Diabetic retinopathy detection using SWIN transformer

Sheetal J. Nagar[1], Nikhil Gondaliya[2]
[1]Department of Computer/IT Engineering, Gujarat Technological University, Ahmedabad, India
[2]Department of Information Technology, G H Patel College of Engineering and Technology,
Charutar Vidya Mandal University, Vallabh Vidyanagar, India

Article Info

Article history:

Received Aug 23, 2025
Revised Apr 4, 2026
Accepted Apr 15, 2026

Keywords:

CLAHE
Diabetic retinopathy
False negative reduction
Medical image classification
SWIN transformer

ABSTRACT

Diabetic retinopathy (DR) is a diabetes related eye disorder that damages the retina. DR is among the most specific complications of diabetes. A vital challenge for automated detection systems in medical image diagnosis is to minimize the false negative rate for patients' timely treatment. This paper presents a novel strategy employing the shifted window (SWIN) Transformer for efficiently modeling local and global visual information to address this challenge. We have proposed our work to maximize the true positive ratio and minimize the false negative ratio for the automated process of diagnosing the level of DR, so that patients with positive signs of DR can be predicted most accurately and can save vision. The results suggest that SWIN Transformer architecture, along with the contrast-limited adaptive histogram equalization (CLAHE) technique, provides a robust option for developing a reliable DR detection system. The results indicate that the proposed approach achieves 96% weighted recall across all the levels of DR detection and 97.45% validation accuracy for the eyePACS DR detection dataset, as well as 99% weighted recall across all the levels of DR detection, along with 99.26% validation accuracy for APTOS 2019 Blindness Detection dataset. Thus, this study aimed to develop a DR detection system focused on minimizing false negatives using the SWIN transformer.

Corresponding Author:

Nikhil Gondaliya
Department of Information Technology, G H Patel College of Engineering and Technology
Charutar Vidya Mandal University, Vallabh Vidyanagar, India
Email: nikhilgondaliya@gcet.ac.in

1. INTRODUCTION

Diabetes is a worldwide disease and diabetic retinopathy (DR), the most common microvascular consequence of diabetes, is becoming an increasing cause of blindness and visual impairment. DR is considered as the most complicated disorder of diabetes and has been used to diagnose diabetes [1]. As per WHO, 642 million people are predicted to have diabetes by 2040, with 224 million (35%) developing DR and 70 million (11%) developing sight-threatening retinopathy. Immediate and significant action is required to control diabetes and with it, diabetic retinopathy. A well-organized approach is required for timely detection and treatment to save patients' vision [2]. Manual identification by ophthalmologists is more time consuming and causes significant discomfort during testing. By leveraging the techniques of deep learning, automated systems can be developed for analysis and classification of severity level for medical images, which helps in early detection of DR. Using a combination of data-efficient image enhancement, SWIN transformer and fine-tuning, our research offers classification of various DR levels. The international classification of diabetic retinopathy (ICDR) is the most commonly used classification of DR [3]. It classifies DR into 5 categories, such as No Apparent DR, Mild non proliferative DR (NPDR), Moderate NPDR, Severe NPDR and proliferative diabetic retinopathy (PDR).

Figure 1 illustrates a retinal image showing both a normal retina [4] and a retina affected by DR [4]. The latter displays abnormal features, such as hemorrhages, cotton wool spots and hard exudates. Additionally, the diseased retina may exhibit irregular blood vessel growth.

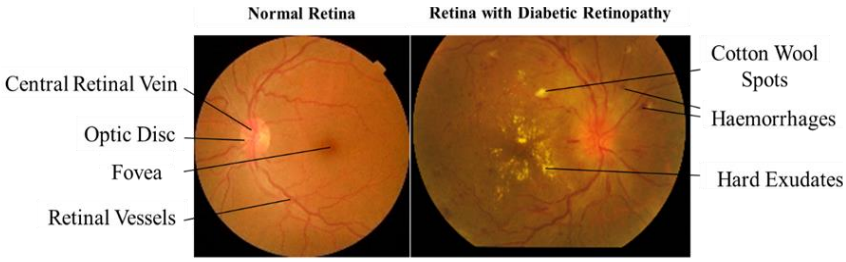

Figure 1. Normal retina and retina with DR [3]

2. RELATED WORKS

Das *et al.* [5] have implemented TesNet50, InceptionV3, EfficientNetB4 and DenseNet201. Thus, they made experiments on different transfer learning approaches and the best result was observed by EfficientNetB4 with 79.11% validation accuracy. Luo *et al.* [6] performed experiments such that could take advantage of relationships among features of local patches of the images and applied long range patches into CNN network. Their experiments exhibit 83.6% accuracy, 86.5% sensitivity, 69.3% specificity, 81.9% precision and 82.6% F1-score. Zhang *et al.* [7] developed a deep learning algorithm of a graph convolutional network to identify relationships of retinal features, which were learnt by layers of CNN. They applied this approach on two datasets which resulted in 89.9% accuracy, 88.2% sensitivity and 91.3% specificity for eyePACS dataset and 91.8% accuracy, 90.2% sensitivity and 93.0% specificity for Messidor dataset. Mutawa *et al.* [8] used datasets from three countries, namely Asia Pacific Tele Ophthalmology Society from India, Eye Picture Archive Communication System from the United States and Ocular Disease Intelligent Recognition from China. They applied a geometric transformation on the dataset. Their experiments showed maximum accuracy and recall as 89.10% for EyePACS with DenseNet121 model and maximum accuracy and recall as 98.50% for APTOS with DenseNet121 and MobileNetV2 models. Sait [9] employed the noise removal technique, YOLO v7 technique to extract features and QMPA technique for feature selection. His suggested model resulted in 98.0% accuracy and 93.7% F1-score for APTOS dataset and 98.4% accuracy and 93.1% F1-score for eyePACS dataset. Suedumrong *et al.* [10] removed unnecessary background from the EyePACS dataset and implemented a smaller and flexible CNN architecture to enhance result of the performance and they concluded with 90.60% of accuracy. Touati *et al.* [11] introduced the Compact Convolutional Transformer model, utilizing convolutional and transformer techniques and accomplished 95% accuracy. Saranya *et al.* [12] utilized the Support Vector Machine classifier for multiclass classification and identified DR level through the presence of red lesions, such as microaneurysms and hemorrhages. They obtained significantly improved accuracy of 94.5% on APTOS dataset and 93.3% accuracy on IDRiD dataset, but low recall compared with accuracy as 75.6% on APTOS dataset and 78.5% on IDRiD dataset. Butt *et al.* [13] performed experiments for both binary classification by combining stage 1 to stage 4 in one class as DR and stage 0 as NDR as well as multiclass classification using APTOS dataset. They applied a hybrid approach using GoogleNet, ResNet-18 and a classifier as support vector machine. By this approach, they exhibited 97.80% binary classification accuracy and 89.29% multiclass classification accuracy. Menaouer *et al.* [14] also applied a hybrid technique of Deep CNN and two VGG models, namely VGG16 and VGG19 and yielded 90.60% accuracy, 95% recall and 94% F1-score for APTOS2019, Messidor-2 and Local public DR image datasets.

Nahiduzzaman *et al.* [15] employed a parallel convolutional neural network to reduce time for feature extraction and extreme learning machine techniques, which showed 91.78% accuracy for eyePACS dataset as well as 97.27% accuracy, 95% recall and 95% precision for APTOS 2019 dataset. Alwakid *et al.* [16] performed image preprocessing using contrast limited adaptive histogram equalization and an Enhanced Super resolution Generative Adversarial Network to feed these images to CNN to improve the prediction accuracy of the model. They represented 97.83% accuracy and 98% recall using their proposed model. Xu *et al.* [17] emphasized on integration of both local and global features for accurate lesion characterization and utilized the strength of EfficientNet models and Swin Transformer. Additionally, to optimize the hybrid model's performance they applied Gaussian blur method and received 95% recall, 98% specificity, 97% accuracy and 97% AUC as their outcome.

3. EXISTING DATASETS AND DR DETECTION METHODS

In this section, two DR image datasets are discussed along with the automated DR detection methods.

3.1. eyePACS dataset from Kaggle

The resized and cropped DR detection competition dataset, a publicly available dataset, has been obtained from Kaggle [18]. The dataset provides a vast number of images, making Deep Learning algorithms learn well. This implementation utilizes the resized version of the Kaggle competition dataset, comprising a total of 35,126 images categorized into five distinct levels. For experiments randomly 70% of these images are taken for training and remaining 30% images are used for testing.

3.2. APTOS dataset from Kaggle

APTOS Eye Preprocessing in DR dataset is downloaded from Kaggle, which was created for APTOS 2019 Blindness Detection competition [4]. This dataset contains DR images with five different diagnostic values ranging from 0 to 4, indicating the severity level of DR. Images are accessible from two separate folders for training and testing. 3662 images for training and 1928 images are available for testing. The images were gathered from multiple clinics using a variety of cameras over an extended period of time, which introduces further variations.

3.3. DR detection methods

Figure 2 shows various methods used for automated detection of DR. Supervised learning plays a vital role in DR detection, with the deep learning methods (CNNs, RNNs, Transformers) being the leading techniques. Models such as VGG16, GoogleNet, ResNet, and EfficientNetB3 have shown great potential in accurately classifying retinal images for DR. Meanwhile, transfer learning has significantly emerged for training, making it easier to deploy these models in real-world clinical settings. By leveraging these advanced techniques, the accuracy of detecting the severity of DR can be significantly improved for timely treatment of the disease.

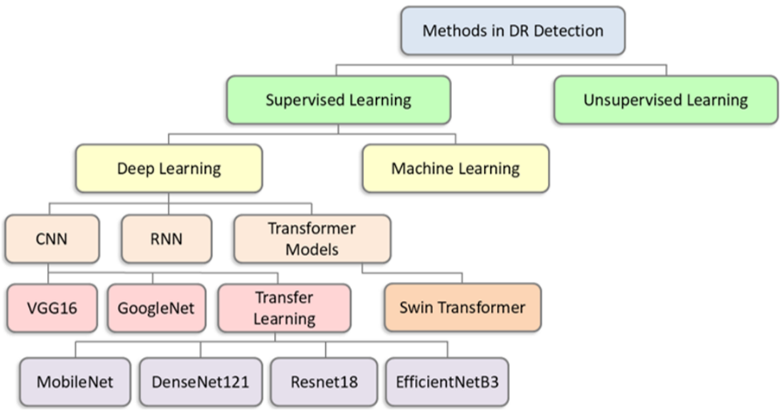

Figure 2. DR detection methods

4. PROPOSED METHODOLOGY

4.1. Architecture of SWIN transformer

The structure of SWIN Transformer is illustrated in Figure 3. The SWIN Transformer is a hierarchical transformer model useful for image processing. It is structured to improve computational efficiency and performance of high-resolution images. It combines several key components that enable local attention, hierarchical representation and effective global context modeling. For patch partition, the input image is first divided into non-overlapping patches. For instance, an image of size H × W × 3 is partitioned into patches of size 4 × 4, creating a sequence of flattened patch vectors. Each patch acts as a visual token. This partitioning reduces the spatial size and increases processing efficiency [1], [19].

Each flattened patch is passed through a linear embedding layer that maps it into a feature vector. This step converts the 2D image content into a format suitable for transformer processing. Then the images can be processed by the attention layer of the SWIN transformer block, consisting of two sub-units, a normalization layer followed by an attention module. Attention is computed on each window made up of patches. Attention is computed within nonoverlapping windows and then by shifting the windows.

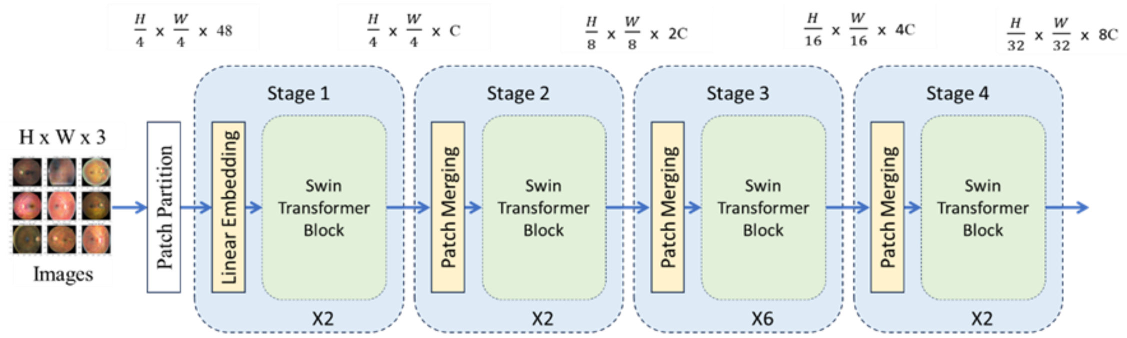

Figure 3. Architecture of SWIN transformer

In the next layer, the window partitions are shifted (e.g., by half the window size) to enable information exchange across window boundaries. It efficiently extracts features by performing self-attention within local windows and then introduces connections between these windows through a shifted window approach. Swin Transformer uses Patch Merging at the end of each stage. As the network progresses to deeper layers, the patch merging layer reduces the number of patch tokens. In each stage, it halves the spatial dimensions of the patch tokens while doubling the projected feature dimension [20]. It concatenates the features of each group of 2×2 neighboring patches, reducing the number of tokens by a factor of 4 while increasing the feature dimension. This process allows the model to progressively learn larger spatial patterns and contextual information across stages. The architecture consists of multiple stages. At each stage, patch merging operation reduces the image size and increases the feature dimensionality. As a result, the SWIN transformer architecture builds a multi-scale hierarchical representation which makes it suitable for image classification.

4.2. Workflow of proposed method using swin transformer

The proposed system aims to reduce false negatives in DR detection by utilizing the Swin Transformer architecture. This model has a hierarchical structure and a shifted window mechanism, which helps to capture both local and global features in images. The methodology includes several key stages as given in Figure 4.

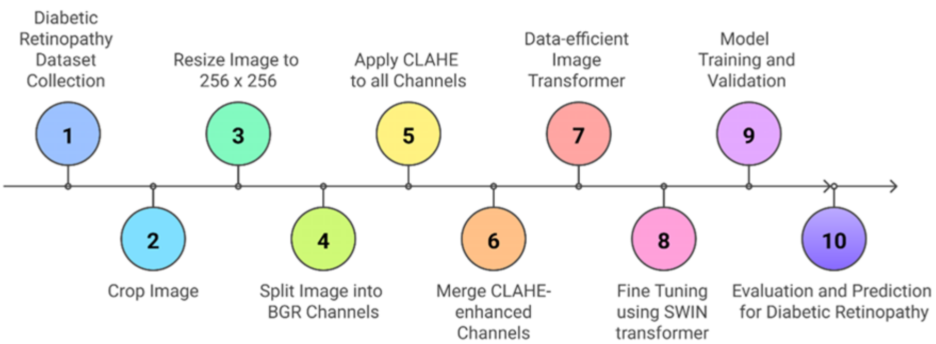

Figure 4. Workflow of proposed method

In the proposed work, we have utilized APTOS 2019 and eyePACS DR datasets for the implementation of a DR detection model. These datasets provide labeled high-resolution fundus photographs representing different stages of DR: no DR, mild, moderate, severe, and proliferative.

Resizing and preprocessing are advantageous for emphasizing the symptoms in DR images. So, images are cropped to isolate the region requiring attention, eliminating the black border and unnecessary background. The cropped region is resized to a uniform resolution of 256 × 256 pixels, to input to the neural network. The resized image is split into its Blue, Red and Green (BRG) channels for contrast enhancement. To expand visibility of critical retinal abnormalities such as microaneurysms and exudates, each channel is independently processed using contrast-limited adaptive histogram equalization (CLAHE). This technique enhances local contrast while limiting noise amplification. The processed channels are then merged to form a single, high contrast image that highlights subtle features relevant to DR grading [21], [22]. For the proposed work, implementation is carried out on two datasets: 1) APTOS: Eye Preprocessing in DR dataset, and 2) eyePACs: DR detection competition dataset. Sample images from both datasets, along with preprocessed images, are represented in Figures 5-8.

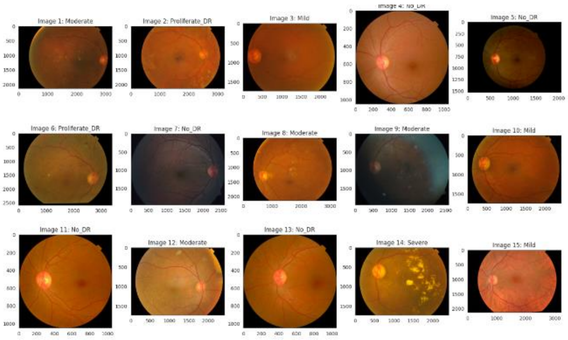

Figure 5. Sample Images of APTOS Dataset

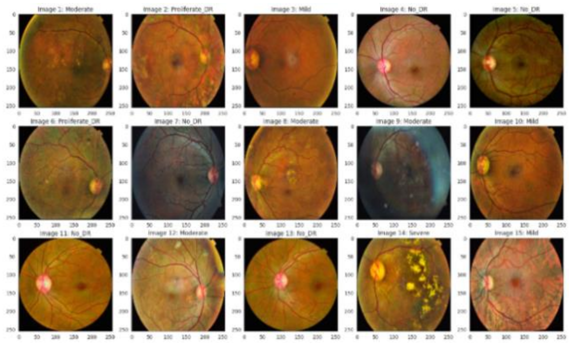

Figure 6. Preprocessed Images of APTOS Dataset

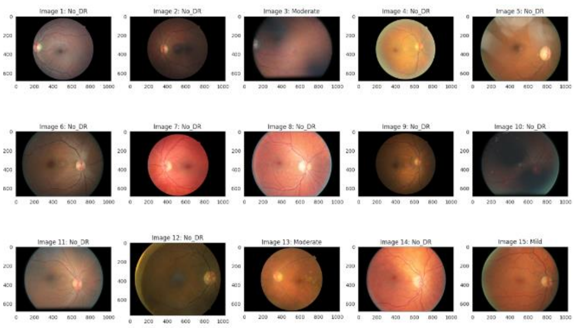

Figure 7. Sample images EyePACS DR dataset

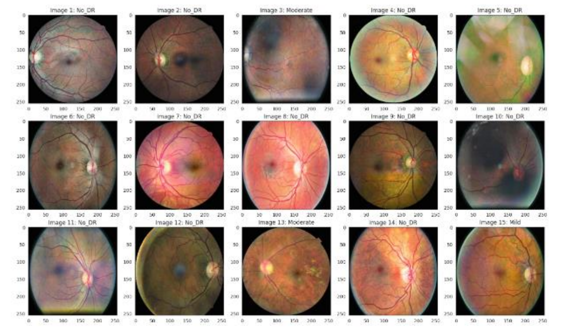

Figure 8. Preprocessed images of EyePACS DR dataset

Vision transformers outperform CNNs by using self-attention technique to extract long-range correlations from raw data [23]. For improved performance and the finest learning, SWIN Transformer is fine-tuned using the preprocessed fundus images. The SWIN Transformer uses a hierarchical structure with shifted window-based self-attention, allowing the model to efficiently learn both local and global patterns. This approach is especially beneficial for retinal images, where detecting small lesions requires both detail sensitivity and spatial awareness [1]. The dataset is randomly divided into training and validation sets, typically using a 70:30 split. Model performance is assessed using evaluation metrics such as accuracy, precision, recall and F1-score. These metrics collectively provide insights into the model's capability across all five DR severity levels. The fine-tuned Swin Transformer achieves robust performance, making it a promising approach for automated DR screening and severity classification in clinical use.

The identification of DR from fundus images requires effective image preprocessing. It also depends on an architectural framework that can extract both fine-grained lesion characteristics and broader retinal patterns. The CLAHE method enhances the visibility of micro-aneurysms, exudates and haemorrhages. It performs localized histogram equalization across image tiles while preventing excessive noise amplification [16], [24]. This preprocessing step enhances the visibility of DR features and enables the model to generalize more effectively. SWIN Transformer enables modelling of fine-grained lesion texture and long-range contextual relationships in retinal images. Its shifted-window self-attention mechanism supports computationally efficient processing of high-resolution images and captures multi-scale anatomical features [24], [25]. The integration of CLAHE preprocessing and the SWIN Transformer provides a robust pipeline. CLAHE refines image input quality and the SWIN Transformer extracts salient hierarchical features for classification. This hybrid pipeline increases classification accuracy and enhances the model’s ability to detect abnormalities, thereby making automated DR screening systems more useful and reliable in clinical practice.

4.3. Implementation and performance metrics

Experiments are performed using PyTorch, a deep learning framework developed by Facebook's AI Research lab [26] and the sklearn libraries of Python. PyTorch can utilize CUDA to perform computations on NVIDIA GPUs, which can significantly accelerate training and inference times for deep learning models. Our focus for the implementation is to maximize recall value, thereby minimizing false negative values while controlling the overall accuracy of the proposed work.

Accuracy: "Ratio of the number of correct classifications to the total number of classifications" [27]. It defines how precisely the model is able to estimate the output for the DR classification from 5 levels, i.e. from level 0 to level 4. Accuracy = Correctly identified samples / Total no. of examined samples

Recall: "Ratio of the number of correctly identified positive samples to the total number of positive samples". For an imbalanced dataset, just accuracy is not preferable. Moreover, false negative ratio must be minimized in the health domain. The higher the recall, the greater the number of positive samples detected, thereby minimizing false negative value. Whenever a false negative is important, the result of recall is preferable. Recall = Correctly identified positive samples / Total no. of actual positive samples.

5. RESULT ANALYSIS

Result of the proposed implementation is observed for the eyePACS and APTOS 2019 datasets. These results are compared with the results of the similar datasets used by other researchers within last three years. Comparison of the results exhibited by various researchers representing the results of accuracy and recall for eyePACS dataset and APTOS-2019 Blindness Detection dataset is shown in Tables 1 and 2 respectively and comparative visual analysis is represented through Figures 9 and 10 respectively.

5.1. Analysis of proposed work on eyePACS dataset

Table 1 presents a detailed comparison of the proposed model's performance against various established approaches on the eyePACS dataset, highlighting both accuracy and recall metrics to demonstrate the model's diagnostic capabilities. The results provide comparison with previous studies along with representing the proposed model's position in current research.

The comparative analysis is further enhanced through visual representation in Figure 9. The bar chart in this figure illustrates the performance evaluation of various DR detection models on the eyePACS dataset, presenting both accuracy and recall metrics across eight different research approaches spanning from 2022 to 2025. The comparison indicates substantial variation in model performance across various studies, with accuracy spanning from 79.11% to 99.40% and recall ranging between 59% and 96.92%. The proposed work exhibits 96.92% for both accuracy and recall, indicating balanced performance across these evaluation metrics. This analysis provides the outcome of different approaches when tested on the EyePACS dataset.

Table 1. Comparisons of accuracy and recall for DR detection using eyePACS dataset

Author	Year	Method	Accuracy	Recall
[5]	2022	EfficientNetB4	79.11%	79.11%
[6]	2022	Deep convolutional neural network by mining local and long-range dependence	83.60%	86.50%
[7]	2022	Diabetic Retinopathy Grading by Deep Graph Correlation Network	89.90%	88.20%
[8]	2023	DenseNet121	89.10%	89.10%
[9]	2023	A hyperparameter-optimized MobileNet V3 model	98.40%	92.60%
[10]	2024	CNN with image preprocessing and data augmentation	90.60%	59.00%
[11]	2025	Convolutional Layers with transformer techniques	95.00%	93.23%
Proposed work	2025	CLAHE for image contrast enhancement of low-contrast DR images with and fine-tuning using SWIN Transformer	96.92%	96.92%

5.2. Analysis of proposed work on APTOS-2019 blindness detection dataset

Table 2 presents a comparative analysis of different methodologies for DR detection using the APTOS-2019 Blindness Detection dataset. The current study achieves 99% recall and 98.72% accuracy. High recall is particularly important in clinical DR screening, as it reflects a very low false negative rate. Minimizing false negatives is critical because undetected cases can result in delayed treatment and potential vision loss for patients. The combination of strong recall and accuracy metrics suggests the model's potential suitability for clinical deployment. Figure 10 displays a visual comparison of accuracy and recall metrics across these methods. The comparison includes various research approaches presented during 2022 and 2025.

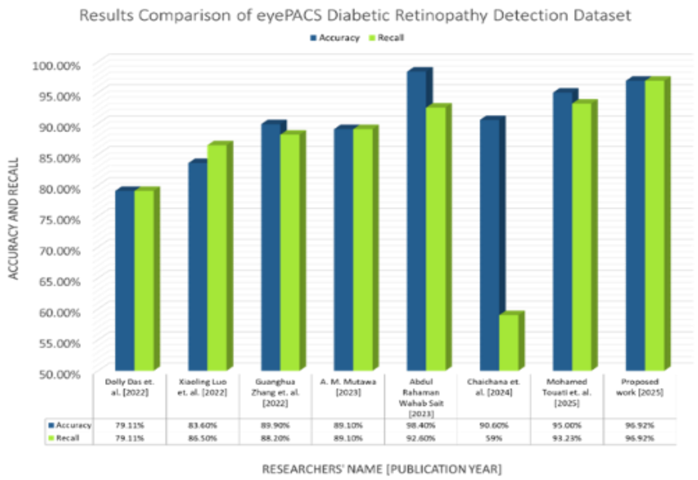

Figure 9. Results comparison of eyePACS DR detection dataset

Table 2. Comparisons of accuracy and recall for DR detection using APTOS-2019 blindness detection dataset

Author	Year	Method	Accuracy	Recall
[12]	2022	multiclass support vector machine classifier	94.50%	75.60%
[13]	2022	GoogleNet Model with SVM classifier	89.29%	80.00%
[14]	2022	hybrid deep learning approach using deep convolutional neural network (CNN) method and two VGG network models (VGG16 and VGG19)	90.60%	95.00%
[9]	2023	A hyperparameter-optimized MobileNet V3 model	98.00%	93.90%
[15]	2023	parallel convolutional neural network-based feature extractor and ELM classifier	97.27%	95.00%
[16]	2023	Deep Learning, CLAHE and ESRGAN	97.83%	98.00%
[17]	2024	The hybrid neural network model based on EfficientNet and Swin Transformer	97.00%	95.00%
Proposed work	2025	CLAHE for image contrast enhancement of low-contrast DR images with and fine-tuning using SWIN Transformer	98.72%	99.00%

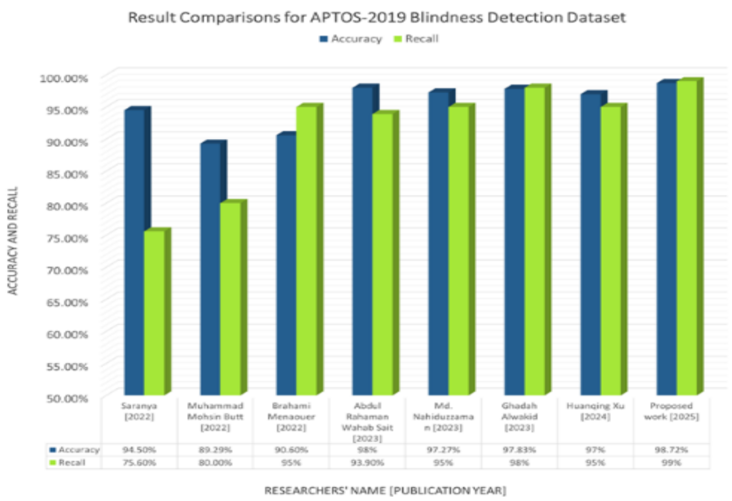

Figure 10. Results comparison for APTOS-2019 blindness detection dataset

5.3. Dataset factors affecting model performance

The APTOS images are generally cleaner, more centered and better illuminated, while EyePACS images vary significantly in size, color tone and illumination. High-quality and well-focused images help the model detect subtle retinal lesions more effectively, resulting in improved sensitivity. EyePACS is a highly imbalanced dataset providing a larger number of small-sized images with variable dimension like 1024 x 913, 1024 x 786 and 1024 x 684. APTOS offers fewer but more detailed, large-sized and better-preprocessed images with uniform dimensions 640 x 480. This helps to improve model performance and achieve higher recall in experiments. Such consistency in preprocessing reduces intra-class variability and minimizes false negatives.

6. FUTURE WORK

Future work can focus on enhancing the ability of the model to perform well on new and unseen data. Clinical readiness of the proposed work can be examined to ensure that it can be deployed safely and effectively in medical practice. The proposed model can be improved through testing on multiple clinical datasets to ensure its reliability across different imaging conditions. A systematic evaluation of CLAHE parameters and alternative preprocessing methods may optimize lesion contrast and help to improve the performance of the model. Additionally, implementing explainable AI techniques like Grad-CAM and attention map visualization, can highlight the regions of the retinal images used by the model for the classification. This interpretability can help to make the model's decision-making process more transparent.

7. CONCLUSION

This work presents an effective solution for DR detection using a combination of CLAHE-based preprocessing for efficient contrast enhancement in the images and fine-tuning with the SWIN Transformer. The model consistently achieved high validation accuracy and recall across two standard datasets, which can be considered as a powerful tool for DR detection. The reduced false negative rate makes this approach well-suited for medical screening programs. Identifying the positive DR cases early and accurately is essential to prevent patients from losing their vision. With the recall rate of 96.92% for eyePACS and 99% for APTOS dataset, the method performs especially well for identifying positive cases. This is crucial because missing a positive case can have serious consequences for patients' health. The model maintains performance across both accuracy and recall metrics. This indicates that the proposed work can perform reliably in the real-world clinical environments. This research focuses on the DR classification. However, the proposed methodology has broader applicability. The framework can be extended to other diagnostic challenges in medical imaging, where similar analytical approaches are needed.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Sheetal J. Nagar	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓
Nikhil Gondaliya	✓	✓			✓	✓		✓		✓	✓	✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] Z. Liu *et al.*, "Swin transformer: hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE International Conference on Computer Vision*, IEEE, Oct. 2021, pp. 9992–10002. doi: 10.1109/ICCV48922.2021.00986.

[2] R. Pradeepa and V. Mohan, "Epidemiology of type 2 diabetes in India," *Indian Journal of Ophthalmology*, vol. 69, no. 11, pp. 2932–2938, Nov. 2021, doi: 10.4103/ijo.IJO_1627_21.

[3] "Strengthening diagnosis and treatment of Diabetic Retinopathy in SEA Region," World Health Organization South-East Asia. Accessed: Jul. 27, 2025. [Online]. Available: https://www.who.int/publications/i/item/9789290227946

[4] "APTOS : Eye Preprocessing in Diabetic Retinopathy." Accessed: Mar. 31, 2025. [Online]. Available: https://www.kaggle.com/code/ratthachat/aptos-eye-preprocessing-in-diabetic-retinopathy/data

[5] D. Das, S. K. Biswas, and S. Bandyopadhyay, "Detection of diabetic retinopathy using convolutional neural networks for feature extraction and classification (DRFEC)," *Multimedia Tools and Applications*, vol. 82, no. 19, pp. 29943–30001, Aug. 2023, doi: 10.1007/s11042-022-14165-4.

[6] X. Luo *et al.*, "A deep convolutional neural network for diabetic retinopathy detection via mining local and long-range dependence," *CAAI Transactions on Intelligence Technology*, vol. 9, no. 1, pp. 153–166, Feb. 2024, doi: 10.1049/cit2.12155.
[7] G. Zhang *et al.*, "Diabetic retinopathy grading by deep graph correlation network on retinal images without manual annotations," *Frontiers in Medicine*, vol. 9, Apr. 2022, doi: 10.3389/fmed.2022.872214.
[8] A. M. Mutawa, S. Alnajdi, and S. Sruthi, "Transfer learning for diabetic retinopathy detection: a study of dataset combination and model performance," *Applied Sciences*, vol. 13, no. 9, p. 5685, May 2023, doi: 10.3390/app13095685.
[9] A. R. Wahab Sait, "A lightweight diabetic retinopathy detection model using a deep-learning technique," *Diagnostics*, vol. 13, no. 19, p. 3120, Oct. 2023, doi: 10.3390/diagnostics13193120.
[10] C. Suedumrong, S. Phongmoo, T. Akarajaka, and K. Leksakul, "Diabetic retinopathy detection using convolutional neural networks with background removal, and data augmentation," *Applied Sciences*, vol. 14, no. 19, p. 8823, Sep. 2024, doi: 10.3390/app14198823.
[11] M. Touati, R. Touati, L. Nana, F. Benzarti, and S. Ben Yahia, "DRCCT: enhancing diabetic retinopathy classification with a compact convolutional transformer," *Big Data and Cognitive Computing*, vol. 9, no. 1, p. 9, Jan. 2025, doi: 10.3390/bdcc9010009.
[12] P. Saranya, K. M. Umamaheswari, S. C. Patnaik, and J. S. Patyal, "Red lesion detection in color fundus images for diabetic retinopathy detection," in *Proceedings of International Conference on Deep Learning, Computing and Intelligence*, 2022, pp. 561–569. doi: 10.1007/978-981-16-5652-1_50.
[13] M. M. Butt, D. N. F. A. Iskandar, S. E. Abdelhamid, G. Latif, and R. Alghazo, "Diabetic retinopathy detection from fundus images of the eye using hybrid deep learning features," *Diagnostics*, vol. 12, no. 7, p. 1607, Jul. 2022, doi: 10.3390/diagnostics12071607.
[14] B. Menaouer, Z. Dermane, N. El Houda Kebir, and N. Matta, "Diabetic Retinopathy Classification Using Hybrid Deep Learning Approach," *SN Computer Science*, vol. 3, no. 5, p. 357, Jul. 2022, doi: 10.1007/s42979-022-01240-8.
[15] M. Nahiduzzaman *et al.*, "Diabetic retinopathy identification using parallel convolutional neural network based feature extractor and ELM classifier," *Expert Systems with Applications*, vol. 217, p. 119557, May 2023, doi: 10.1016/j.eswa.2023.119557.
[16] G. Alwakid, W. Gouda, M. Humayun, and N. Z. Jhanjhi, "Deep learning-enhanced diabetic retinopathy image classification," *Digital Health*, vol. 9, Jan. 2023, doi: 10.1177/20552076231194942.
[17] H. Xu, X. Shao, D. Fang, and F. Huang, "A hybrid neural network approach for classifying diabetic retinopathy subtypes," *Frontiers in Medicine*, vol. 10, Jan. 2023, doi: 10.3389/fmed.2023.1293019.
[18] "Diabetic retinopathy detection competition dataset resized/cropped." Accessed: Mar. 31, 2026. [Online]. Available: https://www.kaggle.com/tanlikesmath/diabetic-retinopathy-resized
[19] Z. Liu *et al.*, "Swin transformer V2: scaling up capacity and resolution," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2022, pp. 11999–12009. doi: 10.1109/CVPR52688.2022.01170.
[20] G. J. Ferdous, K. A. Sathi, M. A. Hossain, and M. Ali Akber Dewan, "SPT-Swin: a shifted patch tokenization swin transformer for image classification," *IEEE Access*, vol. 12, pp. 117617–117626, 2024, doi: 10.1109/ACCESS.2024.3448304.
[21] M. Hayati *et al.*, "Impact of CLAHE-based image enhancement for diabetic retinopathy classification through deep learning," *Procedia Computer Science*, vol. 216, pp. 57–66, 2022, doi: 10.1016/j.procs.2022.12.111.
[22] O. Sule and S. Viriri, "Contrast enhancement of RGB retinal fundus images for improved segmentation of blood vessels using convolutional neural networks," *Journal of Digital Imaging*, vol. 36, no. 2, pp. 414-432, 2023, doi: 10.1007/s10278-022-00738-0.
[23] I. Pacal, "A novel Swin transformer approach utilizing residual multi-layer perceptron for diagnosing brain tumors in MRI images," *International Journal of Machine Learning and Cybernetics*, vol. 15, no. 9, pp. 3579-3597, 2024, doi: 10.1007/s13042-024-02110-w.
[24] S. Zhang, C. A. B. Webers, and T. T. J. M. Berendschot, "Computational single fundus image restoration techniques: a review," *Frontiers in Ophthalmology*, vol. 4, Jun. 2024, doi: 10.3389/fopht.2024.1332197.
[25] Z. Yao *et al.*, "FunSwin: A deep learning method to analysis diabetic retinopathy grade and macular edema risk based on fundus images," *Frontiers in Physiology*, vol. 13, Jul. 2022, doi: 10.3389/fphys.2022.961386
[26] A. Paszke *et al.*, "PyTorch: an imperative style, high-performance deep learning library," in *Advances in neural information processing systems 32*, 2019. [Online]. Available: https://pytorch.org.
[27] S. J. Nagar and N. Gondaliya, "Identification of severity level for diabetic retinopathy detection using neural networks," in *International Conference on Data Science and Applications*, 2023, pp. 205–220. doi: 10.1007/978-981-99-7862-5_16.Z.

BIOGRAPHIES OF AUTHORS

Sheetal J. Nagar serves as an Assistant Professor in the Computer Engineering Department at Government Engineering College (GEC), Rajkot, Gujarat, India. She obtained Bachelor of Engineering in Computer Engineering and Master of Engineering (ME) in Computer Engineering. She has 15 years of experience in the academic field. Her areas of interest are deep learning, medical image processing, data visualization, computer algorithms and compiler design. She has published 6 papers in national and international journals and conferences. She is one of the authors of Lecture Notes in Networks and Systems, Volume 818, published by Springer Nature Singapore. She can be contacted at email: nagar.sheetal@gmail.com.

Nikhil Gondaliya serves as the Professor and Head of the Information Technology Department at G H Patel College of Engineering & Technology (GCET), Vallabh Vidyanagar, Gujarat, India. He obtained Doctor of Philosophy in Computer Engineering from Gujarat Technological University and Master Degree in Computer Engineering from Sardar Patel University. He has 20+ years of experience in the academic field. He has published more than 20 papers in various National/International Conferences and international conferences and journals. He has guided more than 40 projects at Under Graduate level and more than 12 Dissertation at Post Graduate level. His areas of interest are Adhoc wireless networks, data mining and IoT, machine learning, and deep learning. He can be contacted at email: nikhilgondaliya@gcet.ac.in.