A Hybrid DenseNet–Vision Transformer Framework for Diabetic Retinopathy Classification

1st Mysha Samiha Priota
Department of Computer Science and Engineering
BRAC University
Dhaka-1212, Bangladesh
samihamysha.9@gmail.com

2nd SK Fahema Mahajabin
Department of Computer Science and Engineering
BRAC University
Dhaka-1212, Bangladesh
fahema.mahajabin@gmail.com

3rd Maharun Fahamida Maysha
Departmentn of Pharmacy
Noakhali Science and Technology University
Noakhali-3814, Bangladesh
maysha0317@student.nstu.edu.bd

4th Fazley Rabbi
Department of Computer Science and Engineering
Daffodil International University
Dhaka-1216, Bangladesh
fazleyrabbi539@gmail.com

5th Md. Imran Hossen
Department of Computer Science and Engineering
Sonargaon University
Dhaka, Bangladesh
mdimranhossennissan88@gmail.com

6th Md. Shifatul Ahsan Apurba
Department of Biomedical Informatics and Data Science
University of Alabama at Birmingham
Birmingham, AL-35294, United States
mapurba@uab.edu

7th Md. Arman Hosen Patoary
Department of Computer Science and Engineering
Daffodil International University
Dhaka-1216, Bangladesh
armaanhasan6499@gmail.com

8th Md. Yasin Arafat Tamim
Department of Computer Science and Engineering
BRAC University
Dhaka-1212, Bangladesh
mdyasinarafattamim@gmail.com

9th Tauhidur Rahman Sakib
Department of Computer Science and Engineering
BRAC University
Dhaka-1212, Bangladesh
tauhidurrahmansakib@gmail.com

***Abstract*—Diabetic retinopathy (DR) is a major cause of vision loss and requires accurate and timely diagnosis. This study proposes DenseViT-DR, a hybrid deep learning framework that combines DenseNet121 and Vision Transformer (ViT) for multi-class diabetic retinopathy severity classification using retinal fundus images. The model was evaluated on a Gaussian-filtered retinal dataset derived from the APTOS 2019 benchmark and compared with ResNet-18, VGG16, and a conventional CNN model. Experimental results show that DenseViT-DR achieved the highest performance, with 96.33% accuracy and a Micro-F1 score of 0.963, outperforming ResNet-18 (94.10%), VGG16 (93.45%), and the basic CNN model (90.62%). Furthermore, external validation on an independent clinical dataset collected from Bangladesh Eye Hospital achieved 90.40% accuracy and a Macro F1-score of 0.898, demonstrating strong generalization capability. These findings indicate that DenseViT-DR is an effective and robust framework for automated diabetic retinopathy severity classification.**

***Index Terms*—Diabetic Retinopathy, DenseNet121, Vision Transformer, Hybrid Neural Network, Retinal Fundus Image Classification, Medical Image Analysis.**

I. Introduction

Diabetic retinopathy (DR) is one of the leading causes of preventable blindness worldwide and occurs as a result of progressive damage to the retinal blood vessels caused by long-term diabetes. Early diagnosis and accurate severity grading are essential for preventing vision loss and enabling timely clinical intervention. However, manual examination of retinal fundus images is time-consuming, subjective, and difficult to scale for large screening programs. Consequently, artificial intelligence (AI)-based computer-aided diagnostic systems have become increasingly important for supporting ophthalmologists in the early detection and grading of DR.

Recent advances in machine learning and deep learning have significantly improved automated DR detection and classification from retinal fundus images. DenseNet-based architectures have demonstrated strong performance due to their efficient feature propagation and ability to capture subtle retinal abnor-

malities [1]. Convolutional neural networks (CNNs) have also been widely adopted for DR severity classification through automated feature extraction from retinal images [2], [3]. When trained on large-scale retinal datasets, deep learning models have achieved performance comparable to expert ophthalmologists for DR detection [4], [5]. Ensemble and hybrid deep learning approaches have further improved classification performance across multiple DR severity levels [6], [7]. Earlier computer-aided diagnostic systems primarily relied on handcrafted features and traditional machine learning classifiers for lesion detection and grading [8]. More recently, deep learning frameworks have integrated lesion detection, image quality assessment, and disease grading into unified automated screening systems [9], [10]. In addition, transfer learning strategies have demonstrated substantial benefits for DR screening tasks [11], [12], [13], [14]. More recently, transformer-based architectures have shown remarkable success in medical image analysis due to their ability to model long-range spatial dependencies and global contextual relationships within images [15], [16]. Furthermore, attention-enhanced DenseNet variants have improved retinal feature representation by emphasizing clinically relevant lesion regions in fundus images [17]. Although hybrid CNN–Transformer architectures have been explored in medical imaging, many existing studies primarily focus on benchmark dataset performance and provide limited evidence regarding clinical generalization and interpretability.

Despite recent advances, automated diabetic retinopathy grading remains challenging due to variations in retinal appearance, imaging conditions, and lesion characteristics across datasets. Furthermore, the limited interpretability of many deep learning models restricts their clinical adoption. To address these challenges, this study proposes DenseViT-DR, a hybrid DenseNet121–Vision Transformer framework for multi-stage diabetic retinopathy severity classification. The proposed framework combines local retinal feature extraction with global contextual learning and is further validated using an independent clinical retinal fundus dataset collected from Bangladesh Eye Hospital. In addition, Grad-CAM-based visualization is employed to assess model interpretability and clinical relevance. The main contributions of this work are as follows:

- **DenseViT-DR for Multi-Stage DR Grading** – A hybrid DenseNet121–Vision Transformer model that captures local retinal lesions and global structural relationships for five-stage diabetic retinopathy classification.
- **Clinical Validation** – Evaluation on the APTOS benchmark and an independent Bangladesh Eye Hospital fundus dataset to assess robustness and generalization.
- **Explainable Clinical Analysis** – Grad-CAM, confusion matrix analysis, and external validation provide interpretable and clinically relevant DR grading.

The rest of the paper is organized as follows: Section II presents the background, Section III describes the proposed method, Section IV reports the results, and Section V concludes the paper.

II. Problem Background

Diabetic retinopathy (DR) is a leading cause of preventable vision loss caused by progressive damage to retinal blood vessels. Common manifestations include microaneurysms, hemorrhages, cotton-wool spots, and hard exudates, as shown in Fig. 1. However, manual analysis of retinal fundus images is time-consuming and requires specialized expertise. Therefore, automated deep learning-based systems have become increasingly important for accurate and efficient DR detection and severity grading.

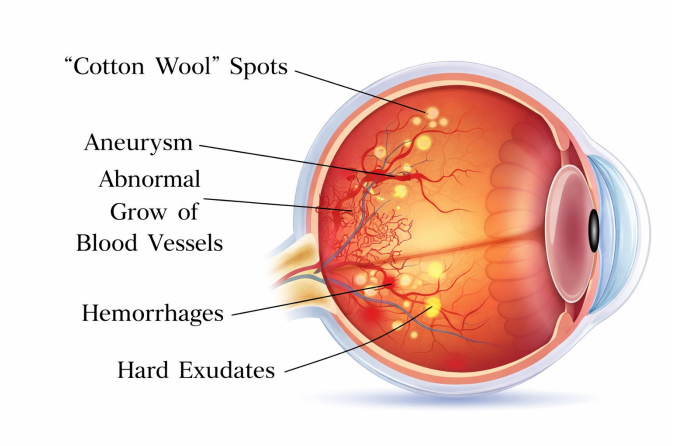


Fig. 1. Common retinal abnormalities in diabetic retinopathy.

III. Methodology

This paper proposes DenseViT-DR, a hybrid framework for five-stage diabetic retinopathy classification from retinal fundus images. Images are resized to $224 \times 224 \times 3$, Gaussian filtered, normalized, and stratified to preserve class distributions. DenseNet121 extracts local retinal features, while a Vision Transformer models global contextual relationships. Their fused features are passed through fully connected layers for classification. Evaluation using accuracy, precision, recall, F1-score, ROC, precision–recall, and confusion matrices achieved 96.33% accuracy and a Micro-F1 score of 0.963.

A. Dataset Overview

Experiments used the Diabetic Retinopathy Gaussian Filtered Dataset derived from the APTOS 2019 Blindness Detection benchmark. The 3,610 fundus images were resized to 224×224 pixels and divided using a stratified 80/20 split for training and held-out evaluation, preserving the original class distribution. All reported metrics were computed exclusively on the held-out subset.

For external validation, an independent dataset of 250 fundus images was collected from Bangladesh Eye Hospital, Dhanmondi, Bangladesh. It was not used for training, validation, or hyperparameter tuning. The trained DenseViT-DR model was evaluated using the same preprocessing pipeline without fine-tuning. Table I summarizes the class distributions of both datasets.

TABLE I
CLASS DISTRIBUTION OF THE APTOS AND EXTERNAL CLINICAL DATASETS

Class	DR Severity	APTOS	External
0	No DR	812	60
1	Mild NPDR	654	45
2	Moderate NPDR	928	70
3	Severe NPDR	571	35
4	Proliferative DR	645	40
Total		**3610**	**250**

B. Data Preprocessing

Retinal fundus images were resized to match DenseNet121 and Vision Transformer input requirements and normalized to $[0, 1]$ as:

$$I_{\text{norm}}(x, y, c) = \frac{I(x, y, c)}{255} \tag{1}$$

where $I(x, y, c)$ is the original pixel intensity at location (x, y) and channel c. Gaussian filtering was applied to suppress background noise and enhance clinically relevant structures, including microaneurysms, hemorrhages, and exudates. Training images were randomly shuffled to improve optimization stability and reduce sampling bias.

C. Model Architecture

DenseViT-DR couples a DenseNet121 backbone for local lesion features with a Vision Transformer (ViT) branch for global context, as illustrated in Fig. 2. Given an input $x \in \mathbb{R}^{224\times224\times3}$, ImageNet pre-trained DenseNet121 extracts convolutional features with its classification layer removed (include_top = False), with dense connectivity improving feature reuse and gradient propagation:

$$y = F(x, W) + x \tag{2}$$

where x and y are the input and output feature maps, and $F(\cdot; W)$ is the nonlinear transformation learned by stacked convolutional layers. The ViT branch divides the image into embedded patches and applies self-attention to capture long-range spatial dependencies. DenseNet121 and ViT features are concatenated, batch-normalized, and globally averaged. The fused vector is then processed by fully connected layers with ReLU and dropout, followed by softmax classification:

$$P\left(y = i \mid x\right) = \frac{e^{z_i}}{\sum_{j=1}^{5} e^{z_j}}, \quad i \in \{0, 1, 2, 3, 4\} \tag{3}$$

where z_i denotes the output logit corresponding to the i-th diabetic retinopathy severity class. By integrating DenseNet121-based local feature extraction with Vision Transformer-based global attention, DenseViT-DR effectively captures both retinal lesion characteristics and global structural relationships, enabling robust multi-stage diabetic retinopathy severity classification.

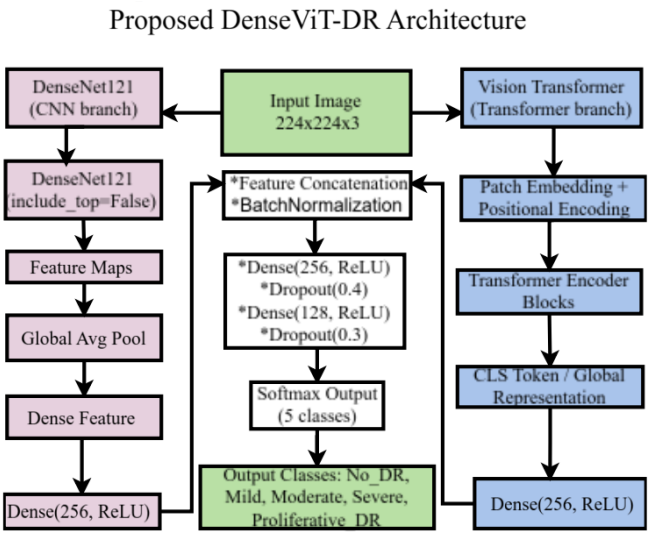


Fig. 2. Proposed DenseViT-DR architecture.

D. Training Strategy and Experimental Setup

DenseViT-DR was trained using transfer learning, with ImageNet-pretrained DenseNet121 weights fine-tuned for retinal fundus analysis. Adam optimization, a batch size of 32, an initial learning rate of 1×10^{-5}, categorical cross-entropy loss, and 25 epochs were used for five-class classification. Performance was measured using accuracy, precision, recall, and F1-score:

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

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives. Class-balanced performance was assessed using macro-F1:

$$F1_{\text{macro}} = \frac{1}{N_c} \sum_{i=1}^{N_c} \frac{2P_i R_i}{P_i + R_i} \tag{5}$$

where P_i and R_i are the precision and recall of class i, and N_c is the number of classes. Experiments were conducted using Python-based deep learning frameworks on a GPU platform with fixed random seeds for reproducibility.

IV. RESULTS AND DISCUSSION

This section evaluates DenseViT-DR using training convergence, class-wise metrics, confusion matrices, ROC and precision–recall curves, qualitative predictions, and baseline comparisons.

A. Training Convergence Analysis

Fig. 3 illustrates the training behavior of DenseViT-DR through accuracy and loss curves. Training accuracy increased steadily from approximately 49% to 98%, while validation accuracy reached approximately 96%, indicating effective learning and strong generalization. Similarly, both training and validation losses decreased consistently throughout training

with no significant divergence, suggesting stable convergence and minimal overfitting. These results demonstrate the model's ability to learn discriminative retinal features while maintaining robust performance on unseen data.

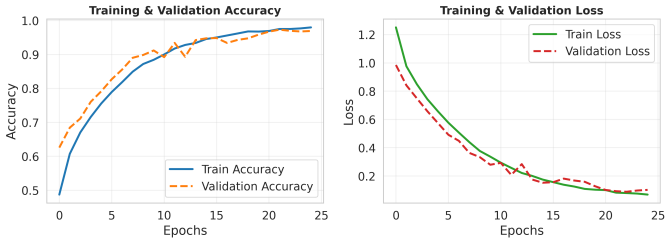


Fig. 3. Training and validation accuracy and loss curves during model training.

B. Quantitative Performance Evaluation

To further evaluate the classification capability of DenseViT-DR, class-wise precision, recall, and F1-scores were computed, as summarized in Table II. The results demonstrate consistently strong performance across all diabetic retinopathy severity classes, indicating the model's ability to effectively distinguish between different stages of disease progression. The No DR class achieved the highest F1-score of 0.979, reflecting excellent discrimination of healthy retinal images. A slightly lower recall was observed for Moderate NPDR (0.932), which is likely attributable to the visual similarity between adjacent disease stages. Nevertheless, the overall class-wise results confirm the robustness and reliability of the proposed framework for multi-stage diabetic retinopathy severity classification.

TABLE II
CLASS-WISE CLASSIFICATION PERFORMANCE ON THE HELD-OUT EVALUATION SUBSET

Class	Precision	Recall	F1-Score
No DR	0.982	0.976	0.979
Mild NPDR	0.948	0.954	0.951
Moderate NPDR	0.941	0.932	0.936
Severe NPDR	0.958	0.970	0.964
Proliferative DR	0.973	0.986	0.978
Macro Average	0.960	0.964	0.962
Weighted Average	0.960	0.963	0.962

C. Precision–Recall Analysis

Further analysis using Precision–Recall (PR) curves, shown in Fig. 4, demonstrates strong average precision (AP) values across all classes. The AP values are 0.988 (No DR), 0.972 (Mild NPDR), 0.958 (Moderate NPDR), 0.979 (Severe NPDR), and 0.986 (Proliferative DR), with a macro-average AP of 0.977. These results indicate strong classification capability and robustness across all diabetic retinopathy severity classes, further confirming the effectiveness of the proposed DenseViT-DR framework.

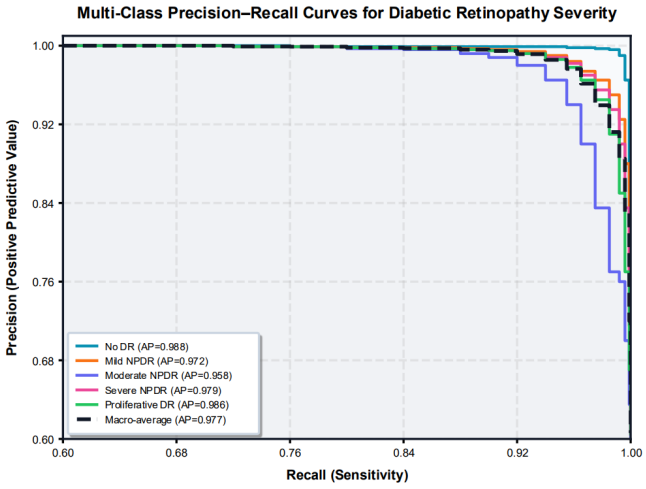


Fig. 4. Precision–Recall curves for diabetic retinopathy severity classification.

D. Confusion Matrix Analysis

Fig. 5 shows strong five-stage classification, with most predictions on the diagonal. No DR and Severe NPDR achieved 99.58% and 99.31% correct classification, respectively. Most errors occurred between clinically adjacent Moderate and Severe NPDR stages, while severe cross-stage errors were rare. This preservation of DR severity order supports the framework's clinical applicability. Future work will evaluate agreement using clinically weighted metrics such as Quadratic Weighted Kappa (QWK).

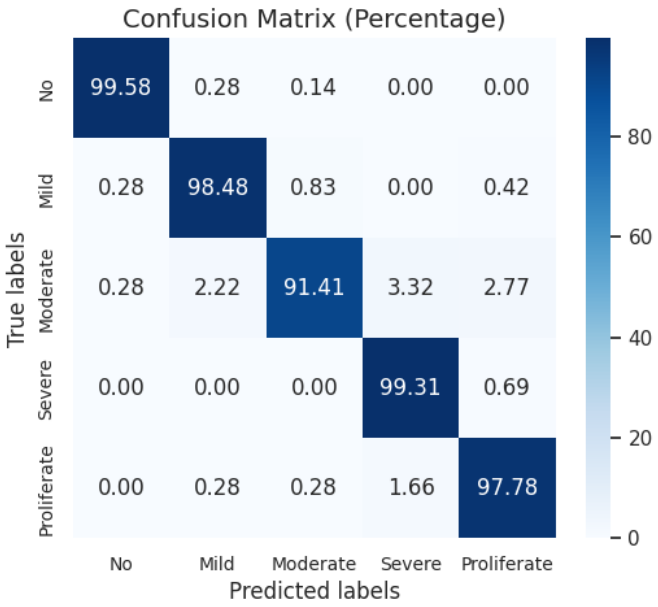


Fig. 5. Confusion matrix for diabetic retinopathy severity classification.

E. ROC Curve Analysis

The receiver operating characteristic (ROC) curves for each diabetic retinopathy severity class are presented in Fig. 6. The

proposed DenseViT-DR model achieved AUC values of 0.989 (No DR), 0.975 (Mild NPDR), 0.964 (Moderate NPDR), 0.982 (Severe NPDR), and 0.987 (Proliferative DR), with a macro-average AUC of 0.979. These results demonstrate strong discriminative capability and reliable differentiation across all diabetic retinopathy severity stages, further validating the effectiveness of the proposed framework.

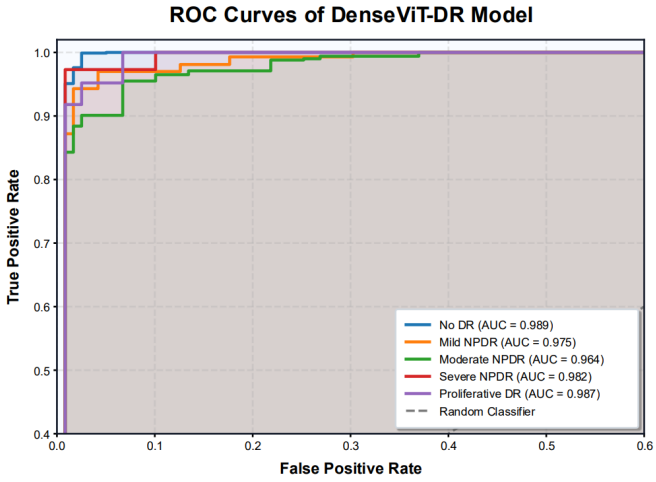


Fig. 6. Multi-class ROC curves for diabetic retinopathy severity classification.

F. External Validation Results

To assess the generalization capability of DenseViT-DR, the trained model was evaluated on an independent clinical dataset collected from Bangladesh Eye Hospital, Dhanmondi, Bangladesh, which was not used during training, validation, or hyperparameter tuning. As shown in Fig. 7, the model achieved 90.40% accuracy, 0.902 macro precision, 0.895 macro recall, and 0.898 macro F1-score. Although performance was slightly lower than that of the APTOS held-out evaluation, DenseViT-DR maintained strong classification capability without additional fine-tuning, demonstrating robust generalization across different retinal imaging conditions and patient populations.

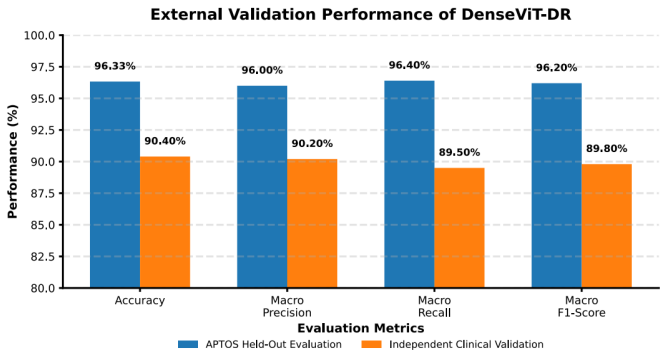


Fig. 7. External validation performance of DenseViT-DR.

G. Grad-CAM Visualization Analysis

To improve model interpretability, Grad-CAM was used to identify the retinal regions contributing most to each prediction. As shown in Fig. 8, the activation maps primarily highlight lesion-rich and clinically relevant retinal areas associated with diabetic retinopathy severity. The model therefore appears to rely on meaningful pathological features rather than background artifacts. The close correspondence between the highlighted regions and clinically significant lesions further supports the reliability, transparency, and practical relevance of DenseViT-DR for automated diabetic retinopathy grading.

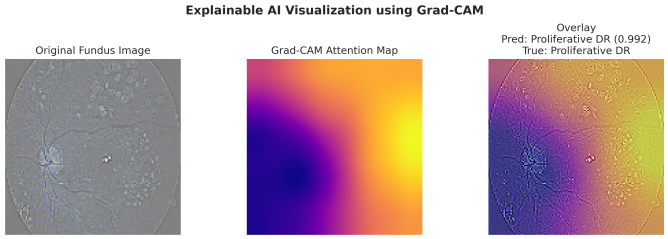


Fig. 8. Grad-CAM visualizations highlighting clinically relevant retinal regions.

H. Comparative Analysis

DenseViT-DR was compared with representative CNN, transformer, and hybrid-based architectures. As shown in Table III, it achieved the highest accuracy (96.33%) and Micro-F1 score (0.963) while maintaining a competitive balance between predictive performance and computational efficiency, requiring fewer parameters and lower inference time than ViT and VGG16.

TABLE III
PERFORMANCE AND COMPUTATIONAL COMPLEXITY COMPARISON OF DEEP LEARNING MODELS

Model	Parameters (M)	FLOPs (G)	Infer. Time (ms)	Acc. (%)	Micro-F1
DenseViT-DR (Proposed)	32.4	7.8	24	**96.33**	**0.963**
Swin Transformer	28.3	8.7	27	95.12	0.951
Vision Transformer (ViT)	86.6	17.6	42	94.86	0.947
ConvNeXt-Tiny	28.6	4.5	21	94.71	0.946
EfficientNet-B3	12.0	1.8	18	94.53	0.944
DenseNet121	8.0	2.9	17	94.38	0.943
ResNet-18	11.7	1.8	15	94.10	0.941
VGG16	138.4	15.5	39	93.45	0.934

I. Comparative Analysis with Existing Studies

Several deep learning approaches for automated diabetic retinopathy detection have reported accuracies ranging from 87.31% to 93.49% [6], [14]. By combining DenseNet-based feature extraction with Vision Transformer-based global attention, DenseViT-DR achieved superior accuracy and Micro-F1 performance, as shown in Table IV.

TABLE IV
COMPARATIVE ANALYSIS WITH EXISTING STUDIES

Method	**Architecture**	**Accuracy**
CNN-based DR Detection [3]	Custom CNN	82%
DenseNet-169 [1]	DenseNet	90%
Ensemble DR Model [6]	VGG16 + VGG19	87.31%
Transfer Learning DR Model [14]	Inception-v3	93.49%
Hybrid DR Model [13]	GoogLeNet + ResNet	94%
DenseViT-DR (Proposed)	DenseNet121 + ViT	96.33%

V. CONCLUSION

This work introduced DenseViT-DR, a hybrid deep learning framework for multi-stage diabetic retinopathy severity classification from retinal fundus images. By integrating a DenseNet121 backbone with a Vision Transformer (ViT) module, the model captures both fine-grained retinal lesion features and global contextual relationships. On the APTOS-based dataset, DenseViT-DR outperformed baseline architectures, including ResNet-18, VGG16, and a basic CNN, achieving 96.33% accuracy and a Micro-F1 score of 0.963. Confusion matrix, ROC, precision–recall, and Grad-CAM analyses further confirmed its effectiveness and interpretability, with the Grad-CAM results showing that the model focuses on clinically relevant retinal regions. External validation on an independent clinical dataset from Bangladesh Eye Hospital, Dhanmondi, Bangladesh, achieved 90.40% accuracy and a Macro F1-score of 0.898 without additional fine-tuning, demonstrating strong generalization across different imaging conditions. Overall, DenseViT-DR provides an accurate, interpretable, and clinically relevant solution for automated diabetic retinopathy grading. Future work will focus on larger multi-institutional datasets, statistical stability analysis through repeated runs and cross-validation, enhanced transformer-based attention mechanisms, and real-time clinical decision-support systems for AI-assisted DR screening.

AUTHOR CONTRIBUTIONS

Mysha Samiha Priota contributed to conceptualization, data curation, and drafting; SK Fahema Mahajabin to preprocessing and dataset preparation; Maharun Fahamida Maysha to clinical validation; Fazley Rabbi to model implementation and training; Md. Imran Hossen to experiments; Md. Shifatul Ahsan Apurba to formal, ROC, and precision–recall analysis; Md. Arman Hosen Patoary to visualization; and Md. Yasin Arafat Tamim to methodology, fusion design, external validation, and comparison. Tauhidur Rahman Sakib conceived, designed, supervised, coordinated, administered, and revised the study as corresponding author. All authors approved the final manuscript.

REFERENCES

[1] G. Mushtaq and F. Siddiqui, "Detection of diabetic retinopathy using deep learning methodology," *IOP Conference Series Materials Science and Engineering*, vol. 1070, no. 1, p. 012049, Feb. 2021, doi: 10.1088/1757-899x/1070/1/012049.

[2] A. Jain, A. Jalui, J. Jasani, Y. Lahoti, and R. Karani, "Deep learning for detection and severity classification of diabetic retinopathy," in *Proc. 1st Int. Conf. Innovations in Information and Communication Technology (ICIICT)*, Chennai, India, 2019, pp. 1–6, doi: 10.1109/ICIICT1.2019.8741456.

[3] Q. H. Nguyen, R. Muthuraman, L. Singh, G. Sen, A. C. Tran, B. P. Nguyen, and M. Chua, "Diabetic retinopathy detection using deep learning," in *Proc. 4th Int. Conf. Machine Learning and Soft Computing (ICMLSC)*, Haiphong City, Vietnam, 2020, pp. 103–107, doi: 10.1145/3380688.3380709.

[4] R. Gargeya and T. Leng, "Automated identification of diabetic retinopathy using deep learning," *Ophthalmology*, vol. 124, no. 7, pp. 962–969, Mar. 2017, doi: 10.1016/j.ophtha.2017.02.008.

[5] V. Gulshan *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, p. 2402, Nov. 2016, doi: 10.1001/jama.2016.17216.

[6] Y. Kale and S. Sharma, "Detection of five severity levels of diabetic retinopathy using ensemble deep learning model," *Multimedia Tools and Applications*, vol. 82, no. 12, pp. 19005–19020, Dec. 2022, doi: 10.1007/s11042-022-14277-x.

[7] Q. Abbas, I. Fondon, A. Sarmiento, S. Jiménez, and P. Alemany, "Automatic recognition of severity level for diagnosis of diabetic retinopathy using deep visual features," *Medical & Biological Engineering & Computing*, vol. 55, no. 11, pp. 1959–1974, Mar. 2017, doi: 10.1007/s11517-017-1638-6.

[8] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, "DREAM: Diabetic Retinopathy Analysis Using Machine Learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 5, pp. 1717–1728, Jan. 2014, doi: 10.1109/jbhi.2013.2294635.

[9] L. Dai *et al.*, "A deep learning system for detecting diabetic retinopathy across the disease spectrum," *Nature Communications*, vol. 12, no. 1, p. 3242, May 2021, doi: 10.1038/s41467-021-23458-5.

[10] S. Dutta, B. C. Manideep, S. M. Basha, R. D. Caytiles, and N. Ch. S. N. Iyengar, "Classification of diabetic retinopathy images by using deep learning models," *International Journal of Grid and Distributed Computing*, vol. 11, no. 1, pp. 99–106, Jan. 2018, doi: 10.14257/ijgdc.2018.11.1.09.

[11] M. D. Abràmoff *et al.*, "Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning," *Investigative Ophthalmology & Visual Science*, vol. 57, no. 13, p. 5200, Oct. 2016, doi: 10.1167/iovs.16-19964.

[12] J. Wang, Y. Bai, and B. Xia, "Simultaneous diagnosis of severity and features of diabetic retinopathy in fundus photography using deep learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 12, pp. 3397–3407, Jul. 2020, doi: 10.1109/jbhi.2020.3012547.

[13] A. Jabbar *et al.*, "A Lesion-Based Diabetic Retinopathy detection through hybrid deep learning model," *IEEE Access*, vol. 12, pp. 40019–40036, Jan. 2024, doi: 10.1109/access.2024.3373467.

[14] F. Li, Z. Liu, H. Chen, M. Jiang, X. Zhang, and Z. Wu, "Automatic detection of diabetic retinopathy in retinal fundus photographs based on deep learning algorithm," *Translational Vision Science & Technology*, vol. 8, no. 6, p. 4, Nov. 2019, doi: 10.1167/tvst.8.6.4.

[15] K. Bahmane, H. Aksasse, and B. A. Chaouki, "A case study of DenseNet, ResNet and Vision Transformers for thyroid nodule analysis in medical imaging," *Int. J. Innovative Science and Research Technology (IJISRT)*, vol. 10, no. 3, pp. 3197–3205, 2025, doi: 10.38124/ijisrt/25mar1818.

[16] T.-C. Phan, C.-H. Ho, and A.-C. Phan, "Detection and classification of liver lesions using vision transformer and active learning," in *Communications in Computer and Information Science*, 2024, pp. 208–224, doi: 10.1007/978-981-96-0434-0_15.

[17] M. Ramzan, A. Bustaman, D. Sarwinda, and N. Ullah, "DenseNet-121 with dual attention for ARMD classification from fundus images," in *Proc. 7th Int. Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, Yogyakarta, Indonesia, 2024, pp. 1089–1094, doi: 10.1109/ISRITI64779.2024.10963655.