



**ARTICLE**

## Deep Learning Empowered Diagnosis of Diabetic Retinopathy

Mustafa Youldash<sup>1</sup>, Atta Rahman<sup>2,\*</sup>, Manar Alsayed<sup>1</sup>, Abrar Sebiany<sup>1</sup>, Joury Alzayat<sup>1</sup>, Noor Aljishi<sup>1</sup>, Ghaida Alshammari<sup>1</sup> and Mona Alqahtani<sup>1</sup>

<sup>1</sup>Department of Computer Engineering, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam, 31441, Saudi Arabia

<sup>2</sup>Department of Computer Science, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam, 31441, Saudi Arabia

\*Corresponding Author: Atta Rahman. Email: aaurahman@iau.edu.sa

Received: 14 September 2024 Accepted: 20 December 2024 Published: 23 January 2025

### ABSTRACT

Diabetic retinopathy (DR) is a complication of diabetes that can lead to reduced vision or even blindness if left untreated. Therefore, early and accurate detection of this disease is crucial for diabetic patients to prevent vision loss. This study aims to develop a deep-learning approach for the early and precise diagnosis of DR, as manual detection can be time-consuming, costly, and prone to human error. The classification task is divided into two groups for binary classification: patients with DR (diagnoses 1–4) and those without DR (diagnosis 0). For multi-class classification, the categories are no DR, mild DR, moderate DR, severe DR, and proliferative diabetic retinopathy (PDR). To achieve this, the proposed model utilizes two pre-trained convolutional neural networks (CNNs), specifically ResNet50 and DenseNet-121. Both models were trained and evaluated on fundus images sourced from the widely recognized APTOS dataset, a publicly available resource, and achieved impressive training and testing accuracies. For binary classification, DenseNet-121 achieved an accuracy of 98.1%, while ResNet50 attained an accuracy of 97.4%. In multi-class classification for DR, DenseNet-121 achieved an accuracy of 82.0%, and ResNet50 reached an accuracy of 80.8%. The results are promising and comparable to state-of-the-art techniques in the literature for both binary and multi-label classification of DR.

### KEYWORDS

Diabetic retinopathy; deep learning; fundus images; DenseNet-121; convolutional neural networks; ResNet50

## 1 Introduction

Diabetes is a serious medical condition characterized by elevated blood sugar levels, primarily resulting from the pancreas's inability to produce sufficient insulin. Uncontrolled high blood sugar can lead to severe health complications, such as heart disease, kidney failure, nerve damage, and ultimately, blindness [1]. A significant complication of diabetes is diabetic retinopathy (DR), a condition that damages the delicate blood vessels in the retina of the eye. When these small blood vessels become blocked due to high blood sugar, they may leak fluids or bleed, leading to serious damage to the retina [2].



To compensate for this blockage, the retina attempts to develop new blood vessels; however, these vessels often fail to function properly and may result in bleeding or leakage. According to estimates from the World Health Organization (WHO), diabetic retinopathy contributes to around 37 million cases of blindness worldwide. Proactively managing diabetes is essential to prevent such severe outcomes, particularly in Middle Eastern regions [3].

Current approaches to treating diabetic retinopathy (DR) mainly aim to slow down or prevent further vision deterioration, highlighting the critical importance of routine eye examinations facilitated by advanced computational systems. Artificial intelligence (AI) and its associated technologies have demonstrated exceptional potential in accurately diagnosing diseases through medical imaging. For instance, recent research has successfully linked the biological factors of DR with cutting-edge AI models [4].

It is critical to identify and treat Vision-Threatening Diabetic Retinopathy (VTDR) as early as possible to effectively reduce the risk of irreversible vision loss. This proposed study aims to make a significant impact by facilitating early detection and proactive prevention of DR. AI has established itself as a vital asset in healthcare and medicine, fueled by innovations in computing power, deep learning (DL) algorithms, and the explosion of big data. By leveraging training datasets, learning algorithms enhances accuracy, minimize the chances of misdiagnosis, and significantly reduce time, effort, and costs. Thus, the development of a sophisticated DL model dedicated to the early diagnosis of diabetic retinopathy, assessing its severity, and delivering tailored treatment plans is not just beneficial—it is imperative for advancing patient care and preserving vision [5].

This study utilizes the latest machine learning technologies to improve diabetic retinopathy (DR) detection in retinal image analysis by leveraging two pre-trained Convolutional Neural Network (CNN) architectures: ResNet50 and DenseNet-121. ResNet50, a widely recognized deep CNN architecture with 50 layers, demonstrates remarkable accuracy and efficiency in analyzing retinal images [6]. Meanwhile, DenseNet-121, a more complex network with 121 layers, excels in extracting intricate features from these images, allowing it to identify subtle abnormalities that may indicate DR [7].

Upon reviewing the literature, it has become evident that there are relatively few studies conducted on DR detection using artificial intelligence (AI) and deep learning in the Kingdom of Saudi Arabia (KSA). However, several studies related to diabetes detection indicate a clear and increasing trend in disease prevalence after the COVID-19 pandemic, largely due to lifestyle changes [8–10]. This evidence suggests a significant likelihood of an increase in DR cases across the kingdom. This study aims to establish an efficient mechanism for the early detection of the disease, enabling timely remedial actions before it is too late.

This study is committed to significantly advancing the accuracy of disease diagnosis through the power of deep learning, with the primary goal of enhancing patient care. We introduce an innovative diagnostic approach for diabetic retinopathy (DR) that utilizes state-of-the-art deep learning models. Our approach strategically employs pre-trained models, integrating two essential classification tasks: a binary classification to confirm the presence or absence of DR, and a multi-class classification to determine disease severity across five distinct categories: no DR, mild, moderate, severe, and proliferative. This dual strategy empowers us to achieve remarkable detection accuracy for DR. Rigorous experiments were carried out using the popular and authentic APTOS 2019 dataset, made available by the Asia Pacific Tele-Ophthalmology Society (APTOS) via Kaggle for their 2019 competition. Our comprehensive results indicate that our proposed method surpasses existing techniques, based on a range of evaluation metrics from prior studies. By enabling earlier detection of DR, our pioneering approach holds promise for substantially improving patient outcomes.

The motivation driving our study is twofold and highly impactful. First, we seek to seamlessly integrate binary classification (assessing presence vs. absence) with multi-class classification (reflecting different degrees of disease severity). This innovative approach directly addresses a significant gap in current research, which has primarily concentrated on isolated aspects. Second, we recognize the urgent necessity to optimize hyperparameters within our transfer learning models, enabling us to achieve substantial improvements over existing models. This crucial optimization is at the forefront of our research, ensuring a powerful and effective solution for diabetic retinopathy (DR) detection. Furthermore, our study aims to make a meaningful contribution to the Kingdom of Saudi Arabia's (KSA) vision for improved healthcare facilities by 2030.

The rest of the paper is organized as follows: [Section 2](#) provides related work on DR. [Section 3](#) provides insights into data acquisition and pre-processing utilized in the study. Model development, training, and testing steps are provided in [Section 4](#). The model is evaluated and analyzed in [Section 5](#), and discussion is made in [Section 6](#), while [Section 7](#) concludes the paper.

## 2 Related Work

Innovative deep-learning techniques are making significant strides in the diagnosis of diabetic retinopathy. This section highlights cutting-edge developments in the early detection of this condition, focusing on the powerful capabilities of deep learning models and methodologies that promise to transform patient outcomes.

Qummar et al. [11] designed an ensemble of deep convolutional neural network (CNN) models to improve the accuracy of classifying various stages of diabetic retinopathy (DR) using retinal images. This research specifically addressed the limitations of current models in detecting early stages of DR and utilized the Kaggle Diabetic Retinopathy Detection dataset, which is publicly available. To accomplish this, the team developed an ensemble comprising five CNN models: ResNet50, InceptionV3, Xception, DenseNet-121, and DenseNet-169. Preprocessing and data augmentation steps involved resizing images to  $786 \times 512$  pixels, generating five random  $512 \times 512$  patches per image, applying flipping, rotating by 90 degrees, and mean normalization. The results showed that the proposed model achieved 80.8% accuracy, 51.5% recall, 86.72% specificity, 63.85% precision, and an F1-score of 53.74%. Overall, the model demonstrated its effectiveness in detecting all DR stages and outperformed existing techniques on the Kaggle dataset. A study by Yasashvini et al. [12] proposes a deep learning model for automatically classifying the stages of diabetic retinopathy (DR) from fundus images. The research evaluates three deep learning models: a convolutional neural network (CNN), a hybrid CNN with ResNet, and a hybrid CNN with DenseNet. All three models were trained on the same training dataset and evaluated on the same test dataset. The study utilized the Kaggle dataset APTOS 2019 Blindness Detection, which comprises approximately 5500 fundus images labeled with one of the five DR stages: stage zero (no DR), stage one (mild DR), stage two (moderate DR), stage three (severe DR), and stage four (proliferative DR). Among the models evaluated, the hybrid CNN with DenseNet achieved the highest performance, with an accuracy of 96.22%. CNN and the hybrid CNN with ResNet recorded accuracies of 75.61% and 93.18%, respectively. However, the study's scope was limited by the small sample size. Additionally, the authors did not conduct cross-validation to assess the generalization performance of their models.

In a study conducted by Ali et al. [13], the authors proposed a novel deep learning (DL) model for the early classification of diabetic retinopathy (DR). This model utilizes color fundus images and focuses on the most critical aspects of the disease while excluding irrelevant factors to ensure high recognition accuracy. The study utilized the Kaggle Diabetic Retinopathy Detection dataset, which

was divided into 80% for training and 20% for testing. A hybrid deep-learning classification approach was developed using ResNet50 and InceptionV3. ResNet50 was employed to enhance performance without introducing excessive complexity, while InceptionV3 was used to reduce parameters and computational costs through the application of different filter sizes. The images underwent preprocessing that included histogram equalization, intensity normalization, and data augmentation. The results of the study were impressive, achieving an accuracy of 96.85%, sensitivity of 99.28%, specificity of 98.92%, precision of 96.46%, and an F1-score of 98.65%.

A recent study by Alwakid et al. [14] developed a sophisticated deep learning (DL) model to accurately classify the five stages of diabetic retinopathy (DR). The data used for this study were obtained from two different sources: the Diabetic Retinopathy Dataset (DDR) and the APTOS dataset. Both datasets consist of high-resolution retinal images that represent a range of DR stages, from stage 0 (no DR manifestations) to stage 4 (proliferative DR). The study implemented a bifurcated methodological approach, introducing two variants: one with image augmentation and one without. The DenseNet-121 model was employed to identify the five stages of DR within both the APTOS and DDR datasets, achieving exceptional performance when compared to other methods. The methodology differentiated between two enhancement scenarios: case 1 utilized Contrast Limited Adaptive Histogram Equalization (CLAHE) and Enhanced Super Resolution Generative Adversarial Networks (ESRGAN), while case 2 proceeded without any enhancements. The model, leveraging the DenseNet-121 architecture, demonstrated accuracy rates of 98.7% for case 1 and 81.23% for case 2 on the APTOS dataset. In parallel, the DDR dataset achieved accuracy rates of 79.6% for case 1 and 79.2% for case 2.

A study conducted by Shakibania et al. [15] introduced a reliable method for screening and grading the stages of diabetic retinopathy, demonstrating significant potential for enhancing clinical decision-making and patient care. The research presents a deep learning approach for detecting and grading diabetic retinopathy using Discriminative Restricted Boltzmann Machines (DRBMs) with Softmax layers, all based on a single fundus retinal image. By employing transfer learning, the model utilizes two advanced pre-trained models, ResNet50 and EfficientNetB0, as feature extractors, which are then fine-tuned with a new dataset. This dataset was divided into training (70%), validation (10%), and testing (20%) subsets. To address the issue of imbalanced data, the complement cross-entropy (CCE) loss function was implemented. The model was trained using three datasets: APTOS 2019 Blindness Detection, MESSIDOR2, and the Indian Diabetic Retinopathy Image Dataset (IDRID). As a result, the proposed method achieved an accuracy of 98.50%, a sensitivity of 99.46%, and a specificity of 97.51% in binary classification. For stage grading (non-binary classification), the model reached an accuracy of 89.60%, a sensitivity of 89.60%, and a specificity of 97.72%, as well as a quadratic weighted kappa of 93.00%.

Research by Kumari et al. [16] proposed an automated decision-making approach for identifying diabetic retinopathy (DR) using a ResNet-based feed-forward neural network methodology. The study utilized a dual-image multi-layer mapping technique and evaluated the effectiveness of the approach on two datasets: the APTOS 2019 Blindness Detection dataset and EyePACS. The datasets were split into 80% for training, which was used to optimize the model, and 20% for testing, which was used for evaluation. Fundus images from both datasets were preprocessed to create both colored and black-and-white versions, which were subsequently used for DR stage classification and detection. Feature extraction was performed using the ResNet feed-forward neural network, and missing data were addressed through vectorization. A feed-forward neural network classifier was then utilized to categorize images based on the presence of diabetic retinopathy. Methodology evaluation was conducted using 10-fold cross-validation, and the results were stratified according to image quality. For

high-quality images, the testing accuracy, sensitivity, and specificity were found to be 98.9%, 98.7%, and 98.3%, respectively. In contrast, these metrics were 94.9%, 93.6%, and 93.2% for low-quality images.

A study referenced in [17] examined the role of hybrid intelligence by utilizing various machine learning models, including support vector machines (SVM), k-nearest neighbors (kNN), and histogram gradient boosting (HGB). The findings revealed that SVM achieved the highest accuracy at 96.9%. Meanwhile, another study in [18] utilized convolutional neural network (CNN) architectures for the binary classification of diabetic retinopathy (DR), attaining a maximum accuracy of 94.55%.

The literature review indicates that the use of artificial intelligence (AI) for diagnosing diabetic retinopathy (DR) is a leading area of research. Timely and accurate diagnostics can significantly aid in mitigating the severity of the disease. Additionally, there is still potential for improvement in various evaluation metrics, including accuracy, precision, sensitivity (also known as recall), specificity, and F1-score, as well as in the exploration of different datasets. A summary of the literature review is provided in Table 1.

**Table 1:** Summary of literature review

Ref.	Year	Technique/s	Dataset/s	Accuracy	Task
[11]	2019	Ensemble of five models: (Resnet50, InceptionV3, Xception, Dense-121, and Dense169)	Kaggle: Diabetic Retinopathy Detection	80.8%	Multi
[12]	2021	CNN CNN with ResNet CNN with DenseNet	APTOS 2019	75.61% 93.18% 96.22%	Multi
[13]	2023	ResNet50 and InceptionV3 for feature extraction + CNN for classification	Kaggle: Diabetic Retinopathy Detection	96.85%	Multi
[14]	2023	DenseNet-121rendered	APTOS 2019 DDR	98.7% 79.6%	Binary
[15]	2023	Deep dual-branch model (ResNet50 and EfficientNetB0)	APTOS 2019 + Messidor-2 + IDRiD) stage grading APTOS 2019 + Messidor-2 + IDRiD) binary classification	89.60% 98.50%	Multi Binary
[17]	2024	SVN, k-nearest neighbors, and Histogram Gradient Boosting	APTOS	96.9%	Binary
[18]	2024	MobileNet CNN	APTOS	94.55%	Binary

### 3 Data Acquisition and Pre-processing

#### 3.1 Data Acquisition

This study utilized retinal images from the APTOS 2019 Blindness Detection dataset, an open competition on Kaggle [19]. The dataset comprises 3662 retinal images collected from individuals residing in rural areas of India. The images may contain artifacts, blurriness, poor brightness, and other issues. Trained doctors labeled the images and categorized the extent of blindness into five classes ranging from 0 to 4: No DR (Class 0), Mild (Class 1), Moderate (Class 2), Severe (Class 3), and Proliferative DR (Class 4). As a result of adapting this classification into a binary system, we divided the data into two groups: those with DR (diagnoses 1-4) and those without DR (diagnosis 0). The main reason behind this mapping is the highly imbalanced nature of the dataset. We also conducted our analysis in both the standard multi-class and an altered binary format. This allowed us to thoroughly evaluate the dataset's flexibility and reliability across multiple diagnostic approaches. Fig. 1 demonstrates the class distribution of the binary classification, and Fig. 2 demonstrates the multi-classification dataset.

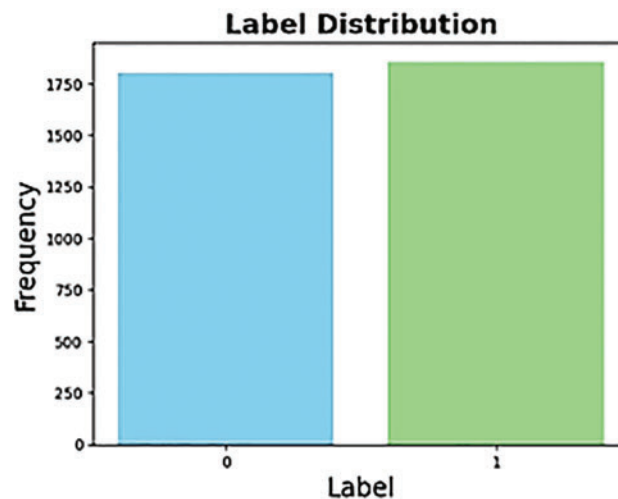
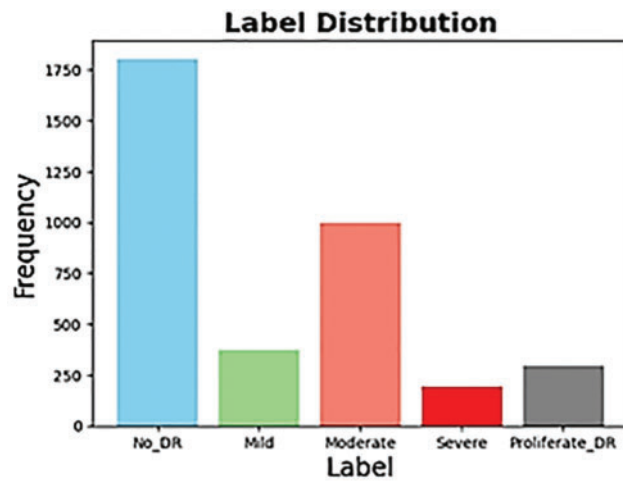


Figure 1: Binary label distribution

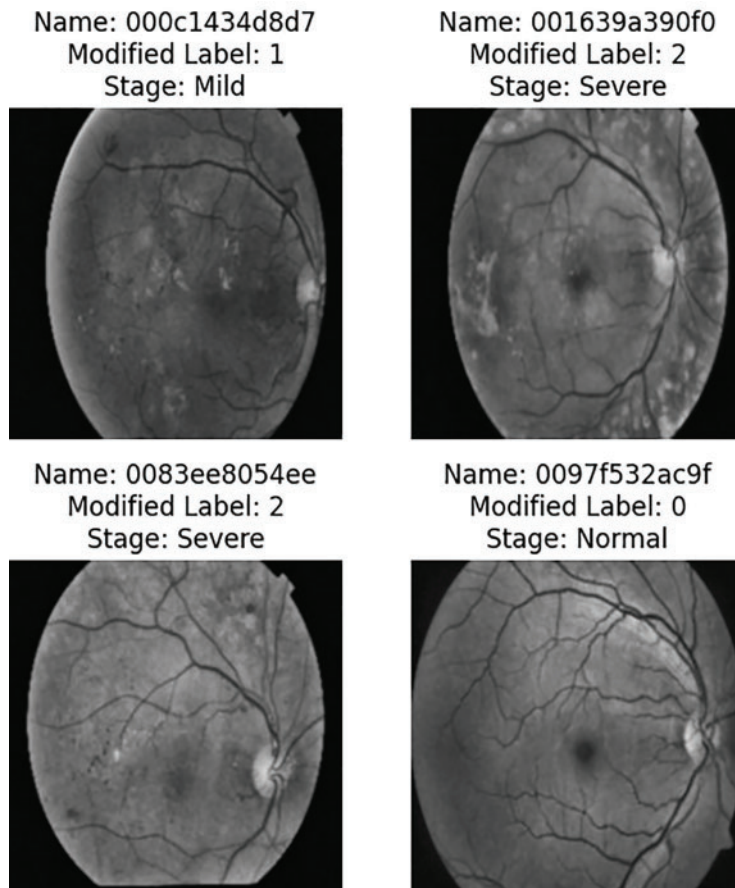
#### 3.2 Data Preprocessing

The quality of the images in the dataset could be improved through a preprocessing phase. Since the images are noisy and contain artifacts, such as being out of focus, having too much exposure, or having extra lighting or a black background, preprocessing is necessary to get them into a standard format. As the first step of preprocessing, all images in the dataset are resized to a uniform size of  $224 \times 224$  pixels to match the size of the input image used in the extraction process. In the second preprocessing step, contrast adjustment techniques are applied to the input images to increase the contrast. Fundus images are typically available as RGB color images. Initially, only the green channel of the image is extracted during preprocessing out of the three red, blue, and green channels. Additionally, to enhance the contrast of the green channel image at the tile level, the Contrast-Limited Adaptive Histogram Equalization (CLAHE) process is applied. Afterward, Gaussian filtering is used to reduce the noise in the images. Fig. 3 displays the preprocessed images.





**Figure 2:** Multi label distribution



**Figure 3:** Images obtained after pre-processing

## 4 Model Development and Training

ResNet50 and DenseNet-121 stand out as top choices for transfer learning models in the detection of diabetic retinopathy [11–15]. These models have been chosen for their remarkable effectiveness in similar challenges, particularly in the field of medical image diagnostics. Their performance has consistently outshined other models across numerous datasets dedicated to diabetic retinopathy detection [11–15]. This impressive capability stems from their exceptional proficiency in extracting critical features from medical images, which are vital for achieving precise disease detection and diagnostics.

### 4.1 Model Development

This section provides insight into the proposed transfer learning models investigated in the current study. Moreover, the values of optimized hyperparameters are also mentioned that are obtained after several trials on the proposed models.

#### 4.1.1 ResNet50

ResNet50 is a type of ResNet architecture that is particularly noteworthy for having 50 layers, making it exceptionally deep and ideal for tasks that involve image classification [20]. The key innovation is the use of residual connections, which help address the vanishing gradient problem that often affects deep networks. With the help of shortcut connections, ResNet50 makes it possible to add convolutional layers seamlessly, which allows for deeper architecture without compromising performance [21]. The architecture is divided into four main segments, including initial convolutional layers for feature extraction, identity, and convolutional blocks for further transformation with feature processing, and fully connected layers for classification [21]. ResNet50's 50 bottleneck residual blocks provide a robust solution to the challenges of vanishing gradients, ensuring efficient training and exceptional performance in complex visual recognition tasks. Fig. 4 refers to the architecture of ResNet50. This study utilized ResNet50, there were two separate trials conducted, one involving binary classification and the other having five classes. In this regard, we have employed the Cross-Entropy loss function and used the Adam optimizer (optim.Adam) to optimize the model parameters during training. The Adam optimizer is an adaptive optimization algorithm that adjusts the learning rates for each parameter individually.

#### 4.1.2 DenseNet-121

In this study, DenseNet-121 has been implemented as the second model. This choice was made to overcome the common obstacles of training deep neural networks such as convergence rate. The architecture of DenseNet is characterized by a feed-forward connection between every layer, resulting in dense connectivity among feature maps. This design enables the efficient transmission of information from earlier to later layers, thereby reducing the total number of parameters required. Additionally, it facilitates the reuse of features and promotes gradient flow during training [22].

DenseNet-121 is a deep convolutional neural network architecture comprising 121 layers and four dense blocks. A dense block in the context of DenseNet-121 refers to a cohesive set of layers where each layer is intricately connected to every other layer within that block. This dense connectivity pattern



promotes feature reuse and enhances information flow throughout the network. To ensure a seamless transition between these dense blocks, DenseNet-121 incorporates transition layers. Decreasing the spatial dimensions of the feature maps is a vital function carried out by the transition layers, facilitating a smooth transition from one dense block to another and contributing significantly to the network's overall efficiency [23]. Fig. 5 refers to the architecture of DenseNet-121.

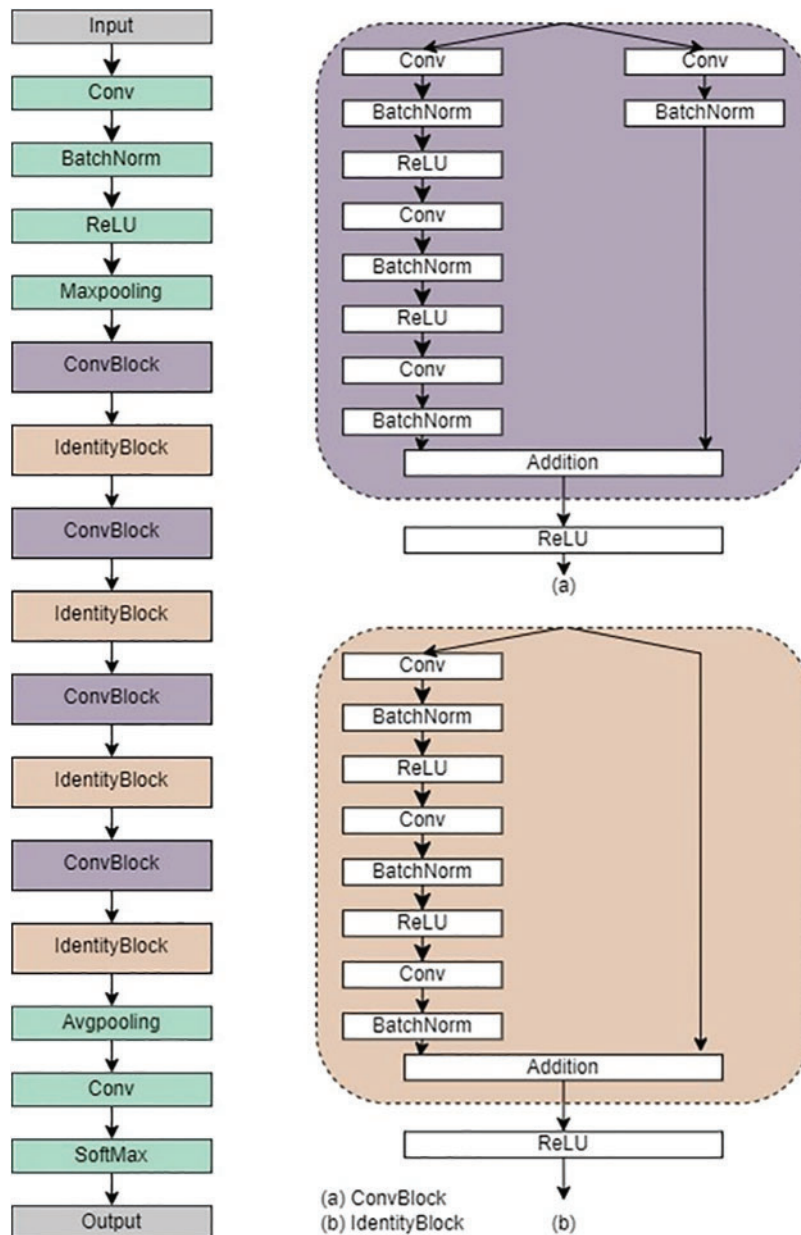
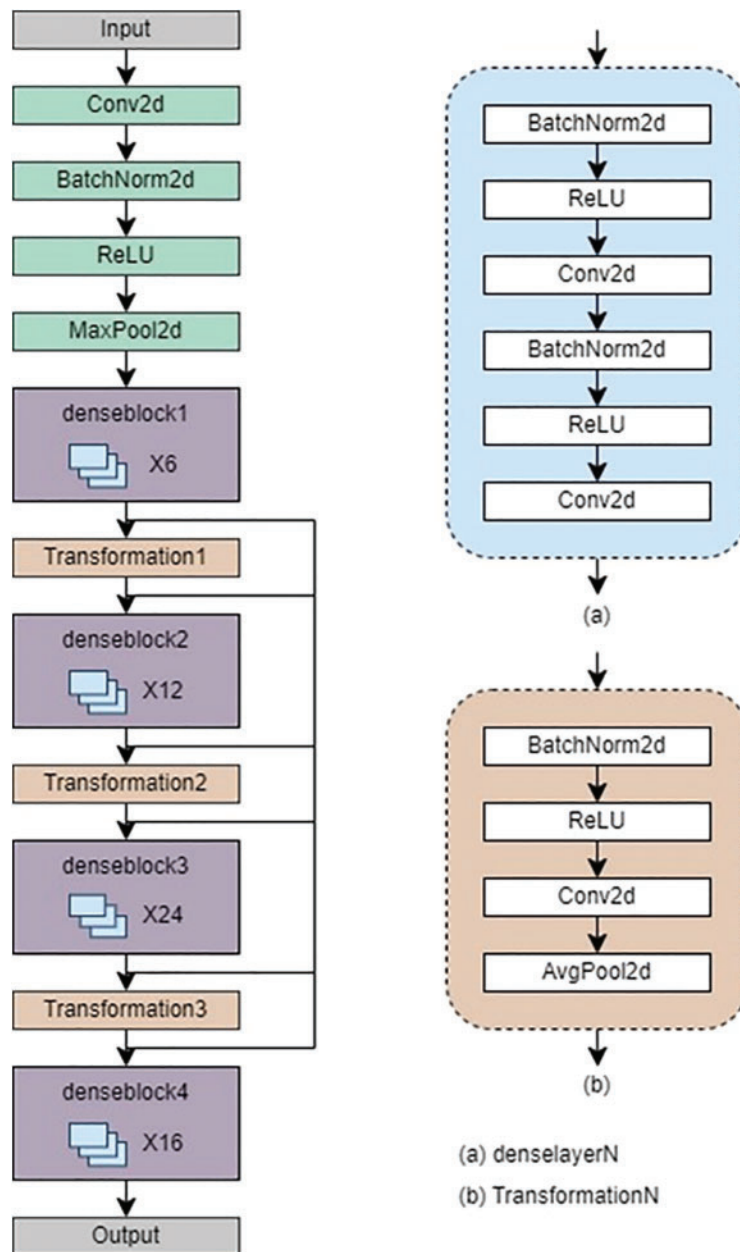


Figure 4: Architecture of ResNet50



**Figure 5:** Architecture of DenseNet-121

#### 4.2 Model Training

To initiate the training process, the input data is divided into training, testing, and validation sets using a 70%–20%–10% split with the `train_test_split` function from the `sklearn.model_selection` module. This separation is crucial in assessing the model's efficacy on unobserved data. The neural network model is formulated with a cross-entropy loss function and the Adam optimizer.

The data is transformed into TensorFlow format, then normalized and resized using a series of transformations. These include resizing the images to (224, and 224) and normalizing them with specific mean and standard deviation values. During the training process, the loop runs for a predetermined number of epochs, typically 30. The model is set to training mode, and the optimizer is initialized to start the forward pass. The loss is calculated using the cross-entropy criterion, and the backward pass computes the gradients. The optimizer then updates the model weights. Once the training phase is complete, the model switches to evaluation mode, and its performance is assessed on the validation set. If the validation accuracy improves, the best model weights are saved.

## 5 Evaluation and Analysis

### 5.1 Model Development

Five key metrics were utilized to evaluate the performance of the classification model, ensuring a comprehensive analysis: accuracy, precision, sensitivity (also referred to as recall), specificity, and the F1-score. Accuracy represents the fraction of correct predictions compared to the total observations. Precision measures the proportion of correctly predicted positive instances out of all positive predictions. Sensitivity, or recall (also known as the True Positive Rate), evaluates the proportion of correctly identified positive instances relative to the total number of actual positives. Specificity assesses the percentage of correctly identified negative predictions among all negatives. Finally, the F1-score combines precision and recall into a single value, providing a balanced measure between these two metrics [24].

#### 5.1.1 Accuracy

Mathematically, accuracy can be represented as [25]. It is referred to as the ratio of true positive and true negative to all cases that are true positive, true negative, false positive, and false negative, respectively.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \quad (1)$$

#### 5.1.2 Precision, Sensitivity (Recall), Specificity, and F1-Score

Mathematically, the precision, sensitivity, specificity, and F1-score can be represented as given in Eqs. (2)–(5), respectively [26].

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (2)$$

$$\text{Sensitivity (Recall)} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (3)$$

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} \quad (4)$$

$$\text{F1 - score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

In medical diagnostics and clinical decision support systems, sensitivity and specificity are crucial factors in addition to accuracy and precision. This is primarily because they effectively report the true

positive and true negative rates, which demonstrate the efficacy of the proposed approach in medical diagnostics [27,28].

## 5.2 Performance Evaluation

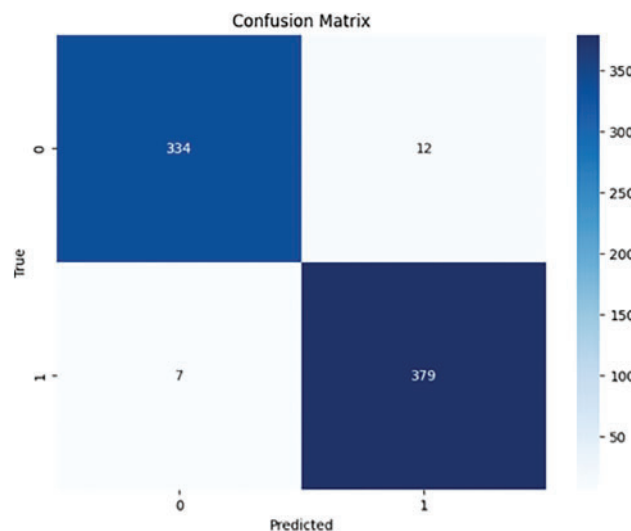
In this research, we employed pre-trained models, ResNet50 and DenseNet-121, to accurately identify diabetic retinopathy. After training the models with the APTOS dataset for 30 epochs, we evaluated their performance for both multi-class and binary-class classification. Despite an unbalanced dataset, we refrained from using specific preprocessing techniques, allowing the data to dictate the outcome. Both ResNet50 and DenseNet-121 achieved impressive accuracy in binary classification, demonstrating the effectiveness of deep learning models for this task. ResNet50 displayed an accuracy of 97.4%, precision of 97.4%, sensitivity (recall) of 97.4, specificity of 96.5%, and F1-score of 97.4%, while DenseNet-121 exhibited an accuracy of 98.1%, precision of 98.1%, sensitivity (recall) of 98.1%, specificity of 99.7%, and F1-score of 98.1%.

Table 2 shows the accuracy, precision, sensitivity (recall), specificity, and F1-score for the binary classification scenario.

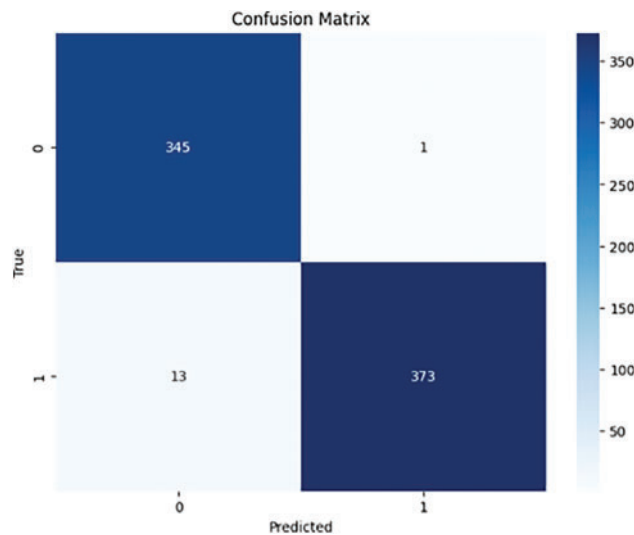
**Table 2:** Models performance in binary classification

Model	Accuracy	Precision	Sensitivity (Recall)	Specificity	F1-score
ResNet50	0.974	0.974	0.974	0.965	0.974
DenseNet-121	0.981	0.981	0.981	0.997	0.981

Figs. 6 and 7 demonstrate the confusion matrices for the binary class for ResNet50 and DenseNet-121, respectively.



**Figure 6:** ResNet50 binary confusion matrix



**Figure 7:** DenseNet-121 binary confusion

Our analysis of the comparison between ResNet50 and DenseNet-121 for the multi-class classification of diabetic retinopathy indicated that both models performed well. ResNet50 achieved an accuracy of 80.8%, whereas DenseNet-121 achieved a slightly higher accuracy of 82.0%. Additionally, ResNet50 obtained 79.7% precision, 80.8% sensitivity (recall), 62.1% specificity, and 79.9% F1-score. On the other hand, DenseNet-121 exhibited 81.0% precision, 82.0% sensitivity (recall), 62.9% specificity, and 81.3% F1-score. [Table 3](#) shows the accuracy, precision, sensitivity (recall), specificity, and F1-score for multi-class classification. [Figs. 8](#) and [9](#) contain the confusion matrices for the multi-class.

**Table 3:** Models performance in multi-class classification

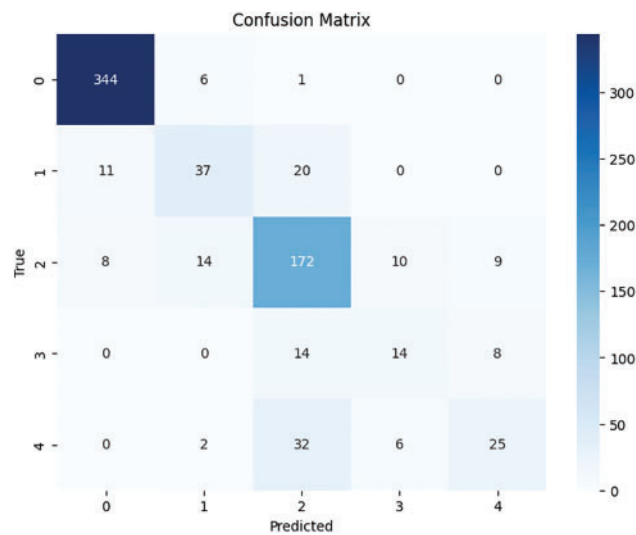
Model	Accuracy	Precision	Sensitivity (Recall)	Specificity	F1-score
ResNet50	0.808	0.797	0.808	0.621	0.799
DenseNet-121	0.820	0.810	0.820	0.629	0.813

### 5.3 Comparison with State-of-the-Art

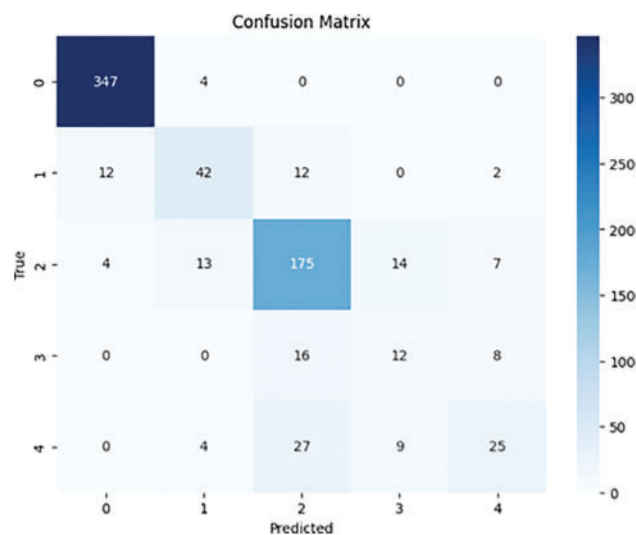
The proposed study is compared with a scheme from the literature that uses the same dataset, called APTOS. A separate comparison is made for binary and multi-classification.

#### 5.3.1 Binary Classification

First, we will analyze our scheme against the state-of-the-art methods for binary classification of diabetic retinopathy (DR). For this purpose, we selected recent studies published in 2023 [[14,15](#)] and 2024 [[17,18](#)] that achieved the highest metrics using the same dataset. [Table 4](#) presents a comparison of these studies in terms of accuracy, precision, recall (sensitivity), specificity, and F1-score. However, some studies [[14,15](#)] do not provide value for precision and F1-score.



**Figure 8:** ResNet50 multi-class confusion matrix



**Figure 9:** DenseNet-121 multi-class confusion matrix

The proposed scheme significantly outperforms SVM [17] and MobileNet CNN [18], achieving higher accuracy by 1.1% and 3.6%, respectively. This advantage extends to essential metrics such as sensitivity, F1-score, and specificity, reinforcing its effectiveness. While schemes [14] and [15] that incorporated DenseNet-121 and ResNet50 with EfficientNetB0 showed slight improvements in accuracy—0.6% and 0.4% respectively—the proposed scheme excels in specificity, surpassing them by 1.7% and 2.5%. Moreover, the sensitivity achieved by the proposed scheme is on par with that of [14] and [15], demonstrating their competitiveness across the board.



**Table 4:** Comparison of binary classification

Study	Model	Accuracy	Precision	Sensitivity (Recall)	Specificity	F1-score
Proposed	ResNet50	0.974	0.974	0.974	0.965	0.974
	DenseNet-121	0.981	0.981	0.981	0.997	0.981
[14] 2023	DenseNet-121	0.987	NA	0.98	0.98	0.98
[15] 2023	Deep dual-branch model (ResNet50 and EfficientNetB0)	0.985	NA	0.984	0.975	NA
[17] 2024	SVM	0.969	0.969	0.969	0.969	0.969
[18] 2024	MobileNet CNN	0.945	0.945	0.945	0.945	0.945

### 5.3.2 Multi Classification

The study outlined in [11] investigates an impressive ensemble of five models—ResNet50, InceptionV3, Xception, DenseNet-121, and DenseNet-169—targeting multi-class classification and achieving a remarkable accuracy of 80.8%. However, the proposed scheme leverages two of these techniques without relying on an ensemble approach, surpassing the original results by a significant 1.2% accuracy, specifically with DenseNet-121. In addition to this, the proposed scheme shows notable improvements in key performance metrics, exceeding the previous study by 1.62% in precision, 3.05% in sensitivity, and 2.786% in F1-score. While the study in [11] does hold an edge in specificity, outperforming the proposed scheme by 2.38%, the overall findings underscore the robustness and efficacy of the proposed approach. These compelling results are detailed in [Table 5](#).

**Table 5:** Comparison of multi classification

Study	Model	Accuracy	Precision	Sensitivity (Recall)	Specificity	F1-score
Proposed	ResNet50	0.808	0.797	0.808	0.621	0.799
	DenseNet-121	0.820	0.810	0.820	0.629	0.813
[11]	Ensemble transfer learning model	0.808	0.638	0.515	0.867	0.537

This study emphasizes the compelling benefits of using transfer learning with pre-trained models for interpretable diabetic retinopathy classification. Instead of merely striving for the highest accuracy, our findings highlight the critical importance of interpretability, especially in the medical field where understanding decisions can be lifesaving. While our proposed model may not surpass the performance of the latest pre-trained models, particularly in multi-class scenarios, the ability to comprehend the model’s decision-making process is vital. The slight decrease in performance can be attributed to the common issue of class imbalance in medical datasets. Nonetheless, this approach effectively balances performance with interpretability, making it a superior choice for real-world applications where understanding a model’s reasoning holds equal weight to its predictions. A notable limitation of our study is its reliance on a single dataset. To bolster the model’s adaptability and overall effectiveness, upcoming work should expand to include diverse datasets and leverage data augmentation techniques [29,30].

## 6 Discussion

The findings from this study significantly contribute to the existing literature on the use of deep learning models for detecting and classifying diabetic retinopathy (DR). Our innovative approach has practical implications for enhancing healthcare access and quality. By improving the accuracy of DR detection, our deep learning models can accelerate the diagnostic process, allowing for earlier interventions that can greatly improve patient outcomes. Integrating deep learning technologies into healthcare systems can alleviate the burden on medical facilities and ensure that patients receive timely eye care. This integration also provides healthcare professionals with advanced diagnostic tools, empowering them to make well-informed clinical decisions and tailor treatment plans to meet each patient's unique needs. In the study, we investigated the DenseNet-121 model, which achieved outstanding results in the binary classification task. Notably, the model delivered an accuracy of 98.1%, along with precision, sensitivity (recall), and F1-scores all at 98.1%, and a specificity of 99.7%. In the multi-class classification task, the DenseNet-121 model showed a precision of 81.0%, a sensitivity (recall) of 82.0%, a specificity of 62.9%, and an F1-score of 81.3%. These impressive results demonstrate the effectiveness and superiority of our proposed approach. Implementing this model could lead to significantly improved patient outcomes and reduced healthcare risks and costs, particularly regarding late-stage DR complications, thereby enhancing the quality of life for individuals with diabetes. However, to fully realize this potential, it is essential to discuss and address these critical aspects more comprehensively.

Our future research goal is to develop an innovative smartphone app for Apple and Android, as well as a user-friendly website, aimed at the rapid detection of diabetic retinopathy (DR) in hospitals and medical centers. This advanced clinical decision support system is intended for immediate implementation, enhancing the diagnostic process while significantly reducing the workload of healthcare professionals and accelerating timely patient intervention. By seamlessly integrating our predictive models into existing clinical workflows, this tool will empower early detection and effective management of DR, which is crucial for preventing the disease from advancing to more serious stages. This initiative is a vital contribution to the Saudi Vision 2030, focused on enhancing healthcare services around the kingdom. Additionally, we guarantee that our research meets contemporary standards for clinical decisions and actively promotes advancements in the field of ophthalmology [31,32].

### 6.1 Limitations

A key limitation of the research lies in its reliance on a single dataset. Although the APTOS 2019 dataset provided a strong foundation for training and evaluating the proposed models, the results may not be universally applicable, as other datasets could exhibit different image characteristics, demographic variables, or labeling standards. To enhance the robustness and versatility of the proposed models, future studies should consider incorporating multiple datasets and employing data augmentation techniques. In addition, ensemble learning models [33,34], and hybrid intelligence techniques [35,36] must be investigated especially for multi classification scenarios where there is still room for improvement. Moreover, the dataset is imbalanced, and the current study is in the same format (let the data speak). In the future, however, data balancing techniques such as Synthetic Minority Over Sampling Technique (SMOTE) with up and downsampling may also be investigated [37,38]. Moreover, the current study only focuses on critical yet single disease diagnostics from the retinal images. However, other retinal eye diseases can be associated with the same image. In the future, advanced studies can be conducted for joint diagnosis of multiple retinal eye diseases [39,40].

## 7 Conclusions

Numerous deep-learning techniques have shown great promise in the literature for diagnosing diabetic retinopathy. In this section, we explore the recent breakthroughs in the early detection of this serious condition, particularly through innovative deep-learning models and methodologies. This study highlights the groundbreaking potential of deep learning in identifying diabetic retinopathy from image datasets, emphasizing its capability to transform patient outcomes. The evidence demonstrates that deep learning can reliably and accurately classify diabetic retinopathy into distinct stages, providing a valuable and cost-effective resource for medical professionals. By utilizing two state-of-the-art convolutional neural network (CNN) architectures—ResNet50 and DenseNet-121—pre-trained on diverse and extensive datasets, we significantly boost diagnostic accuracy. Moreover, the application of various image enhancement techniques on retinal fundus images has been shown to further elevate the models' performance. The ResNet50 model achieved a remarkable accuracy of 97.4% in binary classification, while DenseNet-121 outperformed this with an accuracy of 98.1%. In the multi-class classification task, ResNet50 reached 80.8% accuracy, while DenseNet-121 exceeded this with an accuracy of 82.0%. While this framework focuses specifically on diabetic retinopathy diagnosis, its methods can readily extend to other retinal conditions, such as glaucoma and age-related macular degeneration (AMD). To fully harness the potential of data augmentation techniques, leveraging a range of diverse datasets is essential. Additionally, exploring advanced preprocessing methods and fine-tuning hyperparameters presents exciting opportunities to further enhance diagnostic performance, ultimately improving patient care.

**Acknowledgement:** We would like to acknowledge the collaborative effort between computer science and medical disciplines to conduct this study.

**Funding Statement:** The authors received no specific funding for this study.

**Author Contributions:** Conceptualization, Mustafa Youldash, and Atta Rahman; Data curation, Manar Alsayed, and Noor Aljishi; Formal analysis, Ghaida Alshammari; Methodology, Manar Alsayed, Abrar Sebiany, Joury Alzayat, and Mona Alqahtani; Resources, Mustafa Youldash; Software, Manar Alsayed, Abrar Sebiany, Noor Aljishi, Ghaida Alshammari, and Mona Alqahtani; Supervision, Mustafa Youldash, and Atta Rahman; Validation, Joury Alzayat; Writing an original draft, Abrar Sebiany, Joury Alzayat, Noor Aljishi, Ghaida Alshammari, and Mona Alqahtani; Writing review editing, Atta Rahman. All authors reviewed the results and approved the final version of the manuscript.

**Availability of Data and Materials:** The data utilized in the current study is publicly available at Kaggle. <https://www.kaggle.com/competitions/aptos2019-blindness-detection/data> (accessed on 19 December 2024).

**Ethics Approval:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest to report regarding the present study.

## References

- [1] "Diabetic Retinopathy | National Eye Institute," Accessed: Oct. 19, 2023. [Online]. Available: <https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/diabetic-retinopathy>

- [2] G. N. Gelcho and F. S. Gari, "Time to diabetic retinopathy and its risk factors among diabetes mellitus patients in Jimma University Medical Center, Jimma, Southwest Ethiopia," *Ethiop J. Health Sci.*, vol. 32, no. 5, Sep. 2022, Art. no. 937.
- [3] W. L. Alyoubi, M. F. Abulkhair, and W. M. Shalash, "Diabetic retinopathy fundus image classification and lesions localization system using deep learning," *Sensors*, vol. 21, no. 11, Jun. 2021, Art. no. 3704. doi: [10.3390/S21113704](https://doi.org/10.3390/S21113704).
- [4] J. Wolf *et al.*, "Liquid-biopsy proteomics combined with AI identifies cellular drivers of eye aging and disease *in vivo*," *Cell*, vol. 186, no. 22, pp. 4868–4884.e12, Oct. 2023.
- [5] Y. Zheng, M. He, and N. Congdon, "The worldwide epidemic of diabetic retinopathy," *Indian J. Ophthalmol.*, vol. 60, no. 5, Sep. 2012, Art. no. 428. doi: [10.4103/0301-4738.100542](https://doi.org/10.4103/0301-4738.100542).
- [6] W. Xu, Y. -L. Fu, and D. Zhu, "ResNet and its application to medical image processing: Research progress and challenges," *Comput. Methods Programs Biomed.*, vol. 240, 2023, Art. no. 107660. doi: [10.1016/j.cmpb.2023.107660](https://doi.org/10.1016/j.cmpb.2023.107660).
- [7] T. Shahzad, M. Saleem, M. S. Farooq, S. Abbas, M. A. Khan and K. Ouahada, "Developing a transparent diagnosis model for diabetic retinopathy using explainable AI," *IEEE Access*, vol. 12, pp. 149700–149709, 2024. doi: [10.1109/ACCESS.2024.3475550](https://doi.org/10.1109/ACCESS.2024.3475550).
- [8] B. F. Wee, S. Sivakumar, K. H. Lim, W. K. Wong, and F. H. Juwono, "Diabetes detection based on machine learning and deep learning approaches," *Multimed. Tools Appl.*, vol. 83, no. 1, pp. 24153–24185, 2024. doi: [10.1007/s11042-023-16407-5](https://doi.org/10.1007/s11042-023-16407-5).
- [9] L. Fregoso-Aparicio, J. Noguez, L. Montesinos, and J. A. García-García, "Machine learning and deep learning predictive models for type 2 diabetes: A systematic review," *Diabetol. Metab. Syndr.*, vol. 13, no. 148, 2021. doi: [10.1186/s13098-021-00767-9](https://doi.org/10.1186/s13098-021-00767-9).
- [10] L. Czupryniak, D. Dicker, R. Lehmann, and G. Schernthaner, "The management of type 2 diabetes before, during and after COVID-19 infection: What is the evidence?," *Cardiovasc. Diabetol.*, vol. 20, no. 198, 2021. doi: [10.1186/s12933-021-01389-1](https://doi.org/10.1186/s12933-021-01389-1).
- [11] S. Qummar *et al.*, "A deep learning ensemble approach for diabetic retinopathy detection," *IEEE Access*, vol. 7, pp. 150530–150539, 2019. doi: [10.1109/ACCESS.2019.2947484](https://doi.org/10.1109/ACCESS.2019.2947484).
- [12] R. Yasashvini, M. V. Raja Sarobin, R. Panjanathan, S. G. Jasmine, and L. J. Anbarasi, "Diabetic retinopathy classification using CNN and hybrid deep convolutional neural networks," *Symmetry*, vol. 14, no. 9, 2022, Art. no. 1932. doi: [10.3390/sym14091932](https://doi.org/10.3390/sym14091932).
- [13] G. Ali, A. Dastgir, M. W. Iqbal, M. Anwar, and M. Faheem, "A hybrid convolutional neural network model for automatic diabetic retinopathy classification from fundus images," *IEEE J. Transl. Eng. Health Med.*, vol. 11, pp. 341–350, 2023. doi: [10.1109/JTEHM.2023.3282104](https://doi.org/10.1109/JTEHM.2023.3282104).
- [14] G. Alwakid, W. Gouda, M. Humayun, and N. Zaman Jhanjhi, "Deep learning-enhanced diabetic retinopathy image classification," *Digit. Health*, vol. 9, 2023. doi: [10.1177/20552076231194942](https://doi.org/10.1177/20552076231194942).
- [15] H. Shakibania, S. Raoufi, B. Pourafkham, H. Khotanlou, and M. Mansoorizadeh, "Dual branch deep learning network for detection and stage grading of diabetic retinopathy," *Biomed. Signal Process. Control*, vol. 93, 2024, Art. no. 106168. doi: [10.1016/j.bspc.2024.106168](https://doi.org/10.1016/j.bspc.2024.106168).
- [16] A. A. Kumari, A. Bhagat, S. K. Henge, and S. K. Mandal, "Automated decision making ResNet feed-forward neural network based methodology for diabetic retinopathy detection," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 5, pp. 303–314, 2023. doi: [10.14569/IJACSA.2023.0140532](https://doi.org/10.14569/IJACSA.2023.0140532).
- [17] A. Rahman *et al.*, "Diabetic retinopathy detection: A hybrid intelligent approach," *Comput. Mat. Contin.*, vol. 80, no. 3, pp. 4561–4576, 2024. doi: [10.32604/cmc.2024.055106](https://doi.org/10.32604/cmc.2024.055106).
- [18] A. H. Khudaier and A. M. Radhi, "Binary classification of diabetic retinopathy using CNN architecture," *Iraqi J. Sci.*, vol. 65, no. 2, pp. 963–978, Feb. 2024. doi: [10.24996/ijsc.2024.65.2.31](https://doi.org/10.24996/ijsc.2024.65.2.31).
- [19] "APTOS 2019 Blindness Detection," Accessed: Nov. 28, 2023. [Online]. Available: [www.kaggle.com/competitions/aptos2019-blindness-detection/data](https://www.kaggle.com/competitions/aptos2019-blindness-detection/data)
- [20] N. Kundu, "Exploring ResNet50: An in-depth look at the model architecture and code implementation," *Medium*, vol. 23, Jan. 23, 2023. Accessed: Apr. 09, 2024. [Online]. Available: <https://medium.com/@nitishkundu1993/exploring-resnet50-an-in-depthlook-at-the-model-architecture-and-code-implementation-d8d8fa67e46f>

- [21] G. Boesch, “Deep residual networks (ResNet, RESNET50)-2024 guide, viso.ai,” Accessed: Apr. 09, 2024, 2024. [Online]. Available: <https://viso.ai/deep-learning/resnet-residual-neural-network/>
- [22] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *2017 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, 2017, pp. 2261–2269.
- [23] A. K. Gangwar and V. Ravi, “Diabetic retinopathy detection using transfer learning and deep learning,” in *Evolution in Computational Intelligence. Advances in Intelligent Systems and Computing*, V. Bhateja, S. L. Peng, S. C. Satapathy, Y. D. Zhang, Eds. Singapore: Springer, 2021, vol. 1176, pp. 679–689. doi: [10.1007/978-981-15-5788-0\\_64](https://doi.org/10.1007/978-981-15-5788-0_64).
- [24] M. N. Alnuaimi *et al.*, “Transfer learning empowered skin diseases detection in children,” *Comput. Model. Eng. Sci.*, vol. 141, no. 3, pp. 1–15, 2024. doi: [10.32604/cmcs.2024.055303](https://doi.org/10.32604/cmcs.2024.055303).
- [25] L. Balyen, “New approaches in the detection and management of diabetic retinopathy in the near future,” *Adv Ophthalmol. Vis. Syst.*, vol. 13, no. 1, pp. 8–9, Jan. 2023. doi: [10.15406/aovs.2023.13.00430](https://doi.org/10.15406/aovs.2023.13.00430).
- [26] L. Dai *et al.*, “A deep learning system for detecting diabetic retinopathy across the disease spectrum,” *Nat. Commun.*, vol. 12, no. 1, 2021, Art. no. 3242. doi: [10.1038/s41467-021-23458-5](https://doi.org/10.1038/s41467-021-23458-5).
- [27] J. Zhou and B. Chen, “Retinal cell damage in diabetic retinopathy,” *Cells*, vol. 12, no. 9, May 2023, Art. no. 1342. doi: [10.3390/cells12091342](https://doi.org/10.3390/cells12091342).
- [28] J. Grauslund, “Diabetic retinopathy screening in the emerging era of artificial intelligence,” *Diabetologia*, vol. 65, no. 9, pp. 1415–1423, Sep. 2022. doi: [10.1007/s00125-022-05727-0](https://doi.org/10.1007/s00125-022-05727-0).
- [29] L. Zhan, “Frontiers in understanding the pathological mechanism of diabetic retinopathy,” *Med Sci. Monitor.*, vol. 29, no. 1, Apr. 2023, Art. no. e939658. doi: [10.12659/MSM.939658](https://doi.org/10.12659/MSM.939658).
- [30] M. Kumar, P. Genter, and E. Ipp, “LBODP048 barriers to diabetic retinopathy screening in hospitalized patients with diabetes,” *J. Endocrine Soc.*, vol. 6, no. 1, p. A271, Nov. 2022.
- [31] W-H. Yang, Y. Shao, and Y. W. Xu, “Guidelines on clinical research evaluation of artificial intelligence in ophthalmology,” *Int. J. Ophthalmol.*, vol. 16, no. 9, pp. 1361–1372, Sep. 18, 2023. doi: [10.18240/ijo.2023.09.02](https://doi.org/10.18240/ijo.2023.09.02).
- [32] M. Gollapalli *et al.*, “Appendicitis diagnosis: Ensemble machine learning and explainable artificial intelligence-based comprehensive approach,” *Big Data Cogn. Comput.*, vol. 8, no. 9, 2024, Art. no. 108. doi: [10.3390/bdcc8090108](https://doi.org/10.3390/bdcc8090108).
- [33] S. M. Alotaibi, Atta-ur-Rahman, M. I. Basheer, and M. A. Khan, “Ensemble machine learning based identification of pediatric epilepsy,” *Comput. Mater. Contin.*, vol. 68, no. 1, pp. 149–165, 2021. doi: [10.32604/cmcs.2021.015976](https://doi.org/10.32604/cmcs.2021.015976).
- [34] J. Yang *et al.*, “Optimizing diabetic retinopathy detection with Inception-V4 and the dynamic version of the snow leopard optimization algorithm,” *Biomed. Signal Process. Control*, vol. 96, 2024, Art. no. 106501. doi: [10.1016/j.bspc.2024.106501](https://doi.org/10.1016/j.bspc.2024.106501).
- [35] B. Menaouer, Z. Dermene, N. El Houda Kebir, and N. Matta, “Diabetic retinopathy classification using hybrid deep learning approach,” *SN Comput. Sci.*, vol. 3, no. 1, 2022, Art. no. 357. doi: [10.1007/s42979-022-01240-8](https://doi.org/10.1007/s42979-022-01240-8).
- [36] M. I. B. Ahmed, “Early detection of diabetic retinopathy utilizing advanced fuzzy logic techniques,” *Math. Modell. Eng. Problem.*, vol. 10, no. 6, pp. 2086–2094, 2023. doi: [10.18280/mmep.100619](https://doi.org/10.18280/mmep.100619).
- [37] A. Bilal *et al.*, “DeepSVDNet: A deep learning-based approach for detecting and classifying vision threatening diabetic retinopathy in retinal fundus images,” *Comput. Syst. Sci. Eng.*, vol. 48, no. 2, pp. 511–528, Jan. 2024. doi: [10.32604/csse.2023.039672](https://doi.org/10.32604/csse.2023.039672).
- [38] M. Youldash *et al.*, “Early detection and classification of diabetic retinopathy: A deep learning approach,” *AI*, vol. 5, no. 4, pp. 2586–2617, 2024. doi: [10.3390/ai5040125](https://doi.org/10.3390/ai5040125).
- [39] Y. Zhou *et al.*, “A foundation model for generalizable disease detection from retinal images,” *Nature*, vol. 622, no. 1, pp. 156–163, 2023. doi: [10.1038/s41586-023-06555-x](https://doi.org/10.1038/s41586-023-06555-x).
- [40] S. Muchuchuti and V. Serestina, “Retinal disease detection using deep learning techniques: A comprehensive review,” *J. Imaging*, vol. 9, no. 4, Apr. 18, 2023, Art. no. 84. doi: [10.3390/jimaging9040084](https://doi.org/10.3390/jimaging9040084).