy, sensitivity, precision, specificity, F1-score, and areas under the curve (AUC). A confusion matrix is then added to each model. In (1), accuracy is the number of occurrences that can be predicted accurately given the total number.

$$Accuracy(Acc) = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} \tag{1}$$

where TP and TN are true positive and negative parameters, whereas FP and FN are false positive and negative values. In (2) gives recall as the number of samples that were both actually positive and projected to be positive. In (3) shows the precision, or positive predictive value, of all expected positive samples. In (4) gives the real negative rate or specificity, which is the number of truly negative samples anticipated to be negative. Last, (6) shows the F1-score, or harmonic mean of precision and sensitivity.

$$Sensitivity(Sen) = \frac{Tp}{Tp + Fn} \tag{2}$$

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

$$Specificity(Spc) = \frac{Tn}{Tn + Fp} \tag{4}$$

$$F1 - score(F1) = \frac{2 * Pre * Sen}{Pre + Sen} \tag{5}$$

Recent research suggests confusion matrix analysis is effective for model validation [8] due to its ability to characterize data linkages and distributions. Information on illustrating categorization models is included.

4.4. Multi-classification RNN_Diabetic framework results

We organized an extensive experiment, and we obtained different outputs corresponding to the performance metrics to discover the significant findings. Therefore, we illustrated the results in Table 4 for the three models. The table shows classifier model performance in accuracy (Acc), sensitivity (Sen), specificity (Spe), precision (Pre), F1-score (F1), and AUC. Most people use the confusion matrix to evaluate categorization errors. We report loss, AUC, precision, recall, and accuracy between training and validation with 300 epochs in all proposal models.

Table 4. The performance metrics for the three proposed models for our RNN_Diabetic framework

Model	Acc	Sen	Spe	Pre	F1	AUC
RNN+ResNet152V2	97.3	96.7	95.9	96.7	96.7	97.4
GRU+ResNet152V2	98.3	97.9	98.1	98.3	98.3	99
Bi-GRU+ResNet152V2	99.8	98.1	98.6	99.8	99.8	99.4

4.4.1. Results of stage 2_model 1 (RNN+ResNet152V2 model)

Figure 6 displays the ResNet152V2 model confusion matrix of the RNN+ResNet152V2 model. The figure shows the effective classification of four patient statuses: DR, DME, glaucoma, and cataract, with DR images having the greatest ratio, followed by cataract, DME, and glaucoma.

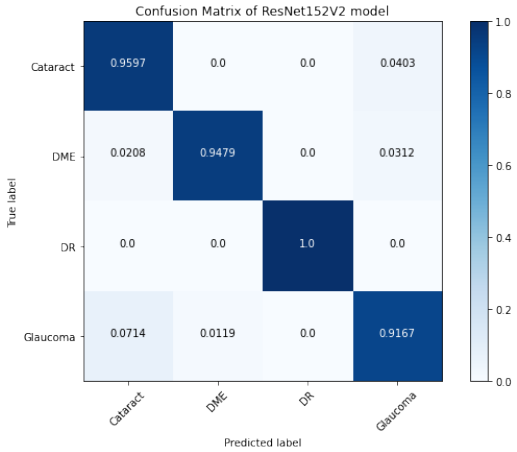


Figure 6. Confusion matrix of stage 2-model 1: RNN+ResNet152V2 model

4.4.2. Results of stage 2_model 2 (GRU+ResNet152V2 model)

The GRU+ResNet152V2 model demonstrated improved performance with GRU, as seen in Table 4; assumed values were higher. In Figure 7, the confusion matrix of the GRU+ResNet152V2 model classifies four patient statuses: DR, DME, glaucoma, and cataract. The greatest ratio is for DME and DR pictures, followed by cataract (0.9839) and glaucoma (0.9542).

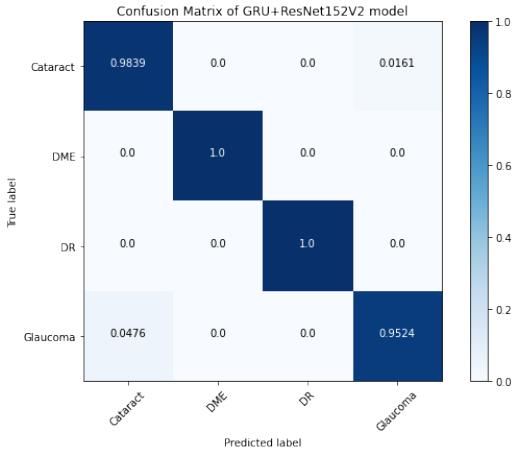


Figure 7. Confusion matrix of stage 2-model 2: GRU+ResNet152V2 model

4.4.3. Results of stage 2_model 2 (Bi-GRU+ResNet152V2 model)

Figure 8 displays the performance indicators of the Bi-GRU+ResNet152V2 model, including loss, AUC, precision, recall, and accuracy throughout training and validation, along with the number of epochs. See Figure 9 for the confusion matrix of the Bi-GRU+ResNet152V2 model. Figure shows four precisely recognized patient conditions: cataract, DR, DME, and glaucoma. The DR pictures have the highest ratio (0.9896), followed by the DME, cataract, and glaucoma. They outperform the GRU+ResNet152V2 model but are inferior to the RNN+ResNet152V2 model.

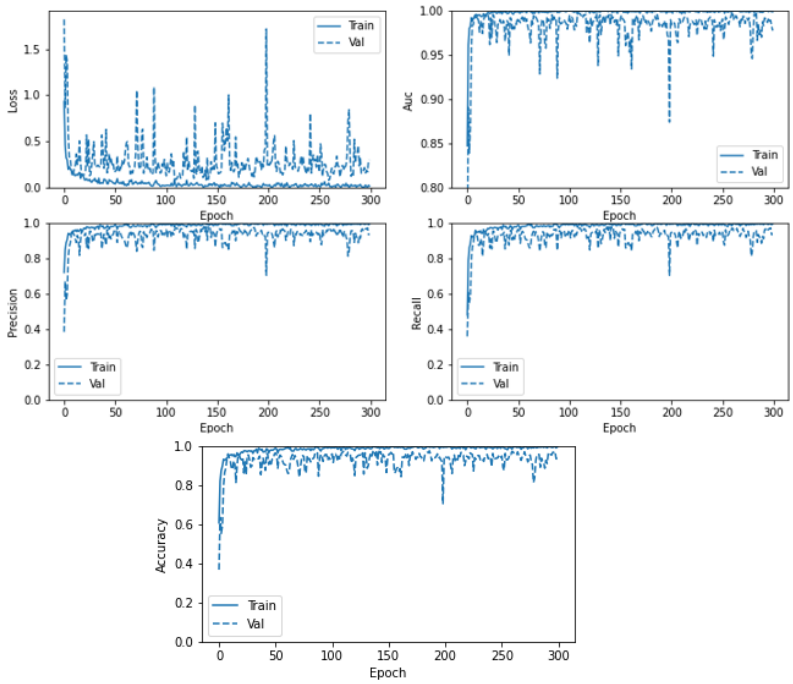


Figure 8. Loss, AUC, precision, recall, and accuracy between the training and validation phases, with number of epochs for stage 2-model 3: Bi-GRU+ResNet152V2 model

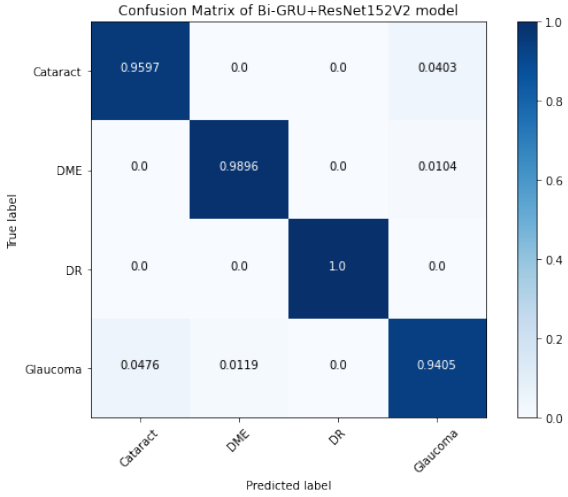


Figure 9. Confusion matrix of stage 2-model 3: Bi-GRU+ResNet152V2 model

4.5. Comparative analysis and discussion

This study recommends three deep RNN models for diabetic eye disease diagnosis: RNN+ResNet152V2, GRU+ResNet152V2, and Bi-GRU+ResNet152V2. DR, DME, cataract, and glaucoma the four most common diabetic eye diseases are classified using these models. The proposed models were evaluated using accuracy, sensitivity, F1-score, and AUC. According to our data, Bi-GRU+ResNet152V2 had the highest classification accuracy (99.8), followed by GRU+ResNet152V2. However, Table 4 and Figure 10 indicate that RNN+ResNet152V2 has the smallest accuracy among all designs.

Table 5 compares the performance evaluation measures of our proposed RNN_Diabetic framework to existing DL models from past publications. This research evaluated the proposed model using AUC, sensitivity, specificity, precision, F1-score, and accuracy. We used two figures to explain the comparisons. Figure 11

illustrates the comparison between our proposed RNN_ Diabetic framework (the three models) and the previous studies in terms of accuracy, sensitivity, and specificity, while Figure 12 represents the comparison between our proposed RNN_Diabetic framework (the three models) and the previous studies in terms of precision, F1-score, and AUC. For simplicity, we used DS1 to represent dataset 1 which is DIRATEDB0 for the study [25]. Similarly, DS2, DS3, and DS4 represent Kaggle, DRIVE, and CHASE, respectively. In comparison with previous studies [10], [20], [25], [30] the results show that our proposed Bi-GRU+ResNet152V2 model with accuracy 99.8% outperforms all the existing models by a significant value, which demonstrates the effectiveness of our predictive model.

The experiments showed that the RNN+ResNet152V2 model achieved results near to the previous studies with accuracy equal to 97.3%, 96.7% sensitivity, 95.9% specificity, 96.7% precision, 96.7% F1-score, and 97.4% AUC. While the GRU+ResNet152V2 model worked better than the RNN+ResNet152V2 model and achieved results with accuracy equal to 98.3%, 97.9% sensitivity, 98.1% specificity, 98.3% precision, 98.3% F1-score, and 99% AUC. Finally, the experiments showed that the Bi-GRU+ResNet152V2 model worked better than the other two proposed models. In addition, we compare these three proposed models with the previous studies and find that the proposed Bi-GRU+ResNet152V2 model achieves the highest results with accuracy equal to 99.8%, 98.1% sensitivity, 98.6% specificity, 99.8% precision, 99.8% F1-score, and 99.8% AUC.

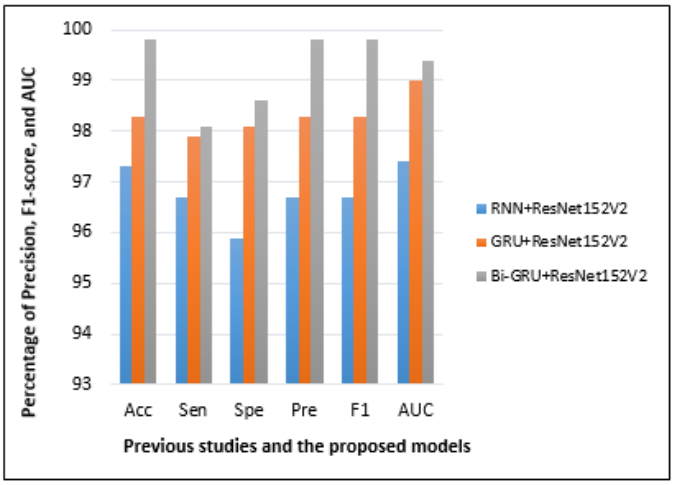


Figure 10. Comparison between the performance metrics for the three proposed models for our RNN_Diabetic framework

Table 5. Performance measures of ResNet152V2 model

Model	Acc	Sen	Spe	Pre	F1	AUC
ResNet50-I [30]	81.4	59	87.9	-	-	85.2
ResNet50-II [30]	90.7	97.6	92.7	-	-	92.3
ResNet50-III [30]	95.8	70.4	98.8	-	-	96.1
ResNet50-IV [30]	91.1	69.7	95.4	-		93.5
ResNet-50-DS1 [25]	93.5	95.2	98.3	99.4	98.4	89
ResNet-152-DS1 [25]	96.9	97.9	98.7	99.5	99.1	97
ResNet-50-DS2 [25]	94.6	94.4	96.8	95.7	97.7	93
ResNet-152-DS2 [25]	99.4	98.2	99.9	99.8	99.9	98
ResNet-50-DS3 [25]	92.4	93.7	95.2	94.8	95.8	93
ResNet-152-DS3 [25]	97.8	98.4	99.2	99.6	99.5	94
ResNet-50-DS4 [25]	93.8	93.2	96.9	95.7	94.6	96
ResNet-152-DS4 [25]	99.1	98.4	99.5	99.9	99.8	97
ResNet34 [10]	90.8	93.1	-	93.7	92.6	-
InceptionResNet [20]	97.3	96.1	-	94.2	94.5	96.1
RNN+ResNet152V2 (Prop1)	97.3	96.7	95.9	96.7	96.7	97.4
GRU+ResNet152V2 (Prop2)	98.3	97.9	98.1	98.3	98.3	99
Bi-GRU+ResNet152V2 (Prop3)	99.8	98.1	98.6	99.8	99.8	99.4

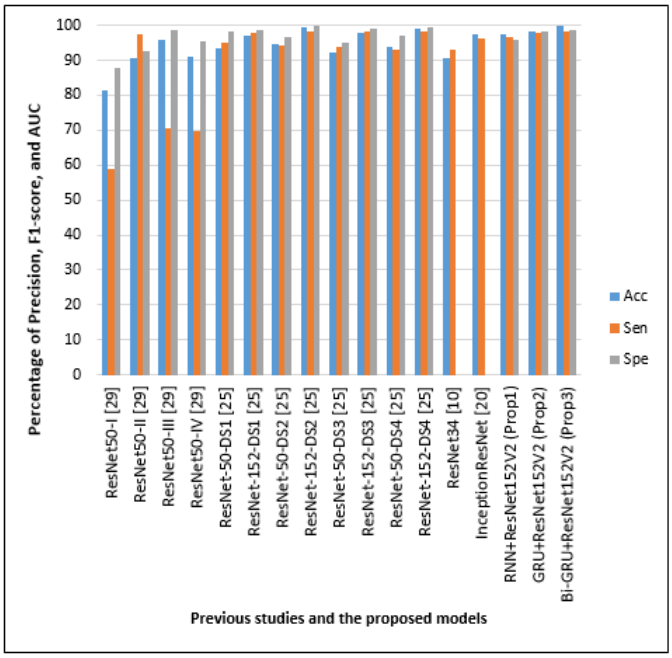


Figure 11. Comparison between our proposed RNN_ Diabetic framework (the three models) and the previous studies in terms of accuracy, sensitivity, and specificity

We found that combining RNN with residual networks in identifying diabetic eye diseases correlates with increasing the performance metrics. The proposed Bi-GRU+ResNet152V2 model in this study tended to have an inordinately higher proportion of accuracy equal to 99.8%, 98.1% sensitivity, 98.6% specificity, 99.8% precision, 99.8% F1-score, and 99.8% AUC.

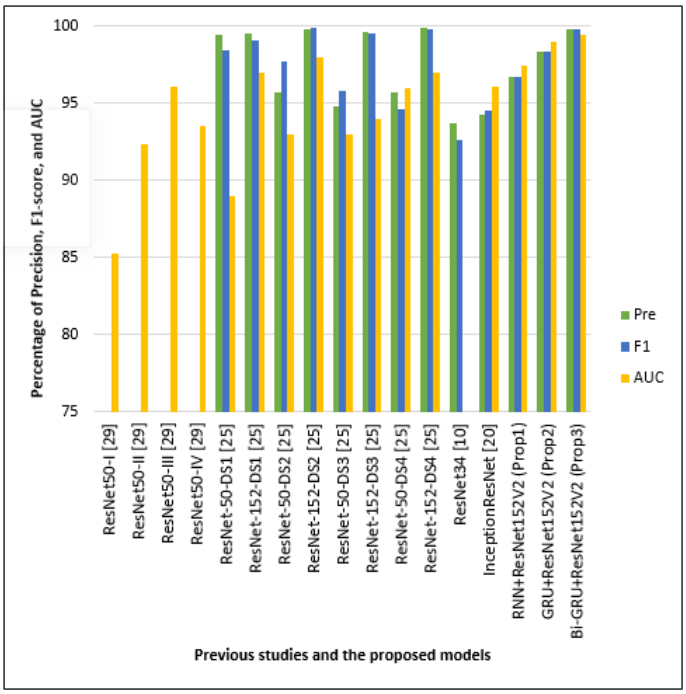


Figure 12. Comparison between our proposed RNN_ Diabetic framework (the three models) and the previous studies in terms of precision, F1-score, and AUC

5. CONCLUSION AND FUTURE WORK

A multi-classification DL model called the RNN_Diabetic framework was created and tested in this study to find DR, DME, glaucoma, and cataracts in fundus images. The main goal of this research is to illustrate the influence of combining RNNs with rsidual networks in identifying diabetic eye diseases. Early diagnosis helps guide treatment for many disorders. Fundus images are difficult to discern, and even competent ophthalmologists can misdiagnose eye lesions. This study gave three model architectures: RNN+ResNet152V2, GRU+ResNet152V2, and Bi-GRU+ResNet152V2. Although most previous research focused on classifying and developing a single fundus disease, the focus was on multi-classifying four diabetes eye diseases and explaining how RNNs and ResNet can identify diabetic eye diseases. This study examined the technique's real-world robustness and flexibility using many datasets. The proposed model was assessed by its accuracies, sensitivities, specificities, precisions, F1-scores, and AUC. With an accuracy of 99.8 percent, our proposed Bi-GRU+ResNet152V2 model significantly outperforms all the existing models, as shown by our comparison with previous studies. Recent observations suggest that the Bi-GRU+ResNet152V2 model worked better than the other two proposed models. In addition, we compare these three proposed models with the previous studies and find that the proposed Bi-GRU+ResNet152V2 model achieves the highest results with accuracy equal to 99.8%, 98.1% sensitivity, 98.6% specificity, 99.8% precision, 99.8% F1-score, and 99.8% AUC. To improve clinical acceptance of DL models, future research should focus on deep neural network features and visualization.

REFERENCES

[1] P. Vashist, S. Singh, N. Gupta, and R. Saxena, "Role of early screening for diabetic retinopathy in patients with diabetes mellitus: an overview," *Indian Journal of Community Medicine*, vol. 36, no. 4, p. 247, 2011, doi: 10.4103/0970-0218.91324.

[2] B. Aljaddouh and D. Malathi, "Trends of using machine learning for detection and classification of respiratory diseases: investigation and analysis," *Materials Today: Proceedings*, vol. 62, pp. 4651–4658, 2022, doi: 10.1016/j.matpr.2022.03.120.

[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," *Computers in Biology and Medicine*, vol. 132, p. 104348, May 2021, doi: 10.1016/j.compbiomed.2021.104348.

[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," *International Journal on Informatics Visualization*, vol. 6, no. 1, pp. 12–18, 2022, doi: 10.30630/joiv.6.1.857.

[5] F. Zia *et al.*, "A multilevel deep feature selection framework for diabetic retinopathy image classification," *Computers, Materials and Continua*, vol. 70, no. 2, pp. 2261–2276, 2022, doi: 10.32604/cmc.2022.017820.

[6] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019, doi: 10.1186/s40537-019-0197-0.

[7] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, Dec. 2014, [Online]. Available: http://arxiv.org/abs/1412.6980.

[8] U. Özkaya, Ş. Öztürk, and M. Barstugan, "Coronavirus (COVID-19) classification using deep features fusion and ranking technique," *Studies in Big Data*, vol. 78, pp. 281–295, 2020, doi: 10.1007/978-3-030-55258-9_17.

[9] F. Demir and B. Taşcı, "An effective and robust approach based on R-CNN+LSTM model and NCAR feature selection for ophthalmological disease detection from fundus images," *Journal of Personalized Medicine*, vol. 11, no. 12, p. 1276, Dec. 2021, doi: 10.3390/jpm11121276.

[10] N. M. Dipu, S. A. Shohan, and K. M. Salam, "Ocular disease detection using advanced neural network based classification algorithms," *Asian Journal of Convergence in Technology*, vol. 7, no. 2, pp. 91–99, Aug. 2021, doi: 10.33130/AJCT.2021v07i02.019.

[11] R. Sarki, K. Ahmed, H. Wang, Y. Zhang, and K. Wang, "Convolutional neural network for multi-class classification of diabetic eye disease," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 9, no. 4, pp. e15–e15, 2022, doi: 10.4108/eai.16-12-2021.172436.

[12] 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, doi: 10.1109/ACCESS.2020.3015258.

[13] 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, doi: 10.1109/ACCESS.2022.3157632.

[14] J.-H. Han, "Artificial intelligence in eye disease: recent developments, applications, and surveys," *Diagnostics*, vol. 12, no. 8, p. 1927, Aug. 2022, doi: 10.3390/diagnostics12081927.

[15] N. Durga, D. Kerana, and A. Muthukumaravel, "A systematic review on diabetic retinopathy and common eye diseases detection through deep learning techniques," *Journal of Positive School Psychology*, vol. 2022, no. 4, pp. 1905–1919.

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

[17] B. Sowmyashree, K. R. Mahesh, and H. K. Chethan, "Approaches for detection of diabetic retinopathy: a review," in *Lecture Notes on Data Engineering and Communications Technologies*, vol. 141, 2023, pp. 201–212.

[18] 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, 2022, doi: 10.3390/diagnostics12071607.

[19] 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, doi: 10.1109/ACCESS.2022.3189374.

[20] A. Bhati, N. Gour, P. Khanna, and A. Ojha, "Discriminative kernel convolution network for multi-label ophthalmic disease detection on imbalanced fundus image dataset," *Computers in Biology and Medicine*, vol. 153, p. 106519, 2023, doi: 10.1016/j.compbiomed.2022.106519.

[21] O. Ouda, E. Abdelmaksoud, A. A. A. El-Aziz, and M. Elmogy, "Multiple ocular disease diagnosis using fundus images based on multi-label deep learning classification," *Electronics (Switzerland)*, vol. 11, no. 13, p. 1966, 2022, doi: 10.3390/electronics11131966.

[22] A. Arora, S. Gupta, S. Singh, and J. Dubey, "Eye disease detection using transfer learning on VGG16," in *Lecture Notes in Networks and Systems*, 2022, vol. 421, pp. 527–536, doi: 10.1007/978-981-19-1142-2_42.

[23] N. Badah, A. Algefes, A. AlArjani, and R. Mokni, "Automatic eye disease detection using machine learning and deep learning models," in *Lecture Notes in Networks and Systems*, 2023, vol. 475, pp. 773–787, doi: 10.1007/978-981-19-2840-6_58.

[24] K. M. Fellah, S. Tigane, and L. Kahloul, "Diabetic retinopathy detection using deep learning," in *International Symposium on Modelling and Implementation of Complex Systems*, 2023, pp. 234–246, doi: 10.1007/978-3-031-18516-8_17.

[25] 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, doi: 10.3390/healthcare11010097.

[26] A. Albelaihi and D. M. Ibrahim, "DeepDiabetic: an identification system of diabetic eye diseases using deep neural networks," *IEEE Access*, vol. 12, pp. 10769–10789, 2024, doi: 10.1109/ACCESS.2024.3354854.

[27] J. Hyma, M. R. Murty, S. R. Mishra, and Y. Anuradha, "Classification of diabetic retinopathy using deep neural networks," *Intelligent System Design: Proceedings of India*, vol. 494, pp. 475–482, 2023, doi: 10.1007/978-981-19-4863-3_47.

[28] 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," *Applied Sciences*, vol. 10, no. 18, p. 6185, Sep. 2020, doi: 10.3390/app10186185.

[29] G. E. Hinton, N. Srivastava, and K. Swersky, "Neural networks for machine learning lecture 6a overview of mini-batch gradient descent.," *Cited on*, vol. 14, no. 8, p. 2, 2012, [Online]. Available: https://www.cs.toronto.edu/ tijmen/csc321/slides/lecture_slides_lec6.pdf.

[30] X. Pan *et al.*, "Multi-label classification of retinal lesions in diabetic retinopathy for automatic analysis of fundus fluorescein angiography based on deep learning," *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 258, no. 4, pp. 779–785, Apr. 2020, doi: 10.1007/s00417-019-04575-w.

[31] 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 Recognition Letters*, vol. 135, pp. 293–298, Jul. 2020, doi: 10.1016/j.patrec.2020.04.026.

[32] W. Zhang *et al.*, "Automated identification and grading system of diabetic retinopathy using deep neural networks," *Knowledge-Based Systems*, vol. 175, pp. 12–25, Jul. 2019, doi: 10.1016/j.knosys.2019.03.016.

[33] T. Kauppi *et al.*, "DIARETDB1 diabetic retinopathy database and evaluation protocol," *Medical image understanding and analysis*, vol. 61, 2007.

[34] E. Decencière *et al.*, "Feedback on a publicly distributed image database: The Messidor database," *Image Analysis and Stereology*, vol. 33, no. 3, pp. 231–234, 2014, doi: 10.5566/ias.1155.

[35] L. Giancardo *et al.*, "Exudate-based diabetic macular edema detection in fundus images using publicly available datasets," *Medical Image Analysis*, vol. 16, no. 1, pp. 216–226, 2012, doi: 10.1016/j.media.2011.07.004.

[36] "Ocular disease intelligent recognition (ODIR)," *Kaggle*, 2020. http://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k (accessed Jun. 01, 2024).

[37] "Retina dataset," *Kaggle*, 2019. http://www.kaggle.com/datasets/jr2ngb/cataractdataset (accessed Jun. 01, 2024).

[38] M. A. Syarifah, A. Bustamam, and P. P. Tampubolon, "Cataract classification based on fundus image using an optimized convolution neural network with lookahead optimizer," *AIP Conference Proceedings*, vol. 2296, p. 020034, 2020, doi: 10.1063/5.0030744.

[39] R. Sarki, K. Ahmed, H. Wang, and Y. Zhang, "Automated detection of mild and multi-class diabetic eye diseases using deep learning," *Health Information Science and Systems*, vol. 8, no. 1, pp. 1–9, 2020, doi: 10.1007/s13755-020-00125-5.

[40] R. Sarki, K. Ahmed, H. Wang, S. Michalska, and Y. Zhang, "Early detection of diabetic eye disease from fundus images with deep learning," in *Databases Theory and Applications: 31st Australasian Database Conference, ADC*, 2020, pp. 234–241, doi: 10.1007/978-3-030-39469-1_20.

[41] F. Grassmann *et al.*, "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, 2018, doi: 10.1016/j.ophtha.2018.02.037.

[42] 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," *Applied Sciences (Switzerland)*, vol. 9, no. 14, p. 2789, 2019, doi: 10.3390/app9142789.

[43] N. Theera-Umpon, I. Poonkasem, S. Auephanwiriyakul, and D. Patikulsila, "Hard exudate detection in retinal fundus images using supervised learning," *Neural Computing and Applications*, vol. 32, no. 17, pp. 13079–13096, 2020, doi: 10.1007/s00521-019-04402-7.

[44] 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, doi: 10.1109/ACCESS.2019.2903171.

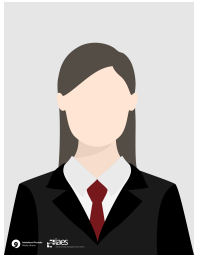
[45] J. de la Torre, A. Valls, and D. Puig, "A deep learning interpretable classifier for diabetic retinopathy disease grading," *Neurocomputing*, vol. 396, pp. 465–476, 2020, doi: 10.1016/j.neucom.2018.07.102.

[46] 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 and Applications*, vol. 79, no. 41–42, pp. 31595–31623, 2020, doi: 10.1007/s11042-020-09630-x.

[47] T. Araújo *et al.*, "DR—GRADUATE: uncertainty-aware deep learning-based diabetic retinopathy grading in eye fundus images," *Medical Image Analysis*, vol. 63, p. 101715, Jul. 2020, doi: 10.1016/j.media.2020.101715.

[48] M. N. Bajwa *et al.*, "Two-stage framework for optic disc localization and glaucoma classification in retinal fundus images using deep learning," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, p. 136, Dec. 2019, doi: 10.1186/s12911-019-0842-8.
[49] Y. Jiang *et al.*, "JointRCNN: a region-based convolutional neural network for optic disc and cup segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 2, pp. 335–343, Feb. 2020, doi: 10.1109/TBME.2019.2913211.

BIOGRAPHIES OF AUTHORS



Arwa Albelaihi is currently pursuing a Ph.D. in data science at the University of Leeds, UK, where she has been actively involved in research since 2023. Concurrently, she serves as a lecturer at the Department of Information Technology within the College of Computer at Qassim University, Saudi Arabia. She holds both B.S. and M.S. degrees from the Information Technology Department at Qassim University. Her research focuses on areas such as reinforcement learning, machine learning, and deep neural networks. For further communication, she can be contacted at email: a.albelaihi@qu.edu.sa or Scaaalb@leeds.ac.uk.

Dina M. Ibrahim has been an associate professor at the Information Technology Department, College of Computer, Qassim University, KSA, since 2015. In addition, she works as an assistant professor at the Computers and Control Engineering Department, Faculty of Engineering, Tanta University, Egypt. She was born in the United Arab Emirates, and her B.Sc., M.Sc., and Ph.D. degrees were obtained from the Computers and Control Engineering Department, Faculty of Engineering, Tanta University, in 2002, 2008, and 2014, respectively. She worked as a consultant engineer, database administrator, and vice manager on the Management Information Systems (MIS) Project, at Tanta University, Egypt, from 2008 until 2014. Her research interests include networking, wireless communications, machine learning, security, and the internet of things. She has published more than 60 articles in various refereed international journals and conferences. She has been serving as a reviewer in the Wireless Network (WINE) Journal since 2015. She has also served as a co-chair of the International Technical Committee for the Middle East Region of the ICCMIT conference since 2020. She can be contacted at email: d.hussein@qu.edu.sa or dina.mahmoud@f-eng.tanta.edu.eg.