Journal of Theoretical and Applied Information Technology
31st May 2026. Vol.104. No.10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN ENHANCED VGG16 BASED DEEP LEARNING FRAMEWORK FOR DIABETIC RETINOPATHY SEVERITY GRADING

ASHOK KUMAR KAVURU[1], RAJESH KUMAR PATJOSHI[2], RAKHEE PANIGRAHI[3]

[1]Research Scholar, Department of ECE, Biju Patnaik University of Technology, Odisha, India

[2]Professor, Department of ECE, NIST University, Odisha, India

[3]Assistant Professor, Electrical Department, Parala Maharaja Engineering College, Odisha, India

E-Mail: [1] ashok.kavuru.gmail.com, [2] rajeshpatjoshi1@gmail.com, [3] rpanigrahi99@gmail.com

ABSTRACT

Diabetic retinopathy (DR) is a common and severe effect of diabetes that contributes significantly to vision impairment worldwide. Retinal fundus imaging is widely used for the detection and monitoring of this condition; however, accurate interpretation of these images requires substantial clinical expertise. Hence, achieving consistent and reliable assessment is challenging, particularly given the limited availability of trained specialists. To address this issue, this work presents a revised VGG16 model for automatically classifying DR types. The proposed approach was assessed using standard evaluation metrics and achieved 95.43% accuracy on the preprocessed APTOS2019 fundus image dataset. The results suggest that the proposed method is well-suited for early screening and severity grading of diabetic retinopathy, enabling the timely diagnosis and lowering the risk of vision loss.

Keywords: *Diabetic Retinopathy, VGG16, Fundus Images, APTOS2019*

1. INTRODUCTION

Over the past few years, there have been many advancements in medical care; however, Diabetic Retinopathy (DR) is one of the primary contributors to impaired vision and blindness, which can create a severe burden on global healthcare systems. Based on data from the International Diabetes Federation (IDF), diabetes affects millions of people, and nearly one in three people with diabetes develops some retinopathy. Recent statistics estimate that there are 589 million people presently living with diabetes. More than one in ten people has some form of Diabetic Retinopathy, and this figure is expected to rise with the rising incidence of diabetes, particularly in developing nations [1]. Early identification of diabetic retinopathy is vital for effective disease management, thereby preventing severe vision loss and blindness. Standard diagnostic methods, including comprehensive eye examinations, Optical Coherence Tomography (OCT), and fundus imaging, are the most widely used and cost-effective for identifying structural abnormalities in the retina [2]. Analyzing retinal images to classify DR was time-consuming and required substantial medical image interpretation. To overcome these challenges, automated diagnostic tools are essential for enabling healthcare professionals to obtain rapid, accurate evaluations. These diagnostic approaches are fundamental in detecting and classifying diabetic retinopathy [3]. As a result, growing attention is being given to advanced technologies, such as deep learning, to support automatic analysis of fundus images for diabetic retinopathy detection [4]. Deep learning represents an essential subdomain of artificial intelligence (AI). It can learn patterns from images and make decisions with minimal human involvement. Convolutional Neural Networks (CNNs) are the most commonly applied deep learning models and achieve strong performance in classification tasks by automatically learning discriminative features from raw images [5]. A widely recognized CNN architecture has achieved significant performance across diverse medical image recognition and classification tasks [6].

Diabetic Retinopathy (DR) was categorized into two broad categories: NPDR and PDR. In NPDR, it is further divided into mild, moderate, and severe stages of DR. These categories were selected based on distinct retinal imaging characteristics. The results of the method demonstrate that the proposed deep architecture framework can accurately classify

the different phases of DR from retinal fundus images. The high classification accuracy achieved highlights the model's ability to improve diagnostic reliability in DR detection. Model performance has been comprehensively evaluated using a confusion matrix and a classification report. The findings of this work highlight the usefulness of a deep neural network for effectively categorizing multiple stages of diabetic retinopathy from fundus images while maintaining efficiency and clinical applicability.

The main contribution of this article:

- To propose a revised VGG16 model for Diabetic Retinopathy classification.
- To address the class imbalance through systematic preprocessing and augmentation of the APTOS2019 dataset.
- To compare the proposed model with VGG16 and EfficientNetB3 using key evaluation metrics.

This paper is organized as follows: Section 2 summarizes the relevant literature on diabetic retinopathy classification; Section 3 details the proposed modified VGG16 model; Section 4 describes the experimental configuration and presents the results; Section 5 discusses the outcomes of the proposed approach; and Section 6 concludes the study.

2. REVIEW OF RELATED WORK

Earlier approaches to diabetic retinopathy analysis relied on manually designed features, such as texture measures, blood vessel extraction, and lesion identification, which were subsequently classified using algorithms like random forests and support vector machines. Although these techniques performed well under controlled conditions, their reliability decreased when applied to large clinical datasets with diverse image quality. Recent advances in convolution-based learning frameworks have transformed DR analysis by enabling the automatic extraction of discriminative patterns directly from retinal images. Clinically relevant factors, including age, diabetes duration, blood glucose levels, and hypertension, were identified as significant predictors. The effectiveness of the proposed approach in improving risk stratification among individuals with diabetes, supporting earlier detection and timely intervention to reduce the risk of severe retinal complications [7].

A study on the APTOS 2019 dataset evaluated VGG16, ResNet50, DenseNet121 and DenseNet169 using a novel loss function and results show that DenseNet169 achieved the best performance (96.68%), demonstrating the strength of deeper CNN models for DR classification [8]. Another method used an original VGG16 configuration and achieved 75.48% accuracy, and by introducing Dropout, and the accuracy increased to 76.57% when regularization was applied only to the classification layers and to 77.11% when regularization was applied to all layers. In contrast, a denser feature extraction in the early stages, followed by a gradual reduction in convolutional layers, achieved a superior accuracy of 81.74%, demonstrating clear performance gains over VGG16-based models [9].

A related investigation used a DenseNet-based approach to extract meaningful features from retinal fundus images. The method reported 96.11% accuracy and a quadratic weighted kappa value of 0.89 for the identification of diabetic retinopathy. When evaluated alongside a VGG16 model, DenseNet121 demonstrated better performance on this classification task [10]. Q. H. Nguyen et al. proposed an automated framework for diabetic retinopathy classification based on VGG16 and VGG19 architectures, which evaluated retinal images acquired under different illumination conditions and viewing angles. This approach achieved an accuracy of 82%, with sensitivity and specificity values of 80% and 82%, respectively, and reported an area under the ROC curve of 0.904 for grading diabetic retinopathy into five severity categories [11].

D. Singh and D. C. Dobhal, using Indian diabetic retinopathy datasets, to categorize retinal fundus images into four major lesion types: microaneurysms, soft exudates, hard exudates, and haemorrhages. In this approach, VGG16 was employed to extract features from fundus images, and logistic regression was used for final classification, yielding an overall accuracy of 90.4% [12]. B. Menaouer et al. introduced a hybrid framework combining VGG16 and VGG19 architectures for severity-based diabetic retinopathy classification. Its performance was evaluated on 5,584 fundus images collected from multiple online sources and reported an accuracy of 90.6%, recall of 95%, and F1-score of 94% [13].

C. Mohanty et al. used the APTOS 2019 dataset and analyzed it under class-imbalanced conditions, with appropriate balancing techniques applied during training. Two learning frameworks were assessed, with the hybrid configuration achieving

79.5% accuracy, while the DenseNet121 model recorded a substantially higher 97.3% accuracy. A comparison with previously reported methods on the corresponding dataset further demonstrated the improved performance of DenseNet121 [14]. In Z. Khan et al.'s study, a VGG-NiN architecture was introduced by integrating a VGG16 backbone with Spatial Pyramid Pooling (SPP) and Network-in-Network (NiN) modules. The SPP component enabled multi-scale analysis of retinal images, while the NiN structure increased model expressiveness through enhanced non-linearity. This approach achieved 85% accuracy in diabetic retinopathy classification [15].

These studies underscore the growing importance of advanced computational approaches in retinal image analysis and lesion segmentation, where automated methods have demonstrated clear improvements in both processing speed and diagnostic accuracy when compared with manual assessment. The incorporation of such frameworks into clinical workflows has the potential to streamline diabetic retinopathy screening by reducing clinician workload and enhancing diagnostic consistency. The reviewed literature indicates that modern convolutional architectures are advancing retinal imaging by enabling efficient and reliable diagnostic solutions for diabetic retinopathy. However, challenges associated with data quality, interpretability of results, and robustness across diverse populations must be addressed before these systems can be broadly implemented in routine clinical practice or large-scale screening programs. Overall, existing evidence suggests that continued methodological refinement will support earlier identification of diabetic retinopathy and improve patient outcomes.

3. PROPOSED MODEL

A convolutional network processes a fixed-size image input. It employs a set of filters to capture spatial patterns within the image, enabling the identification of characteristics such as colour variations, textures, and edge information [16]. By sharing weights and carefully distributing parameters across layers, this structure reduces computational complexity. Feature extraction is performed using kernels of predefined sizes at successive layers. To further refine the extracted information, pooling operations are applied to reduce the dimensionality of feature maps while retaining critical structural details. Among the available pooling techniques, max pooling is most commonly used, in which the input is divided into small regions, and the maximum value from each region is retained. Alternatively, average pooling computes the mean of each area, reducing data size and computational load. When trained with sufficient caution and large-scale datasets, CNN-based models achieve more reliable outcomes in DR disease detection than those trained on limited data [17]. A deep convolutional neural network manages 224 × 224 colour images across multiple convolution operations in the proposed framework. Starting with 64 filters, the network progressively increases filters to 128, 256, and 512, each using 3 × 3 kernels with ReLU activation to capture complex features. Max Pooling layers (2 × 2) follow each block to reduce dimensions and highlight key patterns, while a Global Max Pooling layer gathers the strongest activation into a compact vector. The extracted features are then fed into a stack of dense layers with integrated Dropout for regularisation, and finally, a Softmax output layer provides probabilistic classification predictions.

3.1 Input Layer

Before the extraction of features, retinal fundus images undergo several preprocessing steps to enhance image quality and maintain uniformity across the entire dataset. During this stage, images are resized to the required input resolution, and noise is reduced using filtering techniques such as Gaussian or median filtering. The proposed model is designed to process retinal fundus images of size 224 × 224 pixels with three input channels (RGB). Structural details are enhanced via green-channel extraction, while contrast-limited adaptive histogram equalization improves the visibility of lesions and blood vessels. Pixel intensities are normalized to support stable model training, and data augmentation strategies, including rotation and scaling, are employed to increase sample diversity and limit overfitting. After preprocessing, the refined images are used for feature extraction, which subsequently supports accurate classification of diabetic retinopathy.

3.2 Convolution Layer

Convolutional layers form the fundamental building blocks of the network, where spatial features are progressively learned from input images. The architecture consists of two convolutional blocks with 64 filters, two with 128 filters, three with 256 filters, and two with 512 filters. The Rectified Linear Unit (ReLU) activation is applied after each convolutional block to improve the model's ability to capture complex visual features

Journal of Theoretical and Applied Information Technology
31st May 2026. Vol.104. No.10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

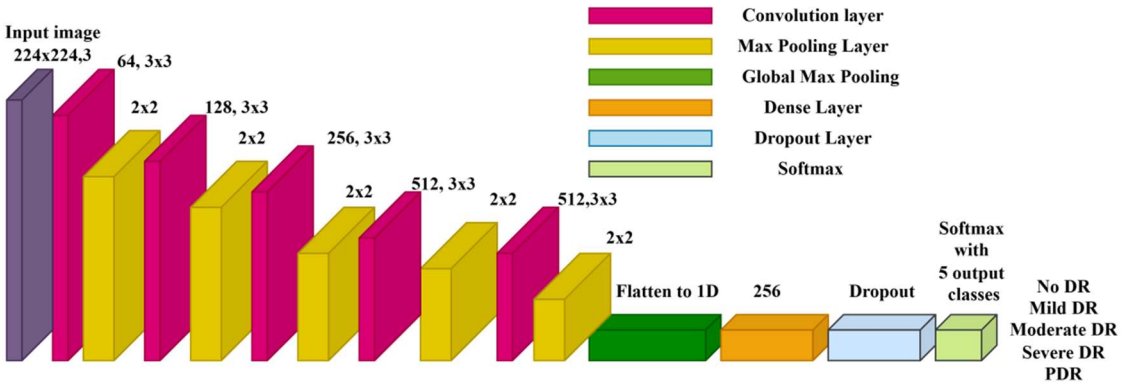

Figure 1. Modified VGG16 Architecture

in fundus images.

$O(i,j) = \sum_m \sum_n I(i+m, j+n) \cdot F(m,n) + b$ (1)

Where I(i,j) is the input image.
F(m,n) is the filter, and b is the bias.
O(i, j) represents the feature value produced at the spatial location (i, j)

3.3 Max Pooling Layer

A max pooling operation with a 2 × 2 filter size was applied after two convolutional blocks. This step down-samples the feature maps, removes redundant details, reduces overfitting, and highlights the most significant patterns in the data.

$$O(i,j) = \max_{(m,n)\epsilon R} I(i+m, j+n) \quad (2)$$

I(i,j) is the input feature
R is the pooling region
O(i,j) is the output after applying the max polling

3.4 Global Max Polling Layer

After the convolutional layers, a global maximum pooling operation is applied to each feature map, reducing it to a single value and producing a one-dimensional representation. This operation aggregates the most significant activations across all channels and passes them to fully connected layers for classification.

$$G_c = \max_{(i,j)} I_c(i,j) \quad (3)$$

I_c is the input feature map for channel c at location (i, j)
G_c is the single output value of channel C,
computed by taking the maximum across the entire feature map.

3.5 Dense Layer

After the convolutional and pooling stages, a dense layer with 256 neurons and ReLU activation is added. To limit the model overfitting, L1 (0.006) and L2 (0.016) regularisation are included. L1 regularisation can push unnecessary weights to zero, while L2 regularisation reduces overfitting by keeping weights smooth and distributed. The final fully connected layer maps these features into class-specific output probabilities for classification. Rectified linear unit activation function is in Eq.(4)

$$R(i) = \begin{cases} i, & if\ i > 0 \\ 0, & if\ i \leq 0 \end{cases} \quad (4)$$

L1 Regularisation: It can add absolute values of the weights to the loss function.

$$L_{1\ total} = L_{original} + \lambda \sum_{i=1}^{n} |\omega_i| \quad (5)$$

L2 Regularisation: It can add the square of the weights to the loss function.

$$L_{2\ total} = L_{original} + \lambda \sum_{i=1}^{n} \omega^2{}_i \quad (6)$$

L denotes the initial loss value.
λ is the regularisation strength, W_i is the weight value.

3.6 Dropout Layer

A Dropout layer with a dropout of 40% is applied after the dense layer to mitigate overfitting. A proportion of neurons is randomly deactivated during training via Dropout to force the network to learn redundant representations, making it more robust and generalizable. The final stage is a Dense layer with neurons equal to the number of classes, followed by a SoftMax activation, which produces a normalized probability distribution across classes, enabling effective multi-class classification. The

architecture sequentially applies convolution and pooling to extract features, then condenses them into a compact representation, which is classified using dense layers with a SoftMax output.

4. RESULTS

The APTOS 2019 Blindness Detection dataset from the Asia Pacific Tele-Ophthalmology Society was used in this method. The dataset includes 3,662 retinal fundus images, annotated across five diabetic retinopathy severity levels ranging from No DR (0) to Proliferative DR (4). The distribution of images across these categories is summarized in Table 1. Representative samples from each class were also examined to provide a visual understanding of the dataset [18]. To ensure reliable model development and evaluation, the dataset is split into training, validation, and test subsets with proportions of 70%, 15%, and 15%, respectively. Stratified sampling was applied to maintain the original class distribution in each subgroup. Before model training, a series of preprocessing steps was carried out to improve image quality and ensure consistency. These include cropping to eliminate dark borders, resizing to a fixed input resolution, normalizing pixel intensities, and contrast enhancement using contrast-limited adaptive histogram equalization to improve the visibility of retinal lesions. Due to the inherent class imbalance, particularly in the Mild, Severe, and Proliferative DR categories, data augmentation was applied to increase their representation. Augmentation techniques included image flipping, small-angle rotations, contrast variation, and localized histogram equalization. An augmentation factor of two was applied to these minority classes, while the No DR and Moderate classes remained unchanged. This approach enhanced data diversity and reduced bias toward dominant classes. All experiments were performed using Python on a system equipped with an NVIDIA RTX 4030 GPU. The network was trained for 100 epochs, with GPU acceleration enabling faster convergence and improved computational efficiency.

Table 1. Distribution of images across each class before and after augmentation

DR Stage	Total Images	Training Images	Validation Images	Test Images	Images after augmentation
No DR	1805	1265	270	270	1265
Mild DR	370	258	56	56	516
Moderate DR	999	699	150	150	699
Severe DR	193	133	30	30	266
PDR	295	205	45	45	410
Total	3662	2560	551	551	3156

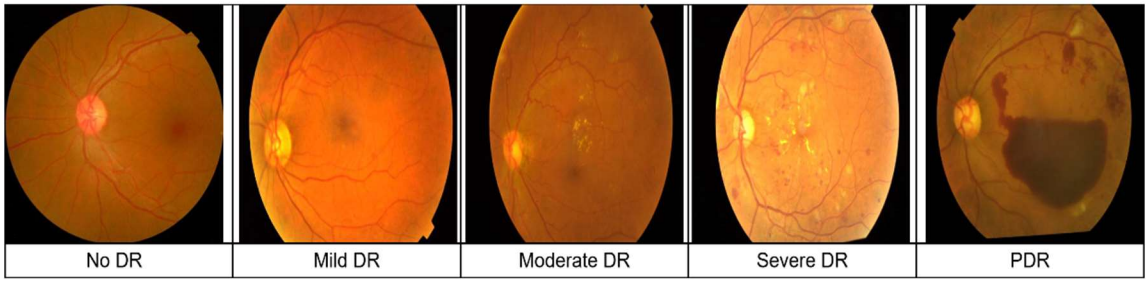

Figure 2. Sample images from the database.

Figure 2 illustrates the sequential progression of diabetic retinopathy as observed in retinal fundus images. In the absence of retinopathy (No DR), the retinal vasculature appears normal, with intact and healthy blood vessels. During the Mild stage, small microaneurysms become visible as early signs of vascular alteration. Moderate diabetic retinopathy is marked by more prominent haemorrhages and the presence of lipid exudates, reflecting increased vascular damage. In the Severe stage, extensive vessel blockage occurs, accompanied by widespread retinal lesions. Proliferative diabetic retinopathy (PDR) is the most advanced stage of the disease, characterized by

Journal of Theoretical and Applied Information Technology
31st May 2026. Vol.104. No.10

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

abnormal neovascularization in which fragile new blood vessels form and are prone to rupture, posing a high risk of vision loss. This stepwise progression underscores the importance of early detection and timely clinical intervention.

4.1. Vgg16

The baseline VGG16 architecture [19] was first evaluated, achieving an accuracy of 92.05%, indicating that the pre-trained network performs well even without architectural modifications. Subsequent experiments explored the impact of incorporating additional layers and regularisation strategies, yielding varying levels of performance enhancement. Model behaviour during training was examined by plotting training and validation accuracy over successive epochs, as shown in figure 3.

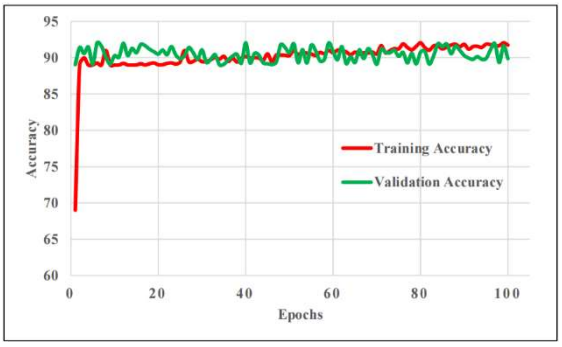

Figure 3. VGG16 Model: Training and Validation Accuracy

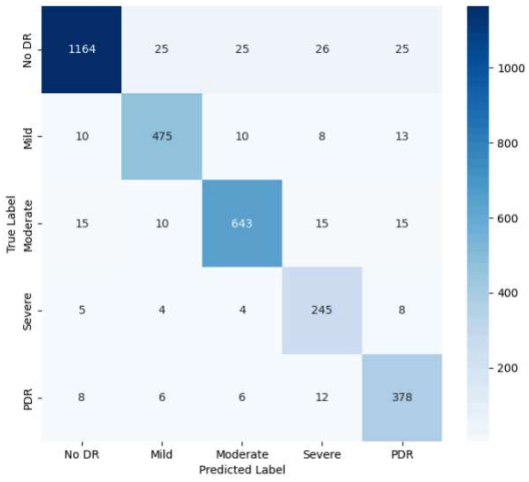

Figure 4. Confusion Matrix of VGG16 Model on APTOS 2019 Dataset

These curves offer insight into the model's learning progression and generalization capability. Furthermore, the confusion matrix shown in Fig. 4 was constructed using the validation dataset to examine class-wise prediction outcomes, thereby allowing identification of categories with firm performance and those requiring further refinement.

4.2. Efficientnetb3

In the second experiment, the EfficientNetB3 architecture [20] was selected for its compact design and favourable balance of accuracy and computational efficiency. The model was initialized with ImageNet-pretrained weights and tested under its original configuration. Under this setting, an accuracy of 93.4% was achieved, exceeding the 92.05% obtained with the VGG16 baseline. As shown in figure 5, the EfficientNetB3 model demonstrates improved learning behaviour and generalization compared with earlier architectures.

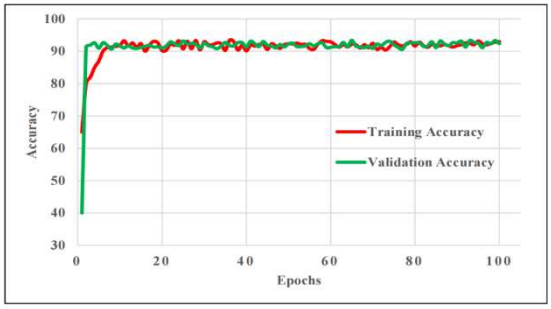

Figure 5. Training and Validation Accuracy of the EfficientNetB3 Model

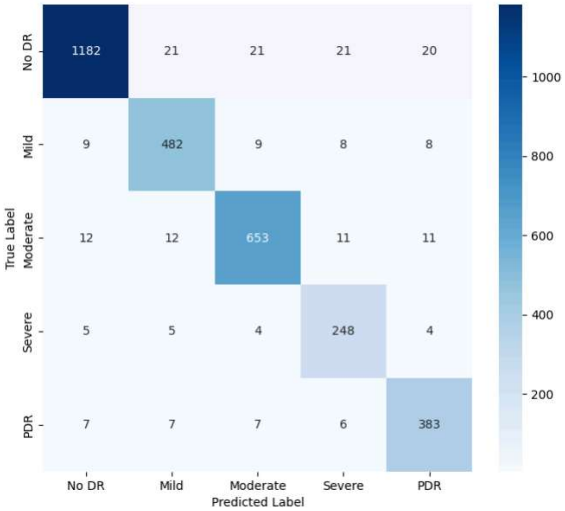

Figure 6. EfficientNetB3 Confusion Matrix on APTOS 2019 Dataset

The corresponding confusion matrix presented in figure 6 summarizes the model's classification performance on the validation set by detailing the distribution of predictions across all diabetic retinopathy classes. This analysis highlights

categories with strong classification performance and those that may benefit from further refinement.

4.3. Proposed Modified Vgg16 Method

The third experiment corresponds to the proposed approach, which incorporates batch normalization, additional dense layers, and a dropout rate of 0.4. The momentum parameter was set to 0.99 to stabilize training further. Under this configuration, the model achieved an accuracy of 95.43%, outperforming the baseline methods. The results indicate that careful adjustment of training parameters can lead to notable performance gains. Consistent with the earlier experiments, the training and validation accuracy trends were examined across epochs in Figure 7, and confusion matrices were constructed using the validation dataset to assess class-wise prediction behaviour.

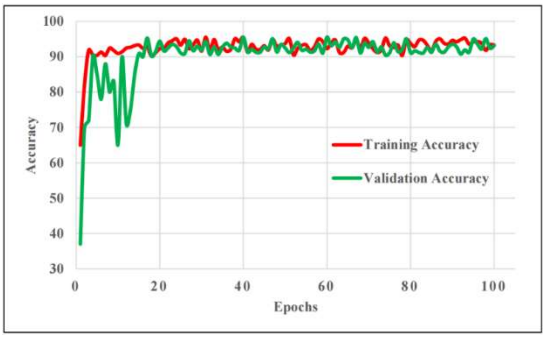

Figure 7. Proposed Model: Training and Validation Accuracy Curve

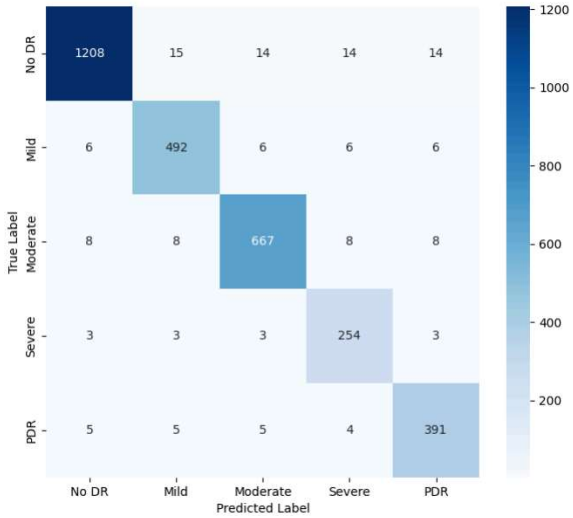

Figure 8. Proposed Model Confusion Matrix on APTOS 2019 Dataset

The confusion matrices provided a detailed view of correct and incorrect predictions, enabling identification of specific categories where classification performance was limited and indicating areas that require further refinement. The observed improvements also demonstrate that slight adjustments to training parameters can lead to meaningful gains in overall accuracy. Consistent with the earlier experiments, the training and validation accuracy trends were plotted across epochs to illustrate performance stability and convergence. As medical image analysis is often time-consuming and depends heavily on expert knowledge, automated tools can help make clinical decision-making quicker and more dependable. Confusion matrix–based evaluation provides a clear summary of class-wise prediction outcomes, highlighting performance gaps and guiding future improvements, as illustrated in Figure 8.

5. DISCUSSION ON RESULTS

The effectiveness of the model was assessed using accuracy, precision, recall, and F1-score, as described in Equations (7)–(10). Precision indicates the proportion of correctly classified positive samples among all predicted positives, highlighting the dependability of diabetic retinopathy detection from fundus images. Recall measures how well the model identifies true positive cases. The F1-score, computed as the harmonic mean of precision and recall, provides a balanced evaluation that considers both detection accuracy and completeness.

$$Precision = \frac{TP}{TP+} \quad (7)$$

$$Recall\ or\ Sensitivity = \frac{TP}{TP+FN} \quad (8)$$

$$F_1 - Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (9)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

In this study, classification outcomes were interpreted using standard definitions: a true positive (TP) corresponds to a retinal fundus image correctly identified as having diabetic retinopathy. In contrast, a true negative (TN) refers to a normal fundus image correctly classified as No DR. A false positive (FP) occurs when a normal image is incorrectly labeled as diabetic retinopathy. In contrast, a false negative (FN) is a case of diabetic retinopathy that is incorrectly classified as non-diabetic. Model performance was further enhanced by carefully selecting training parameters to improve classification reliability. Table 2 presents the

hyperparameters; a dropout rate of 0.4 was used to reduce overfitting by randomly deactivating neurons during training.

Table 2. Hyperparameters and Optimizer settings

S.No	Parameter	Value
1	Dropout rate	0.4
2	Momentum	0.99
3	Optimizer	SDG
4	Learning rate	0.001
5	Batch size	32
6	Number of epochs	100

To promote stable convergence and smoother parameter updates, a momentum value of 0.99 was incorporated. Optimization was performed using stochastic gradient descent with a learning rate of 0.001, resulting in consistent loss minimization. The model was trained for 100 epochs, with a batch size of 32, allowing sufficient iterations for effective learning of complex retinal features and improved classification accuracy. VGG16, EfficientNetB3, and the proposed modified VGG16 were comparatively evaluated for diabetic retinopathy from retinal fundus images. Their performance across the five severity stages is summarized in Table 3. The comparison was based on commonly used evaluation measures, including precision, recall, and F1-score, which collectively provide a comprehensive assessment of classification performance.

5.1 Vgg16

The VGG16 model generally produced satisfactory results; however, performance varied across disease categories. For the No DR class, a high F1-score of 94.37% was achieved, although the recall value of 92.02% suggests that some typical cases were not fully captured. Classification performance for Mild and Moderate DR remained relatively balanced, with F1-scores of 91.70% and 92.79%, respectively.

In contrast, the Severe DR category showed reduced precision (80.07%) despite a high recall (92.11%), indicating an increased number of false-positive predictions. For the Proliferative DR class, the F1-score reached 89.05%, though precision was comparatively lower. Overall, while VGG16 demonstrated reasonable classification capability, its performance was less consistent for the more advanced stages of diabetic retinopathy.

Table 3. Comparison of VGG16, EfficientNet, and the Proposed Model on the APTOS 2019

Model	Disease	TP	FP	FN	TN	Precision	Recall	F1-Score
Vgg16 Model	No DR	1164	38	101	2210	96.84	92.02	94.37
	Mild DR	475	45	41	2952	91.35	92.05	91.7
	Moderate	643	45	55	2770	93.46	92.12	92.79
	Severe	245	61	21	3186	80.07	92.11	85.66
	PDR	378	61	32	3042	86.15	92.2	89.05
EfficientNetB3 Model	No DR	1182	33	83	2215	97.28	93.44	95.32
	Mild DR	482	45	34	2952	91.46	93.41	92.43
	Moderate	653	41	46	2773	94.1	93.23	93.75
	Severe	248	46	18	3201	84.35	93.41	88.57
	PDR	383	43	27	3060	89.91	93.4	91.62
Proposed Modified VGG16 Model	No DR	1208	22	57	2226	98.2	95.5	96.83
	Mild DR	492	31	24	2966	94.07	95.35	94.7
	Moderate	667	28	32	2786	95.97	95.42	95.7
	Severe	254	32	12	3215	88.8	95.49	92.03
	PDR	391	31	19	3072	92.65	95.37	94

5.2 Efficientnetb3model

EfficientNetB3 demonstrated superior performance compared with VGG16 across most disease categories. For the No DR class, an F1-score of 95.32% was achieved, with a precision of 97.28%, indicating a lower false-positive rate. In the Mild and Moderate DR categories, both precision and recall exceeded 91%, resulting in well-balanced classification performance with F1-Scores of 92.43% and 93.75%, respectively. Performance in the Severe DR category showed noticeable improvement over VGG16, achieving an F1-score of 88.57%, although precision remained relatively lower at 84.35%. For Proliferative DR, the model achieved an F1-score of 91.62%, with an overall accuracy of 89.91% and a recall of 93.40%. These findings suggest that EfficientNetB3 provides more reliable classification than VGG16, particularly for early and intermediate stages of diabetic retinopathy.

5.3 Proposed Modified VGG16 Model

The proposed approach demonstrated consistently superior performance compared with both baseline models across all diabetic retinopathy categories. For the No DR class, the method achieved the best overall results, with a precision of 98.20%, a recall of 95.50%, and an F1-score of 96.83%, indicating effective separation between normal and affected retinal images. In the Mild DR category, an F1-score of 94.70% was achieved, primarily driven by a high recall of 95.35%. Performance further improved for Moderate DR, with precision and recall reaching 95.97% and95.42%, respectively, resulting in an F1-score of 95.70%, reflecting the most balanced performance across all classes.

Notably, even in the Severe DR category, classification is typically more challenging, and the proposed method achieved an F1-score of 92.03%, with a strong recall of 95.49%, thereby reducing the likelihood of missed cases. In the Proliferative DR dataset, the model achieved an F1-score of 94.00%, with precision and recall of 92.65% and 95.37%, respectively. These results demonstrate the robustness and clinical relevance of the proposed framework, particularly for detecting advanced stages of diabetic retinopathy. As illustrated in Figure 9, a comparative bar chart summarizes the performance of VGG16, EfficientNetB3, and the proposed method on the test dataset using two evaluation measures. VGG16 recorded the lowest performance levels, with scores of 90.71% and 92.05%, while EfficientNetB3 showed moderate improvement, achieving 92.34% and 93.40%. In contrast, the proposed method outperformed both baselines, gaining 94.65% and 95.43%, indicating greater stability and reliability. Overall, the visual comparison confirms that the proposed framework consistently delivers greater accuracy and robustness than existing architectures.

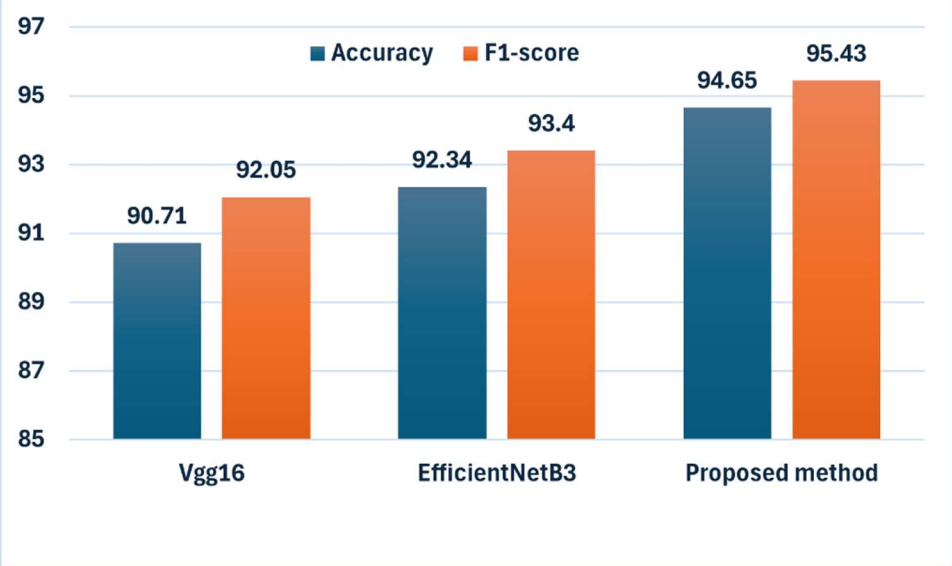

Figure 9. Comparative Performance Analysis Proposed Approach on Test Set

6. CONCLUSIONS

This paper presents a modified VGG-16-based framework for classifying diabetic retinopathy using retinal fundus images. Initial experiments with the baseline VGG16 model yielded an accuracy of 92.18%, demonstrating its suitability for stage-wise analysis of diabetic retinopathy. Subsequent experiments involving the addition of dense layers and regularization showed a temporary decline in performance, highlighting the architecture's sensitivity to hyperparameter selection. By systematically tuning the network, particularly with a dropout rate of 0.4 and a momentum value of 0.99, the proposed revised VGG16 model achieved a test accuracy of 95.43%, outperforming both the original VGG16 and EfficientNetB3 models. These results indicate that appropriate regularization and parameter tuning are crucial for improving generalization and mitigating the overfitting problem. While the proposed approach demonstrates strong performance, certain limitations remain. The evaluation was conducted on a single dataset with inherent class imbalance, and only a limited range of architectural variations was examined.

Future research may address these constraints by exploring alternative network designs, incorporating multi-source retinal datasets, and adopting more systematic optimization strategies to enhance robustness and generalizability. It would also be valuable to validate the system with doctors and real patient data to ensure reliability and practical usefulness in diabetic retinopathy screening.

REFERENCES:

[1] "Diabetes Facts and Figures | International Diabetes Federation." Accessed: Aug. 31, 2025. [Online].https://idf.org/about-diabetes/diabetes-facts-figures

[2] Y. Attiku, Y. He, M. G. Nittala, and S. R. Sadda, "Current status and future possibilities of retinal imaging in diabetic retinopathy care applicable to low- and medium-income countries," Indian J. Ophthalmol., vol. 69, no. 11, pp. 2968–2976, Nov. 2021.
doi: 10.4103/ijo.IJO_1212_21.

[3] D. S. W. Ting et al., "Artificial intelligence and deep learning in ophthalmology," Br. J. Ophthalmol., vol. 103, no. 2, pp. 167–175, Feb. 2019.doi:10.1136/bjophthalmol-2018-313173.

[4] L. Arora et al., "Ensemble deep learning and EfficientNet for accurate diagnosis of diabetic retinopathy," Sci. Rep., vol. 14, no. 1, p. 30554, Dec. 2024, doi: 10.1038/s41598-024-81132-4.

[5] W. H. Lopez Pinaya, S. Vieira, R. Garcia-Dias, and A. Mechelli, "Convolutional neural networks," in Machine Learning, Elsevier, 2020, pp. 173–191. doi: 10.1016/B978-0-12-815739-8.00010-9.

[6] D. Shen, L. Zhou, and M. Liu, "Deep Learning Models with Applications to Brain Image Analysis," in Neural Engineering, B. He, Ed., Cham: Springer International Publishing, 2020, pp. 433–462.doi: 10.1007/978-3-030-43395-6_15.

[7] D. Tarasewicz et al., "Development and Validation of a Diabetic Retinopathy Risk Stratification Algorithm," Diabetes Care, vol. 46, no. 5, pp. 1068–1075, May 2023, doi: 10.2337/dc22-1168.

[8] D. Singh, Dinesh C. Dobhal, and Janmejay Pant, "Diagnostic System Based on Deep Learning to Detect Diabetic Retinopathy" ,J. Ophthalmol., vol. 40, no. 3, July 2024,
doi: 10.36351/pjo.v40i3.1771.

[9] Ö. Çağrı Yavuz, U. Yavuz, and İ. Akgül, "Deep Learning Based Models for Detection of Diabetic Retinopathy," Teh. Glas., vol. 17, no. 4, pp. 581–587, Oct. 2023.
doi: 10.31803/tg-20220905123827.

[10] S.Mishra, S. Hanchate, and Z. Saquib, "Diabetic Retinopathy Detection using Deep Learning," in 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), Bengaluru, India: IEEE, Oct. 2020, pp. 515–520.
doi: 10.1109/ICSTCEE49637.2020.9277506.

[11] Q. H. Nguyen et al., "Diabetic Retinopathy Detection using Deep Learning," in Proceedings of the 4th International Conference on Machine Learning and Soft Computing, Haiphong City, Vietnam: ACM, Jan. 2020, pp. 103–107.
doi: 10.1145/3380688.3380709.

[12] D. Singh and D. C. Dobhal, "A Deep Learning-based Transfer Learning Approach Fine-Tuned for Detecting Diabetic Retinopathy," Procedia Comput. Sci., vol. 233, pp. 444–453, 2024,
doi: 10.1016/j.procs.2024.03.234.

[13] B. Menaouer, Z. Dermane, N. El Houda Kebir, and N. Matta, "Diabetic Retinopathy Classification Using Hybrid Deep Learning Approach," SN Comput. Sci., vol. 3, no. 5, p. 357, July 2022,
doi: 10.1007/s42979-022-01240-8.

[14] C. Mohanty et al., "Using Deep Learning Architectures for Detection and Classification of

Diabetic Retinopathy," Sensors, vol. 23, no. 12, p. 5726, June 2023, doi: 10.3390/s23125726.

[15] Z. Khan et al., "Diabetic Retinopathy Detection Using VGG-NIN a Deep Learning Architecture," IEEE Access, vol. 9, pp. 61408–61416, 2021
doi: 10.1109/ACCESS.2021.3074422.

[16] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in 2017 International Conference on Engineering and Technology (ICET), Antalya: IEEE, Aug. 2017, pp. 1–6.
doi: 10.1109/ICEngTechnol 2017.8308186.

[17] C. Bailer, T. Habtegebrial, K. Varanasi, and D. Stricker, "Fast Feature Extraction with CNNs with Pooling Layers," 2018, arXiv.
doi: 10.48550/ARXIV.1805.03096.

[18] "APTOS 2019 Blindness Detection." Accessed: Aug. 11, 2025. [Online]. Available: https://kaggle.com/aptos2019-blindness-detection

[19] D. A. Da Rocha, F. M. F. Ferreira, and Z. M. A. Peixoto, "Diabetic retinopathy classification using VGG16 neural network," Res. Biomed. Eng., vol. 38, no. 2, pp. 761–772, June 2022
doi: 10.1007/s42600-022-00200-8.

[20] R. Alsohemi and S. Dardouri, "Fundus Image-Based Eye Disease Detection Using EfficientNetB3 Architecture," J. Imaging, vol. 11, no. 8, p. 279, Aug. 2025
doi: 10.3390/jimaging11080279