RESEARCH ARTICLE

Automated Early Diabetic Retinopathy Detection Using a Deep Hybrid Model

Asima Shazia[1], **Fida Hussain Dahri**[1,*], **Asfand Ali**[2], **Muhammad Adnan**[3], **Asif Ali Laghari**[4] and **Tehniyat Nawaz**[5]

[1] School of Computer Science and Engineering, Southeast University, Nanjing 211189, China
[2] Department of Computer, Control and Management Engineering (DIAG), University Sapienza Università di Roma, Italy
[3] Data Science and Business Informatics, University of Pisa, Pisa 56124, Italy
[4] Software College, Shenyang Normal University, Shenyang 110136, China
[5] Department of Computer and Software Engineering, NUST Islamabad, Pakistan

Abstract

Recently, the primary reason for blindness in adults has been diabetic retinopathy (DR) disease. Therefore, there is an increasing demand for a real-time efficient classification and detection system for diabetic retinopathy (DR) to overcome fast-growing disease (DR). We introduced a novel deep hybrid model for automated diabetic retinopathy (DR) disease recognition and classification. Our model leverages the power of CNN architectures: Inception V3 and VGG16 models by combining their strengths to cater to exact requirements. VGG16 model efficiently captures fine features and wide-ranging features such as textures and edges, crucial for classifying initial signs of DR. Similarly, Inception V3's architecture is proficient at detecting multiscale patterns, providing an extensive setting for shaping the occurrence of more complex DR severity stages. Our deep hybrid model allows the extraction of various appearance features in retinal images, which can better assist the classification and detection of DR. Our proposed model was evaluated on diverse datasets, including EyePACS1 and APTOS2019, demonstrating confident performance of 99.63% accuracy in classifying the DR severity levels on EyePACS1 dataset, while 98.70% accuracy on the APTOS2019 dataset, indicating that our proposed deep hybrid model well distinguished different stages and highly efficient in DR detection. This model helps clinicians and medical experts to classify and identify diabetic retinopathy DR stages and severity levels early. This automatic system helps to manage and treat the patient more effectively and introduces timely treatment.

Keywords: retinal images, diabetic retinopathy (DR), deep hybrid model, Inception V3, VGG16.

1 Introduction

Diabetic retinopathy (DR) is the foremost reason for blindness in the young, a significant complication

Submitted: 30 October 2024
Accepted: 17 November 2024
Published: 20 November 2024

Vol. 1, **No.** 1, 2024.
10.62762/TETAI.2024.305743

***Corresponding author:**
✉ Fida Hussain Dahri
223227084@seu.edu.cn

Citation
Shazia, A., Dahri, F. H., Ali, A., Adnan, M., Laghari, A. A., & Nawaz, T. (2024). Automated Early Diabetic Retinopathy Detection Using a Deep Hybrid Model. *ICCK Transactions on Emerging Topics in Artificial Intelligence*, 1(1), 71–83.

© 2024 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

of diabetes, and a most important public health problem [1]. DR impacts one-third of the 537 million people with diabetes worldwide, with projections estimating an increase to 643 million by 2030 and 783 million by 2045 [2]. The global prevalence of DR currently affects over 158.2 million people and is expected to exceed 191 million by 2030. This debilitating eye disease develops due to prolonged diabetes, causing damage in the retina to the blood vessels, the light-sensitive tissue in the rear of the eye [4]. Automated deep learning approaches, including transfer learning-based methods [3], have demonstrated promising capabilities in detecting such retinal damage from fundus images, motivating the development of more accurate automated diagnostic systems. DR progresses through several stages associated with pathological changes and corresponding visual impairments [5]. Non-proliferative Diabetic Retinopathy (NPDR): In this phase, the blood vessels in the retina lead to microaneurysms (small leaks), haemorrhages (bleeding), and cotton wool spots (damaged areas in the nerve fiber layer of the retina) [6]. The NPDR type is further divided into mild, moderate, and severe based on these abnormalities. Proliferative Diabetic Retinopathy (PDR): In this advanced stage, new blood vessels form on the retina's surface, which are fragile and prone to leaking [7], leading to vision loss. These vessels can block fluid flow in the vitreous, causing vitreous haemorrhage and further impairing vision [8].

Timely detection and intervention are critical to prevent severe vision loss, as DR can progress rapidly [9, 10]. However, conventional DR screening methods require significant time and resources [5, 11–14]. The physical inspection of retinal images by ophthalmologists, while standard, has limitations due to subjective interpretation and the need for specialized expertise [16, 17]. These limitations have driven researchers to develop CNN-based automated classification systems that can distinguish multiple DR severity stages with greater consistency and objectivity [15]. This makes large-scale implementation challenging. To address these limitations, an intelligent detection system is essential for accurate and efficient early-stage DR diagnosis, increasing accessibility to early intervention.

Our proposed study introduces a deep hybrid model for automated DR classification and detection. The model combines the strengths of VGG16 and Inception V3 Convolutional Neural Networks (CNNs). VGG16 is proficient at capturing detailed features, such as textures and edges, aiding in early DR identification, while Inception V3 detects multiscale patterns and contributes to identifying complex severity levels. We tested the model on EyePACS1 [18] and APTOS2019 [19] datasets, achieving superior results in DR severity classification. Unlike traditional methods, our approach leverages deep learning to extract detailed features from retinal images, significantly improving DR diagnostic accuracy and efficiency.

- Suggest a novel deep hybrid model that combines the convolutional neural network (CNN) models, VGG16 and Inception V3, leveraging their complementary strengths to surpass the performance of existing state-of-the-art works, demonstrating models' effectiveness in classifying and detecting DR disease severity levels.
- Improve the deep CNN-based early classification and detection system using VGG16 and inception V3 models. The model leads to improved extraction of complex features from retinal images and a more accurate and robust DR diagnosis system.
- Demonstrate the effective implementation of the deep hybrid model in real-time diagnosis, highlighting its effectiveness in detecting diabetic retinopathy (DR) and its potential to significantly enhance patient care and decrease the consequences of DR by offering clinicians a robust and efficient DR classification system.

This paper is organized into five sections. Section 1 provides an introduction to the topic. Section 2 offers a detailed review of prior research on the diagnosis of diabetic retinopathy. Section 3 outlines the materials and methods used in the study. Section 4 presents a comprehensive analysis of the results, followed by a discussion of the findings. Finally, Section 5 concludes the paper with a summary of key insights and implications.

2 Related Work

This review section highlights prior works on diabetic retinopathy classification and detection. Panwar et al. [20] presented an alternative machine-learning approach to analyzing retinal fundus images to diagnose diabetic retinopathy. Despite the traditionally diverse data and the distribution of disease stages, they trained their model on a large sample of retinal images. Models' performance

demonstrated that the overall evaluation of diabetic retinopathy diseases remained high compared to the traditional and prior approaches.

Jain et al. [1] proposed various CNN models, including InceptionV3 [21], VGG19, and VGG16 [32], and assessed the models of DR classification concerning labels for two classes and five classes. Results realized the model's performance would be proportional to increased depth in terms of pooling layers and convolutional in CNN. The top precision for two types of DR classification attained 80.40% using VGG19. In a study conducted by Voets et al. [13], authors used ensemble deep learning for binary classification (NRDR-RDR) and InceptionV3 for feature extraction, the model trained on a publicly available dataset, EyePACS. The proposed study provided an AUC of 85.3% on Messidor-2. Another survey study by Toledo-Cortés et al. [16] used InceptionV3 to classify RDR. The authors have recommended a deep-learning Gaussian Process (GP) model using the EyePACS dataset. A GP regression has an AUC of 87.83% on Messidor2. Masood et al. [22] shows an accuracy of 48.2% in DR detection using the transfer learning model on the EyePACS database Inception-V3. Harun et al. [23] discoursed a Multi-layered perceptron trained along Bayesian regularization, achieving an improved classification presentation compared to using Levenberg–Marquardt in training at 72.11% and testing accuracy at 67.47%. Qummar et al. [24] proposed a deep learning ensemble approach for DR detection on the Kaggle dataset, combining five CNN architectures (VGG16, VGG19, Inception-V3, ResNet50, and InceptionResNetV2) to improve classification robustness, demonstrating that ensemble strategies substantially outperform individual models in DR grading tasks. Zago et al. [25] achieved the Kappa score (K-Sc) of 0.86 using the public Kaggle dataset and the random forest classifier. The CNN model was used to assess the severity of diabetic retinopathy. Saranya and Prabakaran [26] introduced a method for the initial diagnosis of diabetic retinopathy using retinal fundus images, implementing a proposed automatic recognition system. After intensive training on a large corpus of retinal images, the model showed very high accuracy for diagnosis. Yang et al. [27] proposed a two-stage deep convolutional neural network approach for DR lesion detection and grading, demonstrating strong performance in distinguishing multiple severity levels of diabetic retinopathy. On the other hand, Jabbar et al. [28] proposed a hybrid solution to identify DR concentrated image processing and deep learning together for better examinations of diagnosing from a fundus image. For validation of model 400 retinal fundus images in the Messidor dataset used and assessed with various performance metrics, they attained the promised result with an average accuracy of 97%, 94% recall, Specificity of 98%, precision of 96%, F-score95%, and GMean95%. Jiang et al. [29] used three deep learning models (Inception V3, ResNet151, and InceptionResNet-V2). They attained accuracies of 87.91%, 87.20%, and 86.18%. When they combined models with the AdaBoost algorithm, they reached an even better accuracy of 88.21%. Referring to the study [30], the authors applied transfer learning by stacking CNNs network on topmost of ResNet and Inception-based models to perform multiclass classification of fundus images for blindness detection using the (APTOS) dataset. The pictures underwent pre-processed steps through resizing, obscuring, and BBO operations. Data augmentation is utilized to balance the class data distribution. The study achieved a test accuracy of 82.18% using the APTOS dataset. The authors in [17] proposed a multiclass classification method for eye-related diseases, using CNN-based network architecture for fundus image classification and leveraging the power of transfer learning that groups ocular diseases into their respective classes. They used the Peking University Ocular Disease Intelligent Recognition (PKU-ODIR) dataset, which includes class labels for eight types of ocular diseases.

3 Materials and Methods

This methods and materials section focuses on the techniques and materials used in our novel study. The fundamental purpose of this study is to propose a novel deep hybrid model for an automated diagnosis of diabetic retinopathy (DR) and enhance real-time early DR stages classification, surpass the previous classification methods with promised performance, and accomplish the desired outcomes in diabetic retinopathy research (DR). Our study leverages the strengths of CNN architectures, such as VGG16 and Inception V3, employing transfer learning to initialize both models with pre-trained ImageNet weights and fine-tune them for DR classification. The VGG16 model is used because of its ability to identify intricate features like texture and edges, which are crucial for the timely detection of diabetic retinopathy (DR). The Inception V3 model is chosen for its solid pre-training performance and ability to analyze and extract complex and multiple features. It just requires a more remarkable ability to include many variables.

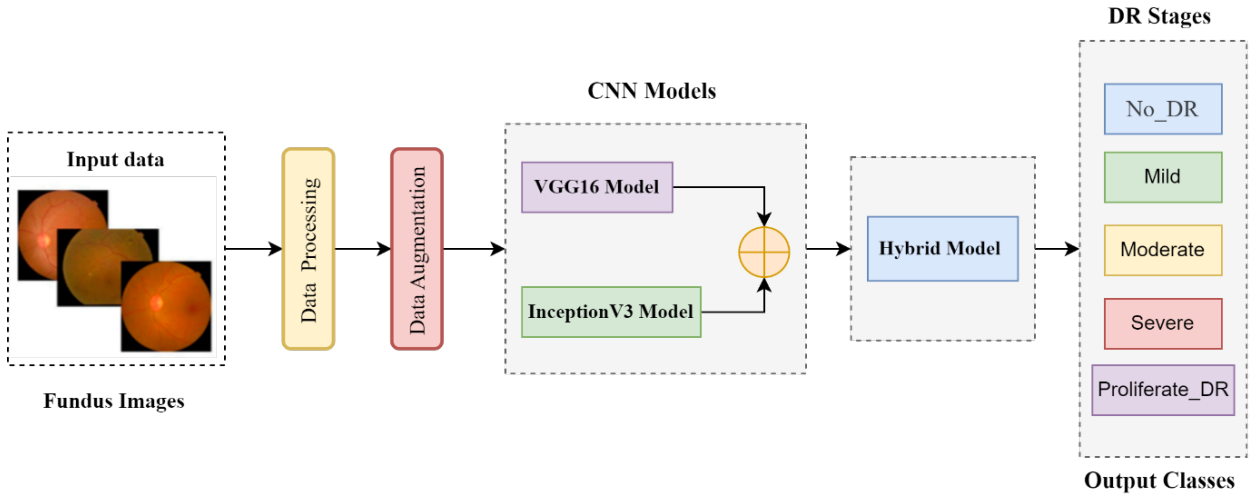

Figure 1. The proposed deep hybrid model for DR detection.

Table 1. Summary of the datasets utilized in experiments.

Datasets	Training		Testing		Validation		Total Images	Classes
	Data Division	Images	Data Division	Images	Data Division	Images		
EyePACS1	80%	28,101	10%	3,513	10%	3,513	35,126	5
APTOS2019	80%	4,467	10%	558	10%	559	5,584	5

Table 2. Distribution of classes in datasets.

Dataset	No DR	Mild	Moderate	Severe	Proliferative DR	Total Images
EyePACS1	18,456	8,712	5,270	1,902	786	35,126
APTOS2019	3,005	1,117	890	377	195	5,584

The performance assessment of our deep hybrid model is conducted by assessing the EyePACS1 and APTOS2019 datasets. These two datasets serve as a reliable basis for training and evaluating models under various situations and levels of severity in diabetic retinopathy. They are extensively used and acknowledged in the field of medicine. We use this extensively employed data to evaluate the real-world effectiveness of our novel hybrid model. We conducted independent evaluations of the accuracy of both VGG16 and Inception V3 on these datasets, enabling us to compare them individually and together. The process included acquiring curated datasets and using transformer learning models (VGG16 and Inception V3) to create a deep hybrid architecture. It also included determining parametric settings, training methods, assessment criteria, and model evaluation. We are comparing the findings of our hybrid model with prior cutting-edge approaches to understand how well it can identify and classify different phases of diabetic retinopathy and its overall efficiency. Figure 1 portrays the research flow diagram of the proposed deep hybrid model to diagnose diabetic retinopathy.

3.1 Data Acquisition

We use multiple data from public databases, such as Kaggle datasets: EyePACS1 and APTOS2019. Both image datasets have five categories: "No DR," "Mild," "Moderate," "Severe," and "Proliferative" DR images. APTOS Asia Pacific Teleophthalmology Association created the APTOS2019 database of 5,584 high-resolution images as part of the 2019 Visual Vision Challenge. Furthermore, the California Health Foundation created the EyePACS1 dataset containing 35,126 high-resolution photos for the Diabetic Retinopathy Detection Challenge. These are well-known and widely used datasets designed to study performance models in assessing diabetic retinopathy. Tables 1 and 2 provide detailed statistics of both datasets.

3.2 Data Processing

The APTOS2019 and EyePACS datasets lay under varying lighting conditions because both datasets

comprise the retinal fundus pictures captured using different cameras with different specifications. In datasets, there are various inconsistencies and unbalanced data. Therefore, to overcome these inconsistencies, all images underwent numerous pre-processing steps to normalize and standardize datasets. Figures 2 and 3 show both datasets' original fundus image samples before applying pre-processing techniques.

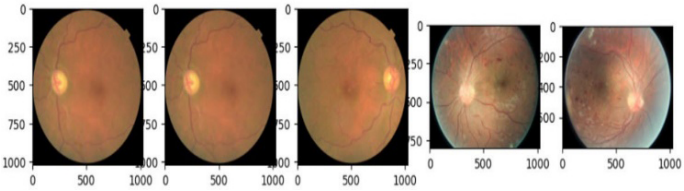

Figure 2. The eye retinal fundus images EyePACS Dataset before pre-processing.

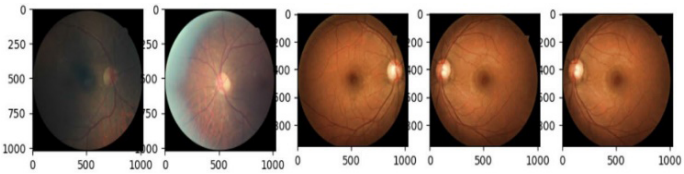

Figure 3. The eye retinal fundus images APTOS2019 Dataset before pre-processing.

Resizing is a crucial step in the pre-processing of diabetic retinopathy fundus images. Bicubic interpolation is often used for this purpose, as it calculates the weighted average of surrounding pixels, leading to smoother and more visually consistent results. This method is favored because it helps minimize artifacts and distortions during resizing, ensuring the images remain clear and free of visual imperfections. When resizing the image, for each new pixel in the output image, the bicubic interpolation formula is applied using the corresponding neighboring pixels in the original image, calculated as:

$$x = \frac{x'}{\text{scale factor}} \tag{1}$$

$$y = \frac{y'}{\text{scale factor}} \tag{2}$$

The scale factor is the ratio of the output size to the input size in either dimension (horizontal or vertical). The resulting $I'(x', y')$ is the interpolated intensity value for the resized image at the location (x', y'). This process ensures that the resized image maintains smooth transitions and reduces the appearance of artifacts and distortions, producing a high-quality, visually consistent output. After resizing, the photos from both datasets are displayed in Figures 4 and 5.

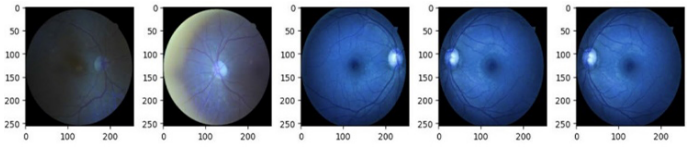

Figure 4. The eye retinal fundus images EyePACS Dataset following image resizing.

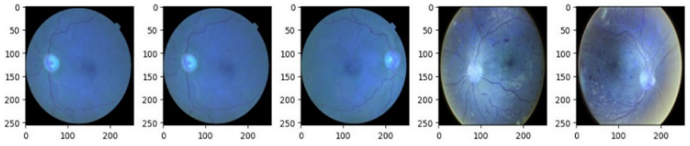

Figure 5. The eye retinal fundus images APTOS2019 Dataset following image resizing.

3.3 Data Augmentation

The performance of deep learning models is greatly affected by the size and variety of the dataset. A large and diverse training dataset is essential for avoiding overfitting and ensuring strong generalization. Various techniques, such as flipping, cropping, rotating, and zooming, are applied to enhance the diversity of the images, as illustrated in Figures 6 and 7. These augmentation Figures 6 and 7 highlight the effects of the augmentation techniques on the training dataset, with operations like cropping, rotating, flipping, and shearing being key contributors. This approach significantly enhances the training data representativeness and comprehensiveness, providing a strong foundation for the model's effectiveness.

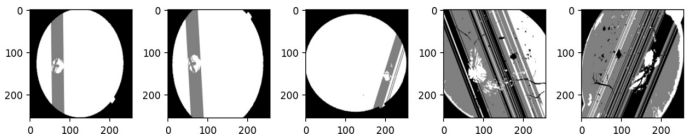

Figure 6. EyePACS Dataset fundus images after applying various data augmentation methods.

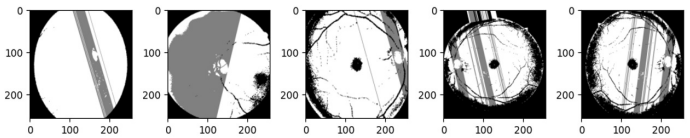

Figure 7. APTOS2019 Dataset fundus images after applying various data augmentation methods.

Table 3. VGG16 parameters.

Layer	Type	Parameters	Example Values
Initial Convolutional Layers	Conv2d	Filters: 64, Kernel Size: (3, 3), Padding: (1, 1)	Filters: 64, Kernel Size: (3, 3), Padding: (1, 1)
Convolutional Blocks (Blocks 1–5)	Conv2d	Number of Layers: 13, Filters: 64, 128, 256, 512	Number of Layers: 13, Filters: 64, 128, 256, 512
Max-Pooling Layers	MaxPool2d	Kernel Size: (2, 2), Stride: (2, 2)	Kernel Size: (2, 2), Stride: (2, 2)
Fully Connected Layers	Linear	Output Size: 4096, 4096, Number of Classes	Output Size: 4096, 4096, Number of Classes
Dropout	Dropout	Dropout Rate: 0.5	Dropout Rate: 0.5

3.4 VGG16 Architecture

We use a CNN based on the VGG16 architecture [32] to detect diabetic retinopathy (DR). The VGG16 model is constructed as a stack of convolutional layers (Conv2D) followed by the ReLU activation function, which is essential for capturing complex patterns in retinal fundus images. This architecture contains 13 convolutional layers with a small 3×3 receptive field, enabling the network to learn fine-grained features such as edges, boundaries, and textures that are critical for early-stage DR detection. Batch normalization is applied after the activation function to stabilize and accelerate the training process, while max-pooling layers are inserted after every two or three convolutional layers to reduce the spatial dimensions of feature maps, thereby downsampling the image, improving parameter efficiency, and preserving essential features, as shown in Table 3. The VGG16 model ultimately includes three fully connected layers, and the final output is classified into five DR severity categories: "No DR," "Mild," "Moderate," "Severe," and "Proliferative DR," which enhances the model's generalization capability. Although the VGG16 architecture is relatively straightforward, its deep structure and well-integrated design enable strong performance in image classification tasks.

3.5 Inception V3 Architecture

The Inception V3 architecture [21] is a more complex and advanced model designed to capture multiscale features from retinal fundus images, which is crucial for accurately detecting numerous levels of DR. The model is schematized by the Inception module, which employs multiple convolutional filters of different sizes (1x1, 3x3, 5x5) simultaneously, allowing the network to recognize features at various levels of granularity. The ReLU function follows each Inception module to highlight non-linearity, permitting the model to acquire complex patterns in the data. Batch normalization is applied to the output of the convolutions to ensure training stability and improve convergence. Inception V3 also adds an auxiliary to the intermediate process to solve the gradient vanishing problem and thus improve early learning. Table 4 displays the flexibility of Inception V3 model parameters. The architecture includes an output max pooling layer to downsample feature maps and decrease computational complexity while preserving important information for classification. The final output is produced by a fully connected process that classifies the input image into one of five DR groups. Inception V3 efficiently processes complex scales of retinal features, making it predominantly useful for DR classification.

Table 4. Inception network parameters.

Layer	Type	Parameters	Example Values
Initial Convolutional Layer	Conv2d	Filters: 32, Kernel Size: (3, 3), Stride: (2, 2), Padding: (0, 0)	Filters: 32, Kernel Size: (3, 3), Stride: (2, 2), Padding: (0, 0)
Inception Modules (Modules 1–11)	Mixed (1x1, 3x3, 5x5)	Filters: 64, 128, 256, 512	Filters: 64, 128, 256, 512
Auxiliary Classifiers	Conv2d + FC	Filters: 128, FC Output Size: 1024, Dropout: 0.7	Filters: 128, FC Output Size: 1024, Dropout: 0.7
GA Pooling	Adaptive AvgPool2d	Size of Output: (1, 1)	Size of Output: (1, 1)
FC Layer	Linear	Size of Output: Number of Classes	Size of Output: Number of Classes

3.6 Deep Hybrid Model (VGG16 + Inception V3)

Table 5. Hybrid architecture parameters.

Layer	Type	Parameters	Example Values
VGG16 Convolutional Blocks	Conv2d	Number of Layers: 13, Filters: 64, 128, 256, 512	Number of Layers: 13, Filters: 64, 128, 256, 512
Inception V3 Modules	Mixed (1x1, 3x3, 5x5)	Filters: 64, 128, 256, 512	Filters: 64, 128, 256, 512
Concatenation Layer	Concat	Input: VGG16 + Inception V3 Feature Maps	Input: VGG16 + Inception V3 Feature Maps
GA Pooling	Adaptive AvgPool2d	Output Size: (1, 1)	Output Size: (1, 1)
FC Layer	Linear	Output Size: Number of Classes	Output Size: Number of Classes
Dropout	Dropout	Dropout Rate: 0.5	Dropout Rate: 0.5

Our deep hybrid model integrates the power of

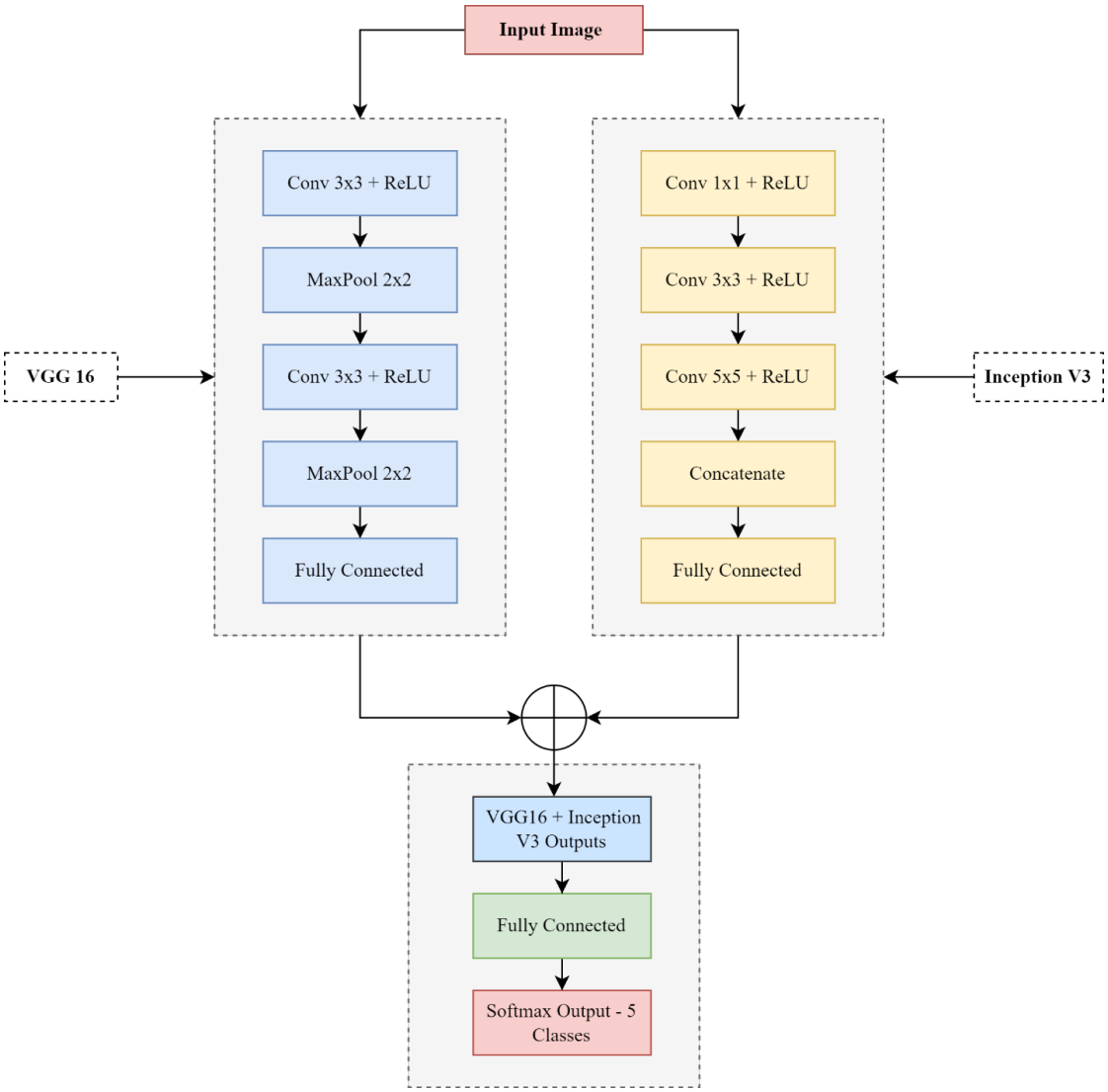

Figure 8. Schematic diagram of Deep hybrid model architecture.

CNN architectures to advance model performance and increase accuracy for diabetic retinopathy (DR) detection: the VGG16 and Inception V3 models. This hybrid approach first inputs pre-processed retinal fundus images independently of the VGG16 and Inception V3 networks. VGG16 is excellent at capturing fine details, such as texture and edges, which are essential for identifying early signs of DR. In contrast, the architecture of Inception V3 excels at analyzing multiple patterns and provides a broader context for analyzing more critical processes. After feature extraction, the outputs of the last convolutional layers of both architectures are combined to form a single feature layer by utilizing the advantages of the two CNN networks, classify the combined features into one of five DR weights, further improving the model's ability to expand widely across different datasets. Combining the fine-grained feature extraction capabilities of VGG16 with the multi-detection capabilities of Inception V3, the hybrid model outperforms all other models given more accurate and robust means to classify DR. Table 5 shows the parametric setting of our deep hybrid model. Figure 8 depicts the detailed architecture of our proposed model for DR detection.

3.7 Performance Metrics

We evaluate and analyze the performance of our proposed deep learning approach using different performance measuring metrics, including confusion matrix (CM), to evaluate the overall proposed model efficiency. The confusion matrix, an error matrix, is mainly helpful in evaluating classification models; its application in DR lesion stratification studies has demonstrated its effectiveness in assessing model performance across multiple severity classes [20]. CM

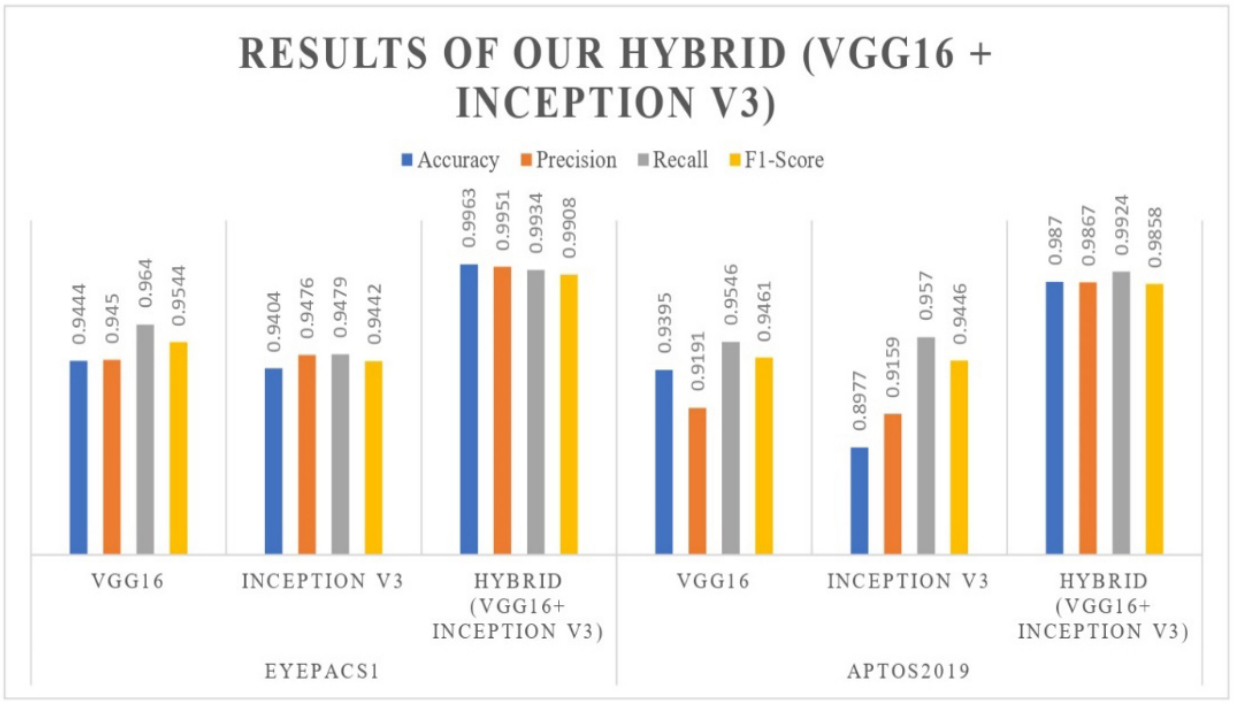

Figure 9. Result of our deep hybrid model.

can be defined as in equations.

$$\begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix} \tag{3}$$

The accuracy metric is the ratio of the sum of true positives (TP) and true negatives (TN) to the total number of TP, TN, FP, and FN, known as ACC.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

The precision metric is defined as the ratio of true positives (TP) to the total number of predicted positive labels, including both true positives (TP) and false positives (FP), which mathematically is defined as.

$$PRC = \frac{TP}{TP + FP} \tag{5}$$

However, recall measures how well a classification model detects positive examples in neural networks. Recall is the dataset's true positives (TP) ratio to positive instances, which is mathematically representative.

$$RC = \frac{TP}{TP + FN} \tag{6}$$

Finally, the F1-score is the harmonic mean of precision and recall, balancing the model's ability to minimize both false positives and false negatives. The equation shows the F1 score.

$$F1 = 2 \times \frac{PRC \times RC}{PRC + RC} \tag{7}$$

4 Results and Discussion

Our study provides an in-depth review of the performance and evaluation of our deep hybrid model in diagnosing diabetic retinopathy (DR). This novel model has been rigorously tested on various databases, including EyePACS1 and APTOS2019, to evaluate its robustness and performance. To assess the reliability and efficiency of our hybrid model, we compare its performance with the previous studies and methods, as shown in Table 7. A comprehensive analysis of results provides insight into the strengths and potential limitations of the deep hybrid model and provides detailed information about its potential. Our study uses various performance metrics to evaluate the model classification. Table 6 and Figure 9 show the results performance of our deep hybrid model on EyePACS1 and APTOS2019 datasets compared to models such as VGG16 and Inception V3.

The VGG16 model achieved 0.9444 accuracy, 0.9450 precision, 0.9640 recall, and 0.9544 F1-score on the EyePACS1 dataset. Similarly, the Inception V3 model achieved 0.9404 accuracy, 0.9476 precision, 0.9479 recall, and 0.9442 F1-score. The hybrid model

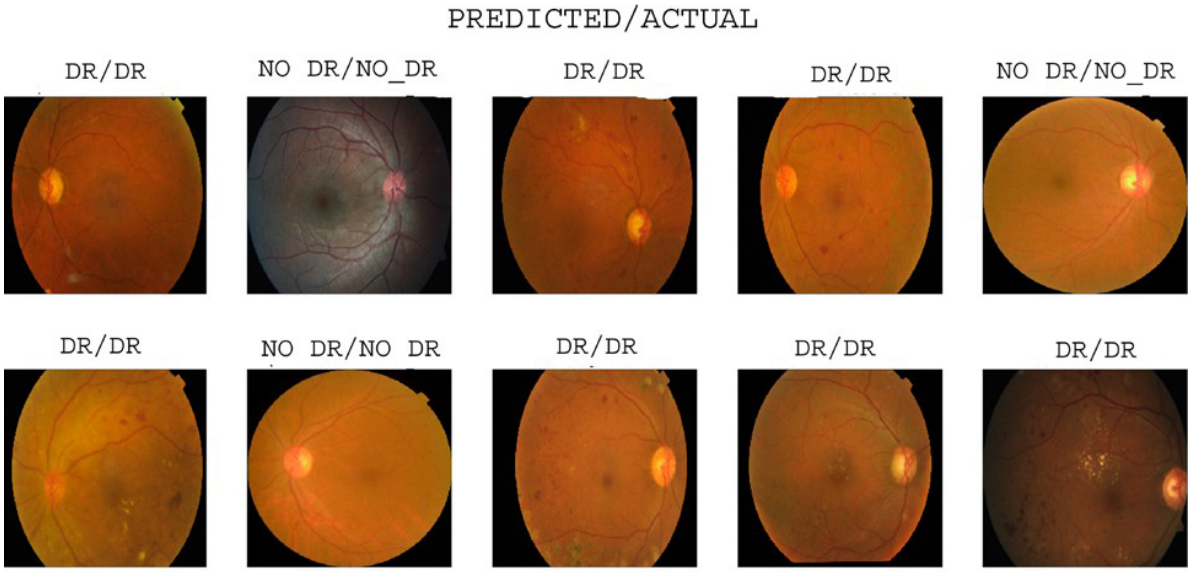

Figure 10. The predicted results of the proposed model on the Eyepacs1.

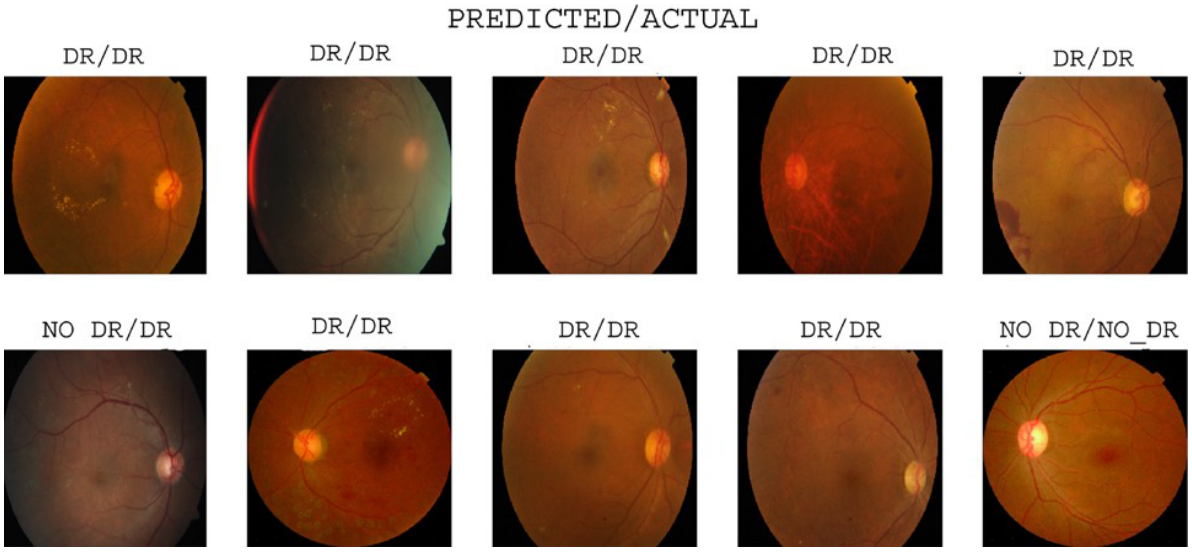

Figure 11. The predicted results of the proposed model on the APTOS2019.

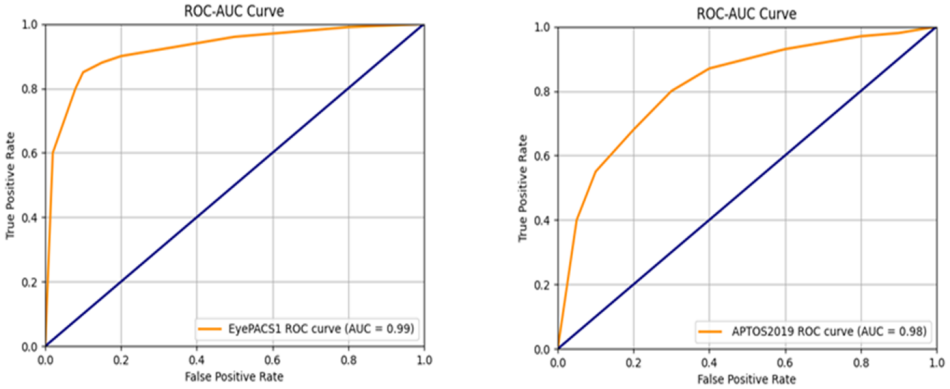

Figure 12. Comparative analysis of ROC-AUC for the hybrid model across all datasets.

combines VGG16 and Inception V3 and outperforms both by achieving 0.9963 accuracy, 0.9951 precision, 0.9934 recall, and 0.9908 F1-score. Correlation confusion matrices [426, 25, 16, 313], [384, 24, 21, 351], and [535, 2, 4, 245] provide detailed information about the performance of the model in detecting and classifying diabetic retinopathy. The model prediction result is shown in Figure 10.

Table 6. Results and discussion of hybrid VGG16 + Inception V3.

Dataset	Model	Accuracy	Precision	Recall	F1-Score
EyePACS1	VGG16	0.9444	0.9450	0.9640	0.9544
	Inception V3	0.9404	0.9476	0.9479	0.9442
	Hybrid	0.9963	0.9951	0.9934	0.9908
APTOS2019	VGG16	0.9395	0.9191	0.9546	0.9461
	Inception V3	0.8977	0.9159	0.9570	0.9446
	Hybrid	0.9870	0.9867	0.9924	0.9858

Note:Hybrid = VGG16 + Inception V3

Table 7. Comparative investigation with prior studies.

Reference	Model	Accuracy	Precision	Recall	F1-Score
[31]	U-Net	0.87	0.89	0.96	0.88
[21]	Inception V3	0.901	0.982	0.89	0.95
[32]	VGGNet	0.803	0.855	0.82	0.98
[33]	AlexNet	0.9680	0.87	0.87	0.86
[8]	Hybrid Model (Hybrid-a, Hybrid-f, Hybrid-c)	0.9137	N/A	0.9390	0.8634
[5]	KNN	0.87	0.66	0.93	N/A
[34]	WP-CNN	0.88	0.719	0.815	N/A
Proposed Deep Hybrid Model	Hybrid Model (VGG16 + InceptionV3)	0.9963 0.9870	0.9951 0.9867	0.9934 0.9924	0.9908 0.9858

The VGG16 model achieved 0.9395 accuracy, 0.9191 precision, 0.9546 recall, and 0.9461 F1-score on the APTOS2019 dataset. Similarly, the Inception V3 model achieved 0.8977 accuracy, 0.9159 precision, 0.9570 recall, and 0.9446 F1-score. The hybrid model combines VGG16 and Inception V3 and outperforms both by achieving 0.9870 accuracy, 0.9867 precision, 0.9924 recall, and 0.9858 F1-score. Correlation confusion matrices [411, 32, 24, 327], [577, 54, 26, 137], and [584, 7, 10, 193] provide detailed information about the performance of the model in detecting and classifying diabetic retinopathy. The model prediction result is shown in Figure 11.

Moreover, we employ many performance-measuring metrics to assess and compare our combined model with prior state-of-the-art research studies conducted on diverse datasets. In Figure 12, the visual analysis displays a comparable ROC-AUC curve. Moreover, Table 6 comprehensively summarizes the model's performance on several datasets, including various measuring metrics. Furthermore, Table 7 compares our hybrid model performance findings with prior research studies to detect and diagnose diabetic retinopathy (DR). Our hybrid model, consisting of VGG16 and Inception V3, provided robust assessment performance on the APTOS2019 and EyePACS1 curated datasets, demonstrating promising results compared to state-of-the-art studies. Our deep hybrid model achieves a notable accuracy of 0.9963 on EyePACS1, demonstrating its effectiveness in categorizing diabetic retinopathy (DR).

5 Conclusion

Our research study significantly contributes to making an automatic and early detection diabetic retinopathy system by proposing a deep hybrid model that integrates the strengths of VGG16 and Inception V3 architectures. This combination enhances the accuracy and robustness of the detection method, particularly in distinguishing between the five stages of diabetic retinopathy: No DR,' 'mild,' 'moderate,' 'severe,' and 'proliferative.' By leveraging the detailed feature extraction capabilities of VGG16 and the multiscale feature recognition power of Inception V3, our hybrid model addresses the challenges posed by varying retinal image characteristics. The model has been tested on two widely recognized datasets, APTOS2019

and EyePACS1, demonstrating its effectiveness in real-world applications. The detailed numerical results, which can be seen in Table 6, highlight the model's capabilities, with accuracy rates of 0.9963 on the EyePACS1 dataset and 0.9870 on the APTOS2019 dataset, along with high precision, Recall, and F1 scores. Compared to recent state-of-the-art models (as shown in Table 7), our hybrid model exhibits superior performance, underscoring its potential as a reliable tool for early and accurate diabetic retinopathy detection. However, it is essential to acknowledge the constraints of our model, such as how it depends on the quality and diversity of the training datasets, which might affect its ability to be applied consistently to other populations or imaging scenarios.

Limitations: Despite the strong performance of our deep hybrid model, there are several limitations. The model's effectiveness depends on the quality of input images, as poor-quality or low-resolution images may reduce diagnostic accuracy. The datasets used exhibit class imbalances, which could introduce bias, particularly for underrepresented classes. While the model performs well on EyePACS1 and APTOS2019, its generalizability to other datasets with different imaging protocols or demographics requires further validation. Additionally, the model's high computational complexity may limit scalability and deployment in resource-constrained environments.

Future Work: To augment the practicality of our model, we suggest conducting a future study that explicitly targets tasks related to these datasets and refines the hybrid model to achieve better performance in real-world scenarios. Future studies could evaluate the model's performance under different conditions and datasets and explore the integration of new advanced features or architectures. Additionally, while data augmentation and preprocessing techniques have shown improvements in model accuracy and generalizability, a direct comparison between the model's performance on augmented and non-augmented data was not performed in this study. As part of future work, we plan to conduct a comprehensive evaluation to assess how data augmentation reduces overfitting and enhances model robustness, providing further insights into its impact on the model's performance.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Jain, A., Jalui, A., Jasani, J., Lahoti, Y., & Karani, R. (2019, April). Deep learning for detection and severity classification of diabetic retinopathy. In *2019 1st International Conference on Innovations in Information and Communication Technology* (*ICIICT*) (pp. 1-6). Ieee. [CrossRef]

[2] International Diabetes Federation. (2021). *IDF Diabetes Atlas* (10th ed.). International Diabetes Federation. Retrieved from https://idf.org/news-and-resources/resources/idf-diabetes-atlas-2021/

[3] Sanjana, S., Shadin, N. S., & Farzana, M. (2021, November). Automated diabetic retinopathy detection using transfer learning models. In *2021 5th International Conference on Electrical Engineering and Information Communication Technology* (*ICEEICT*) (pp. 1-6). IEEE. [CrossRef]

[4] Nagpal, D., Panda, S. N., Malarvel, M., Pattanaik, P. A., & Khan, M. Z. (2022). A review of diabetic retinopathy: Datasets, approaches, evaluation metrics and future trends. *Journal of King Saud University-Computer and Information Sciences, 34*(9), 7138-7152. [CrossRef]

[5] Nair, M., & Mishra, D. S. (2019). Categorization of diabetic retinopathy severity levels of transformed images using clustering approach. *Int J Comput Sci Eng, 7*(1), 642-648. https://www.researchgate.net/publication/331928407

[6] Neelapala, A. K., Satapathi, G. S., & Mosa, S. A. (2023, February). Severity Analysis Automation for Detection of Non-Proliferative Diabetic Retinopathy. In *International Symposium on Sustainable Energy and Technological Advancements* (pp. 301-312). Singapore: Springer Nature Singapore. [CrossRef]

[7] Kropp, M., Golubnitschaja, O., Mazurakova, A., Koklesova, L., Sargheini, N., Vo, T. T. K. S., ... & Thumann, G. (2023). Diabetic retinopathy as the leading cause of blindness and early predictor of cascading complications—risks and mitigation. *Epma Journal, 14*(1), 21-42. [CrossRef]

[8] Butt, M. M., Iskandar, D. A., Abdelhamid, S. E., Latif, G., & Alghazo, R. (2022). Diabetic retinopathy detection from fundus images of the eye using

hybrid deep learning features. *Diagnostics, 12*(7), 1607. [CrossRef]

[9] Patibandla, R. L., Rao, B. T., & Murty, M. R. (2024). Revolutionizing Diabetic Retinopathy Diagnostics and Therapy through Artificial Intelligence: A Smart Vision Initiative. In *Transformative Approaches to Patient Literacy and Healthcare Innovation* (pp. 136-155). IGI Global. [CrossRef]

[10] Qureshi, I., Ma, J., & Abbas, Q. (2019). Recent development on detection methods for the diagnosis of diabetic retinopathy. *Symmetry, 11*(6), 749. [CrossRef]

[11] Kassani, S. H., Kassani, P. H., Khazaeinezhad, R., Wesolowski, M. J., Schneider, K. A., & Deters, R. (2019, December). Diabetic retinopathy classification using a modified xception architecture. In *2019 IEEE international symposium on signal processing and information technology (ISSPIT)* (pp. 1-6). IEEE. [CrossRef]

[12] Hill, L., & Makaroff, L. E. (2016). Early detection and timely treatment can prevent or delay diabetic retinopathy. *diabetes research and clinical practice, 120*, 241-243. [CrossRef]

[13] Voets, M., Møllersen, K., & Bongo, L. A. (2019). Reproduction study using public data of: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *PloS one, 14*(6), e0217541. [CrossRef]

[14] Carrera, E. V., González, A., & Carrera, R. (2017, August). Automated detection of diabetic retinopathy using SVM. In *2017 IEEE XXIV international conference on electronics, electrical engineering and computing (INTERCON)* (pp. 1-4). IEEE. [CrossRef]

[15] Saranya, P., Umamaheswari, K. M., Sivaram, M., Jain, C., & Bagchi, D. (2021, January). Classification of different stages of diabetic retinopathy using convolutional neural networks. In *2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM)* (pp. 59-64). IEEE. [CrossRef]

[16] Toledo-Cortés, S., De La Pava, M., Perdómo, O., & González, F. A. (2020). Hybrid deep learning Gaussian process for diabetic retinopathy diagnosis and uncertainty quantification. In *Ophthalmic Medical Image Analysis: 7th International Workshop, OMIA 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings 7* (pp. 206-215). Springer International Publishing. [CrossRef]

[17] Gour, N., & Khanna, P. (2021). Multi-class multi-label ophthalmological disease detection using transfer learning based convolutional neural network. *Biomedical signal processing and control, 66*, 102329. [CrossRef]

[18] Dugas, E., Jared, Jorge, & Cukierski, W. (2015). Diabetic Retinopathy Detection. Retrieved from https://kaggle.com/competitions/diabetic-retinopathy-detection

[19] Karthik, Maggie, & Dane, S. (2019). APTOS 2019 Blindness Detection. Retrieved from https://kaggle.com/competitions/aptos2019-blindness-detection

[20] Panwar, A., Semwal, G., Goel, S., & Gupta, S. (2022, April). Stratification of the lesions in color fundus images of diabetic retinopathy patients using deep learning models and machine learning classifiers. In *Edge Analytics: Select Proceedings of 26th International Conference—ADCOM 2020* (pp. 653-666). Singapore: Springer Singapore. [CrossRef]

[21] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).

[22] Masood, S., Luthra, T., Sundriyal, H., & Ahmed, M. (2017, May). Identification of diabetic retinopathy in eye images using transfer learning. In *2017 International conference on computing, communication and automation (ICCCA)* (pp. 1183-1187). IEEE. [CrossRef]

[23] Harun, N. H., Yusof, Y., Hassan, F., & Embong, Z. (2019, April). Classification of fundus images for diabetic retinopathy using artificial neural network. In *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)* (pp. 498-501). IEEE. [CrossRef]

[24] Qummar, S., Khan, F. G., Shah, S., Khan, A., Shamshirband, S., Rehman, Z. U., ... & Jadoon, W. (2019). A deep learning ensemble approach for diabetic retinopathy detection. *Ieee Access, 7*, 150530-150539. [CrossRef]

[25] Zago, G. T., Andreão, R. V., Dorizzi, B., & Salles, E. O. T. (2020). Diabetic retinopathy detection using red lesion localization and convolutional neural networks. *Computers in biology and medicine, 116*, 103537. [CrossRef]

[26] Saranya, P., & Prabakaran, S. (2020). Automatic detection of non-proliferative diabetic retinopathy in retinal fundus images using convolution neural network. *Journal of Ambient Intelligence and Humanized Computing*, 1-10. [CrossRef]

[27] Yang, Y., Li, T., Li, W., Wu, H., Fan, W., & Zhang, W. (2017). Lesion detection and grading of diabetic retinopathy via two-stages deep convolutional neural networks. In *Medical Image Computing and Computer Assisted Intervention-MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20* (pp. 533-540). Springer International Publishing. [CrossRef]

[28] Jabbar, A., Liaqat, H. B., Akram, A., Sana, M. U., Azpíroz, I. D., Diez, I. D. L. T., & Ashraf, I. (2024). A lesion-based diabetic retinopathy detection through hybrid deep learning model. *IEEE Access*. [CrossRef]

[29] Jiang, H., Yang, K., Gao, M., Zhang, D., Ma, H., & Qian, W. (2019, July). An interpretable

ensemble deep learning model for diabetic retinopathy disease classification. In *2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 2045-2048). IEEE. [CrossRef]

[30] Gangwar, A. K., & Ravi, V. (2021). Diabetic retinopathy detection using transfer learning and deep learning. In *Evolution in Computational Intelligence: Frontiers in Intelligent Computing: Theory and Applications (FICTA 2020), Volume 1* (pp. 679-689). Springer Singapore. [CrossRef]

[31] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18* (pp. 234-241). Springer international publishing. [CrossRef]

[32] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. [CrossRef]

[33] Birajdar, U., Gadhave, S., Chikodikar, S., Dadhich, S., & Chiwhane, S. (2020, January). Detection and classification of diabetic retinopathy using AlexNet architecture of convolutional neural networks. In *Proceeding of International Conference on Computational Science and Applications: ICCSA 2019* (pp. 245-253). Singapore: Springer Singapore. [CrossRef]

[34] Liu, Y. P., Li, Z., Xu, C., Li, J., & Liang, R. (2019). Referable diabetic retinopathy identification from eye fundus images with weighted path for convolutional neural network. *Artificial intelligence in medicine, 99*, 101694. [CrossRef]



Asima Shazia received her Bachelor's degree from The Islamia University of Bahawalpur, Pakistan, in 2021. She is pursuing her master's in Computer science and Technology at Southeast University, Nanjing, China. Her research areas are machine learning, deep learning, and image processing. (Email: simashazia@seu.edu.cn)



Fida Hussain Dahri received his Bachelor's degree in Information Technology from Quaid-e-Awam University of Engineering, Sciences & Technology Nawabshah, Pakistan, in 2020, securing 3rd position in his faculty. He completed his master's degree in Information Technology from the same university in 2023, graduating in the top 2% of his class. He is pursuing a second Master's in Computer Science and Technology at Southeast University, Nanjing, China, where he is also a Research Assistant at the Data and Intelligence Laboratory (D&Intel Lab). His research focuses on computer vision, machine learning, deep learning, and image processing. He has worked in artificial intelligence and computer vision and continues contributing to advancements in these fields. (Email: 223227084@seu.edu.cn)



Asfand Ali earned his Bachelor's in Telecommunication Engineering from Mehran University of Engineering and Technology, Jamshoro, Pakistan, in 2021. He also completed a Diploma of Associate Engineering (DAE) in Electrical Engineering. During his undergraduate studies, he gained hands-on experience through internships with Pakistan Telecommunication Company Limited (PTCL), Pakistan Broadcasting Corporation (PBC), and the National Telecommunication Corporation (NTC). He was. He is pursuing a Master's in Computer Engineering at Sapienza University di Roma, Italy. His areas of interest include Computer Vision, Natural Language Processing, Artificial Intelligence, and Large Language Models. (Email: ali.2068882@studenti.uniroma1.it)



Muhammad Adnan earned a Bachelor's in Computer Science from Quaid-e-Awam University of Engineering, Science, and Technology in Nawabshah, Pakistan, in 2020. During his undergraduate studies, he gained practical experience through hands-on projects and an internship at a leading software house in Karachi. He is pursuing a Master's in Data Science and Business Informatics at the University of Pisa, Italy. His research interests include Artificial Intelligence, Machine Learning, Deep Learning, Data Modeling, and Data Visualization. (Email: Adnanjatt1998@gmail.com)



Asif Ali Laghari received a B.S. degree in Information Technology from the Quaid-e-Awam University of Engineering Science and Technology Nawabshah, Pakistan, in 2007 and a Master's degree in Information Technology from the QUEST Nawabshah Pakistan in 2014. From 2007 to 2008, he was a Lecturer in the Computer and Information Science Department at the Digital Institute of Information Technology, Pakistan. In 2015, he joined the School of Computer Science & Technology, Harbin Institute of Technology, where he was a Ph.D. student. He is an assistant professor at Sindh Madressatul Islam University, Karachi, Pakistan, and is affiliated with Shenyang Normal University. He has published over 130 technical articles in scientific journals and conference proceedings. His current research interests include Machine Learning, Computer networks, cloud computing, IoT, Fog computing, and multimedia QoE management. (Email: Asiflaghari@synu.edu.cn)



Tehniyat Nawaz received her Bachelor's degree from The Islamia University of Bahawalpur, Pakistan in 2021. She is pursuing her master's degree in computer engineering at the National University of Science and Technology Pakistan. Her research area is Machine learning and Deep learning. (Email: tehniyat809@gmail.com)