AUT Journal of Electrical Engineering
10.22060/EEJ.2024.23500.5621

Diabetic Retinopathy Detection from Retinal Images Using the Pyramid Vision Transformer Method

Samaneh Dehghani[1], Hossein Ebrahimnezhad[2*], Nasrin Rahmani[3]

[1]Master of Science, Faculty of Electrical Engineering, Sahand University of Technology, Tabriz, Iran
[2]Full Professor, Faculty of Electrical Engineering, Sahand University of Technology, Tabriz, Iran
[3]Master of Science, Faculty of Electrical Engineering, Sahand University of Technology, Tabriz, Iran

Abstract:

The development of automated diagnostic tools is essential for efficiently analyzing medical data, particularly for conditions like Diabetic Retinopathy, a leading cause of vision loss and blindness in adults. The Asia Pacific Tele-Ophthalmology Society 2019 blindness detection dataset, which includes detailed retinal images, is crucial for the advancement of these tools. This study focuses on using the Pyramid Vision Transformer to improve both the accuracy and efficiency of Diabetic Retinopathy detection. Unlike the traditional Vision Transformer, which is computationally expensive and produces low-resolution outputs due to its single-scale structure, Pyramid's multi-scale architecture enables more efficient feature representation. This design allows for better management of large feature maps and enhances image resolution, both vital for precise diagnoses. By implementing the Pyramid Vision Transformer, our approach not only increases accuracy but also improves resource efficiency, outperforming conventional Convolutional Neural Networks. Extensive experiments demonstrate that the model significantly boosts detection and classification accuracy, making it a valuable tool for clinical applications. The model achieved 92.38% accuracy and an Area Under the Curve of 99.58%. These results highlight the model's effectiveness in real-world applications. Future research will focus on optimizing the model for even better performance and exploring its clinical integration to further enhance the diagnostic process in healthcare.

Keywords:

Diabetic Retinopathy, Pyramidal Vision Transformer, Detection, Blindness, Retinal Images

*- Corresponding Author: Prof. Hossein Ebrahimnezhad
Address: Computer Vision Res. Lab., Faculty of Electrical Engineering, Sahand University of Technology, Tabriz, Iran. Email: ebrahimnezhad@sut.ac.ir.

10.22060/EEJ.2024.23500.5621

1. Introduction

Diabetic Retinopathy (DR), a complication of diabetes, damages the blood vessels in the retina and is one of the leading causes of blindness among individuals aged 20 to 74. In the early stages of the disease, there are no symptoms; however, as it progresses, vision problems begin to emerge, eventually leading to complete blindness. High blood sugar levels can damage the light-sensitive tissue at the back of the eye, the retina, resulting in DR. Damaged blood vessels may leak or bleed, causing blurred vision [1].

DR is generally classified into four stages: mild, moderate, severe, and proliferative. In the first stage, known as mild DR, small bulges called microaneurysms appear in parts of the retinal blood vessels. This stage rarely leads to vision loss and does not require treatment, but it indicates diabetic damage and the potential for disease progression. In the second stage, moderate DR, some retinal blood vessels are damaged, causing blood and fluid to leak into the retinal tissue, which can result in vision loss. The third stage, severe DR, occurs when diabetes becomes increasingly uncontrolled, causing more blood vessels to become damaged and blocked, leading to a significant amount of blood and fluid entering the retina, which greatly affects vision. The final stage, proliferative DR, represents an advanced stage of the disease that severely threatens vision. Further damage to the eye's blood vessels disrupts blood flow within the eye, prompting the formation of new, abnormal blood vessels in the retina. These new vessels can lead to more serious complications, potentially causing vision loss and even blindness [2]. Deep Learning (DL) approaches are widely used in image classification across various fields, with predictions made through neural networks. This method assists in the accurate diagnosis of diseases due to its utilization of millions of parameters and mathematical computations [3]. This research aims to enhance the early diagnosis of DR by using the Pyramid Vision Transformer (PVT) to create an automated detection system. This system can help ophthalmologists provide more effective and preventive treatments by precisely analyzing images. It is a reliable tool for diagnosing DR, reducing treatment costs, and alleviating economic burden. Additionally, the system can be used in educational programs to enhance the diagnostic

skills of ophthalmologists and has potential applications beyond the medical field, highlighting its significant value in the scientific and technological community.

We will review some previous works. Murad Canayaz [4] examined various methods for improving the accuracy of DR diagnosis. These methods include DL techniques such as EfficientNet and DenseNet, optimization algorithms, and Support Vector Machines (SVMs). The advantages of these approaches include high accuracy and proven success with a kappa value greater than 0.8. However, they require a GPU-supported graphics card and more memory. The highest accuracy achieved was 96.32%, with a kappa value of 0.98. Purna Chandra Reddy and Kumar Gurrala [5] introduced a Graph Convolutional Network (GCN) based on graph learning and a Relation-Aware Channel-Spatial Attention (RACSA) model, along with a modified Deer Hunting Optimization Algorithm (MDHOA) for feature selection. These methods were applied to the IDRID dataset. Despite minor differences between the DME, DR, and common DR-DME categories, which pose challenges and indicate a need for further improvement in classification performance for common DR-DME, these models provided higher accuracy in classifying common DR-DME and outperformed conventional methods. Specifically, the classification accuracy improved by 5.11%, 3.88%, and 5.47% for DME, DR, and common DR-DME, respectively, and outperformed the leaders in the ISBI-2 sub-challenge. Zulaikha Beevi [6] introduced methods such as SqueezeNet and Deep Convolutional Neural Network (DCNN), along with FrWSO and FrWRO optimization, for classifying the severity of DR in retinal images. These methods, despite requiring ophthalmologists' expertise and DL models' weight adjustment, improved the accuracy, sensitivity, and specificity of DR classification. An accuracy of 91.6%, sensitivity of 92.2%, and specificity of 91.9% were achieved at the first level, while an accuracy of 91.1%, sensitivity of 89.8%, and specificity of 91.3% were achieved at the second level. These methods have the potential for early detection and reducing the risk of blindness. Rakesh Chandra Joshi et al [7] introduced the VisionDeep-AI framework, which includes a weighted bidirectional feature pyramid network, a U-Net backbone architecture, DL-based segmentation and classification, and data augmentation. This method was applied to a

comprehensive dataset of retinal color images. Despite the need for further validation across diverse datasets and imaging conditions and the necessity for integration with additional modalities such as Optical Coherence Tomography (OCT) for a comprehensive description of the disease, and subjective assessments by domain experts for validating practical applicability, this method enabled early detection of eye diseases with robust segmentation and classification. High accuracy in blood vessel segmentation (97.73%) and multi-class classification (81.50%) was achieved, demonstrating the generalizability and efficiency of the VisionDeep-AI framework. This method has the potential for significant improvement in the early detection of various eye diseases, leading to timely and effective treatments. Blood vessel segmentation accuracy of 97.73%, multi-class classification test accuracy of 81.50%, and specificity of 93.83% were reported. Gabriel Tuzato Zago et al [8] introduced a CNN and deep patch-based approach for lesion localization and DR labeling. This model achieved an Area Under the Curve (AUC) of 0.912 across various datasets, outperforming other models. Despite drawbacks like reduced lesion localization accuracy and the inability to provide precise segmentation, this approach offers advantages such as early detection, efficient training patch selection, and reduced processing time. This model can immediately classify images without retraining on different databases. Javed Mehdi Shamrat and colleagues [9] used CNN models, specifically DRNet13, to classify diabetic retinopathy. These models were applied to a dataset of 7,500 fundus images, achieving 97% accuracy in DR detection. DRNet13's advantages include automatic classification of different stages of diabetic retinopathy, improved image preprocessing with median filtering and gamma correction, and extensive evaluation with multiple metrics. This model is suitable for resource-limited environments and enhances clinical processes through rapid and reliable diagnosis. Phridviraj and colleagues [10] presented a method based on bidirectional long short-term memory (Bi-LSTM) with multi-scale Retinex with color preservation (MSRCP) for DR detection, evaluated on the MESSIDOR dataset. This approach faces challenges such as difficulty selecting MSRCP algorithm parameter values, potential overfitting on small datasets without proper data augmentation, and difficulty detecting small lesions in the early stages of DR. However, DL methods for DR detection offer significant advantages, including providing more accurate results, better performance in DR detection

compared to modern methods, and effectiveness in differentiating between different DR stages. The Bi-LSTM-MSRCP model achieved the best performance with 96.77% accuracy, while other methods like CNN, DCNN, ResNet 50, and RCNN achieved accuracies ranging from 92.38% to 95.42%. Additionally, models trained without data augmentation performed better in a multi-class scenario, and models trained solely with MSRCP showed the best overall performance. Douglas Abreu da Rocha et al [11] used the VGG16 neural network for classifying diabetic retinopathy. This model was evaluated on retinal images from the DDR, EyePACS/Kaggle, and IDRID databases. Although specific drawbacks were not mentioned in the text, the obtained results indicated significant advantages. This approach improved DR classification performance, with sensitivity, specificity, accuracy, and F1-score metrics used for model evaluation. Evaluation results showed that the VGG16 model achieved the best performance on the DDR database in terms of accuracy, precision, sensitivity, specificity, and F1-score. These findings demonstrate the high capability of the VGG16 neural network in diagnosing and classifying diabetic retinopathy, making it an effective tool in medical diagnosis. Overall, this research indicates that using the VGG16 neural network can effectively improve the accuracy and efficiency of DR classification, aiding doctors in providing more accurate diagnoses and timely treatment. These results can pave the way for further development of Artificial Intelligence (AI) based diagnostic systems in the medical field.

2. Proposed Method

Our proposed method aims to diagnose DR using the overall architecture of the PVT from retinal fundus images [12]. Given that a large number of patients suffer from DR, it is essential to diagnose the condition effectively. Although pyramid-based feature extraction has been used in previous studies, our work uniquely adapts PVT specifically for DR detection. By optimizing PVT's hierarchical structure, this study enables improved detection of small lesions and classification of the disease's various stages, highlighting our novel contribution to the field. The overall process is illustrated in the figure, and the steps are explained below.

10.22060/EEJ.2024.23500.5621

2-1- Overall Architecture of the PVT

The purpose of this section is to introduce a hierarchical structure to the Transformer framework, allowing the creation of multi-scale feature maps for dense prediction tasks, such as diagnosing DR from Retinal images. Following the hierarchical structure, the output resolution decreases gradually from the top (4 steps) to the bottom (32 steps). Our method comprises four stages that generate feature maps at different scales. All stages share a similar architecture, consisting of a patch embedding layer and multiple Transformer encoder layers. The Fig. 1 illustrates the overall architecture of the PVT [12].

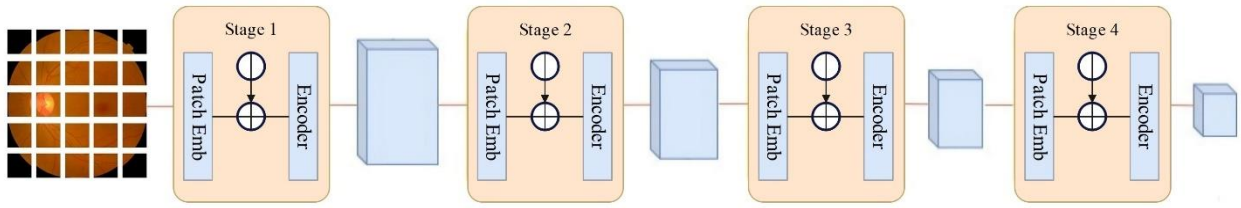

Fig. 1. Overall Architecture of the PVT

In the first stage, an input image with dimensions $H\times W\times 3$ (height (H), width (W), and 3 color channels) is divided into $\frac{16}{HW}$ patches, with each patch being $4\times 4\times 3$ in size. The flattened patches are then passed through a linear projection to obtain embedded patches with dimensions $\frac{HW}{16}\times C_1$. These embedded patches, along with positional embeddings, are then fed through a Transformer encoder with L_1 layers. The output is reshaped into a feature map F_1 with dimensions $\frac{H}{4}\times\frac{W}{4}\times C_1$. The subsequent stages follow a similar process.

2-2- Feature Pyramid for Transformer

Our PVT employs a gradual reduction strategy to control the scale of feature maps using patch embedding layers. At the beginning of each stage, the feature map from the previous stage is divided into patches of a specified size, and after linear projection, multi-scale feature maps are generated. This hierarchical

10.22060/EEJ.2024.23500.5621

structure enables the use of PVT in most downstream tasks, including image classification, object detection, and semantic segmentation.

2-3-Transformer Encoder

The Transformer encoder at each stage consists of multiple encoding layers, each comprising an attention layer and a feed-forward layer. Since PVT needs to process high-resolution feature maps, a Spatial Reduction Attention (SRA) layer is introduced to replace the Multi-Head Attention (MHA) layer. Similar to MHA, SRA takes a query, key, and value as inputs and produces a refined feature as output, as shown in Fig. 2. The difference is that SRA reduces the spatial scale of the key and value before the attention operation, significantly reducing computational and memory costs.

To gain a better understanding, it is useful to delve into the internal details of one of the stages, which remain consistent across all four stages. Fig. 3 illustrates the internal details of a single stage. The image is divided into segments of size $(3 \cdot P_i^2) \times \frac{W}{P} \times \frac{H}{P}$. Each segment is embedded into a feature space through a linear layer. The features are then reconstructed into the shape $C_i \times \frac{W}{P_i} \times \frac{H}{P_i}$, followed by the processes of Transformer Encoder, Normalization, SRA, Feed Forward, and Reshape. Key aspects include positional embedding to retain spatial information, element-wise addition to combine embeddings, and the feature map at each stage. This architecture is designed to progressively reduce the resolution of features at different stages, enabling the extraction of diverse features at various levels of abstraction, which can be highly beneficial for computer vision models.

AUT Journal of Electrical Engineering
10.22060/EEJ.2024.23500.5621

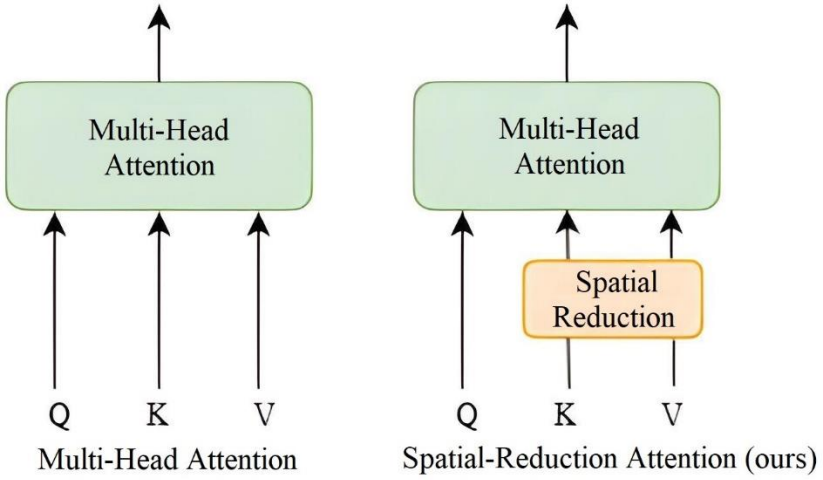

Fig. 2. Comparison of MHA with SRA.

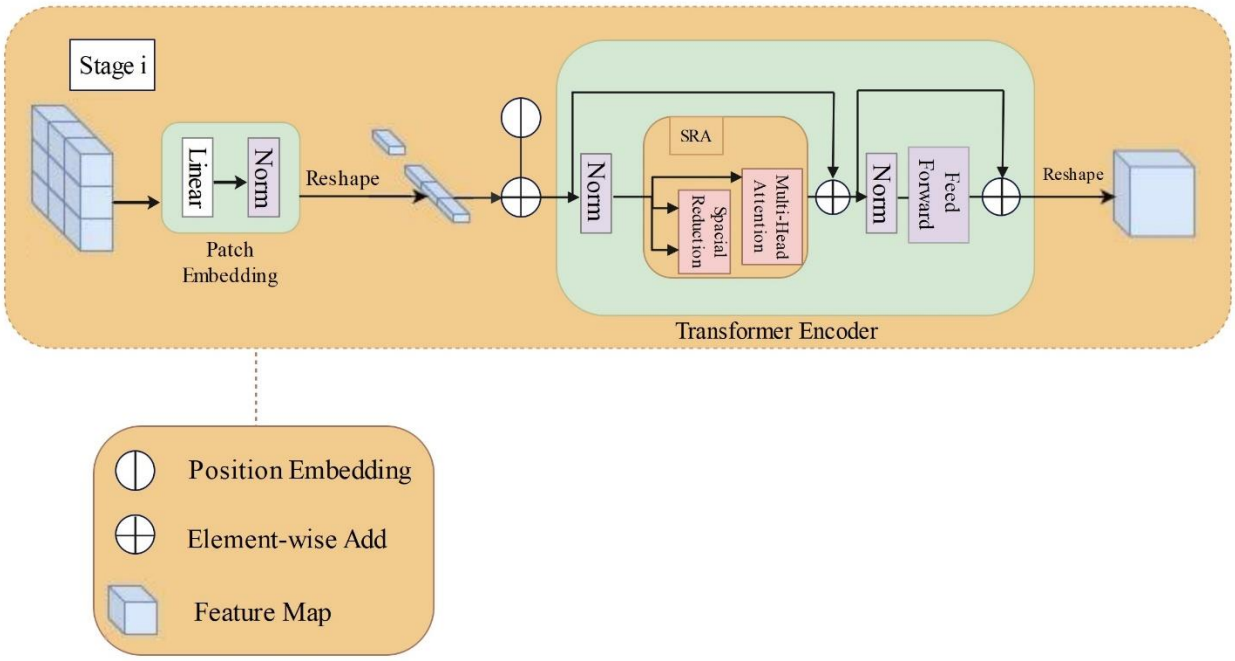

Fig. 3. Internal details of a stage

The SRA layer minimizes computational cost while preserving accuracy. By carefully selecting SRA configurations, we optimized model performance to maintain sensitivity in detecting small lesions. This approach ensures that the reduction in spatial resolution does not significantly affect the model's ability to identify minor anomalies, crucial for early detection in medical imaging.

Our proposed model leverages the hierarchical structure of the PVT, which stands out from the Vision Transformer (ViT) [17] due to its ability to generate multi-scale feature maps through a progressive pyramid structure. This architecture not only enhances resolution but also mitigates the computational strain associated with large medical images. Additionally, the SRA layer optimizes memory and computational resources, making PVT highly efficient for our application in DR detection.

3. Dataset

In this study, we utilized a dataset consisting of fundus images from patients in rural areas of India, collected by technicians from the Aravind Eye Hospital. After collection, these images were evaluated by specialized physicians and categorized into different classes. This dataset was published on Kaggle under the name "aptos2019-blindness-detection" [13].

The dataset we used includes 3,662 fundus images, which are divided into five classes based on the severity of the disease: Class 0 - No DR, Class 1 - Mild, Class 2 - Moderate, Class 3 - Severe, and Class 4 - Proliferative DR [5]. Examples of these images are shown in Fig. 4.

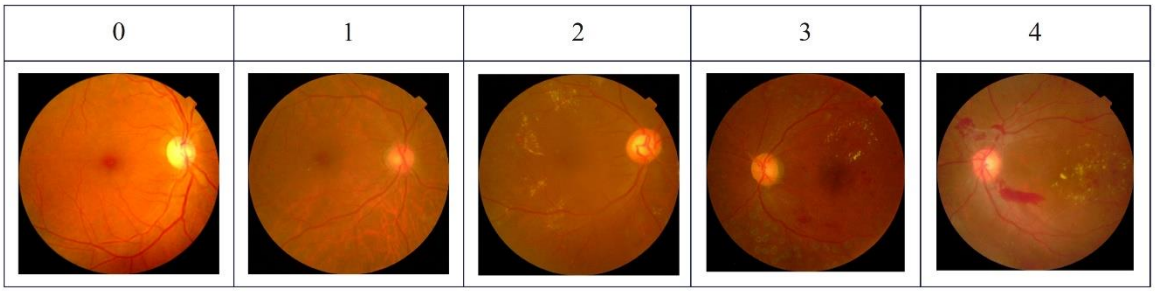

Fig. 4. Classes 0 to 4 based on DR severity

4. Results

4-1- Image Preprocessing

As shown in the Fig. 5, the samples are not evenly distributed across the classes. To address class imbalance, data augmentation techniques such as resizing, rotation, and contrast adjustment were applied before splitting the dataset. This approach ensured that both training and testing sets contained

10.22060/EEJ.2024.23500.5621

representative samples from each class, promoting consistent model generalization across all DR stages. These methods play a crucial role in creating a more balanced distribution of images across the five different classes. The data augmentation process generates additional images, helping to improve the balance of the dataset and reduce the challenges posed by class imbalance. This smart approach to data augmentation enhances the model's ability to generalize better and make more accurate predictions across all classes [14].

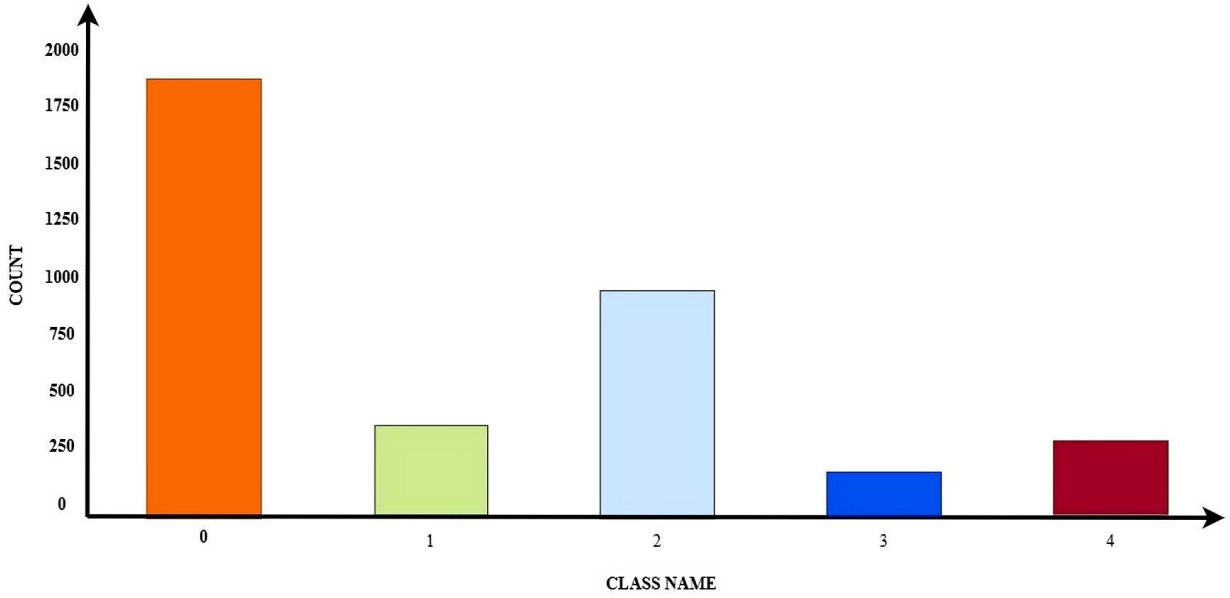

Fig. 5. Class imbalance in data distribution

4-2- Performance Parameters

4-2-1 AUC

AUC is a performance metric for binary classification models. AUC refers to the area under the Receiver Operating Characteristic (ROC) curve and indicates the model's ability to correctly distinguish between positive and negative classes. The ROC curve is a graph obtained by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) at various thresholds. This curve shows how well the model can differentiate between different classes [15].

10.22060/EEJ.2024.23500.5621

The TPR is the ratio of correctly predicted positive instances to the total actual positive instances. TPR is also known as Sensitivity or Recall. The FPR is the ratio of negative instances that are incorrectly predicted as positive to the total actual negative instances. Eq. 1 denotes the AUC.

$$AUC = \int_{+\infty}^{-\infty} TPR(T)FPR(T)dT \tag{1}$$

One of the advantages of using AUC is that, unlike metrics such as accuracy, AUC is not affected by the choice of threshold and provides an overall measure of the model's performance. It can also be a better metric than accuracy in situations where the data is imbalanced.

4-2-2 Accuracy

Accuracy is one of the primary metrics for evaluating the performance of disease detection models. Accuracy is defined as the ratio of correctly classified samples (both patients and healthy individuals) to the total number of samples. This metric indicates how well the model has made correct predictions [16]. Eq. 2 shows the formula for accuracy.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

TP (True Positive) is the number of correctly identified patient samples.

TN (True Negative) is the number of correctly identified healthy samples.

FP (False Positive) is the number of healthy samples incorrectly identified as patients.

FN (False Negative) is the number of patient samples incorrectly identified as healthy.

In other words, accuracy represents the ratio of total correct predictions (both correctly identifying patients and healthy individuals) to the total number of predictions. Accuracy is a general metric for evaluating the performance of a model and indicates how well the model performs overall.

4-2-3 Comparison of Performance Metrics

10.22060/EEJ.2024.23500.5621

In general, the PVT model is designed in various versions, each with different scales of parameters and complexity. These versions are typically named PVT-Tiny, PVT-Small, PVT-Medium, and PVT-Large. These names reflect the number of layers, model dimensions, and architectural complexity. In this study, we focus on the PVT-Tiny model. PVT-Tiny is the version of this model with the fewest parameters and layers, thus requiring fewer computational resources. This version is suitable for applications that demand high speed and lower memory consumption, while still benefiting from the transformer architecture's capability to extract rich features. The PVT-Tiny model has four stages and approximately 13 million parameters. Table 1 presents the number of layers and parameters per stage for the PVT-Tiny model.

Table 1. Comparison of results

Stage	Number of Layers	Multi-Head Attention	Number of Parameters (m)
1	2	1	0.7
2	2	2	2.1
3	2	5	4.9
4	2	8	5.3
Total	8	-	13

We have analyzed the results on data that have been augmented using various techniques to achieve a balanced distribution. This analysis aims to more accurately evaluate the performance of the PVT model under conditions where the data are balanced. For this evaluation, the data were randomly split into two different ratios of 80-20 and 90-10. This splitting allowed us to assess the model's performance under different conditions and to evaluate the results with greater precision. Table 2 presents all the obtained results as percentages, providing a comprehensive and detailed comparison of the experimental outcomes. This table enables us to clearly observe the differences and impacts of various data augmentation techniques and splitting ratios on the model's performance, leading to a more in-depth analysis of the model's effectiveness in different scenarios.

Table 2. Performance of the PVT-Tiny method

Validation method	Evaluation metrics	PVT-Tiny
80:20	Accuracy	84.32
	AUC	98.12

10.22060/EEJ.2024.23500.5621

90:10	Accuracy	92.38
	AUC	99.58

In conclusion, the performance of the proposed PVT-Tiny method has been thoroughly evaluated across several metrics, as depicted in the figures. Fig. 6 illustrates the relationship between epochs and accuracy, demonstrating the method's ability to achieve high accuracy over time. Fig. 7 presents the epoch and AUC graph, further confirming the model's effectiveness in distinguishing between classes. Fig. 8 shows the epoch and loss graph, indicating the method's consistent reduction in loss, which signifies model convergence. These figures collectively underscore the robustness and efficacy of our approach in handling the given tasks.

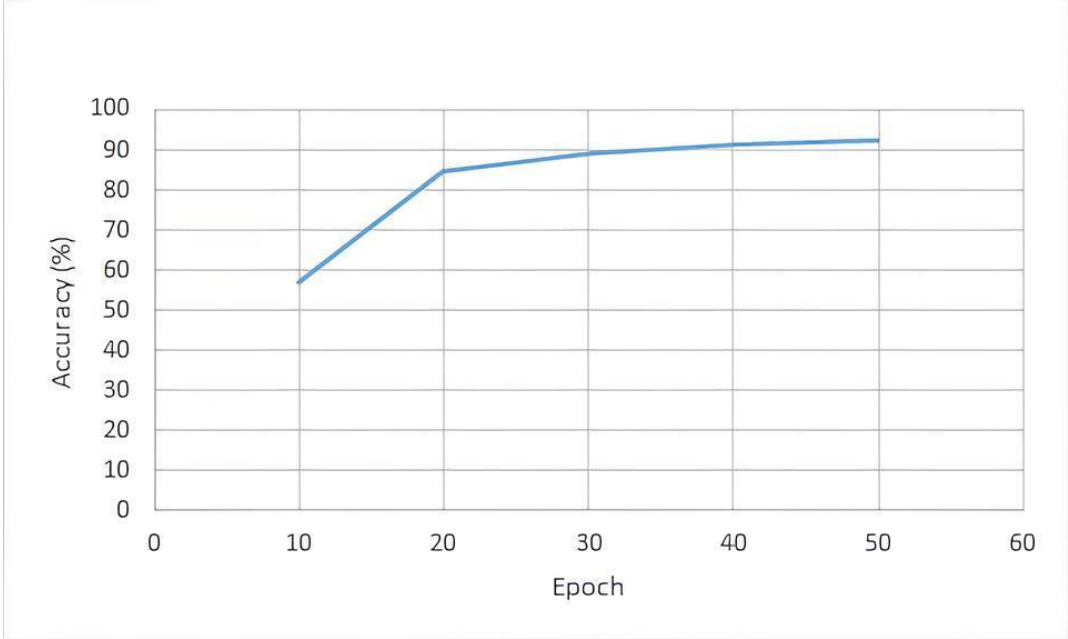

Fig. 6. Epoch and accuracy graph of the proposed method.

10.22060/EEJ.2024.23500.5621

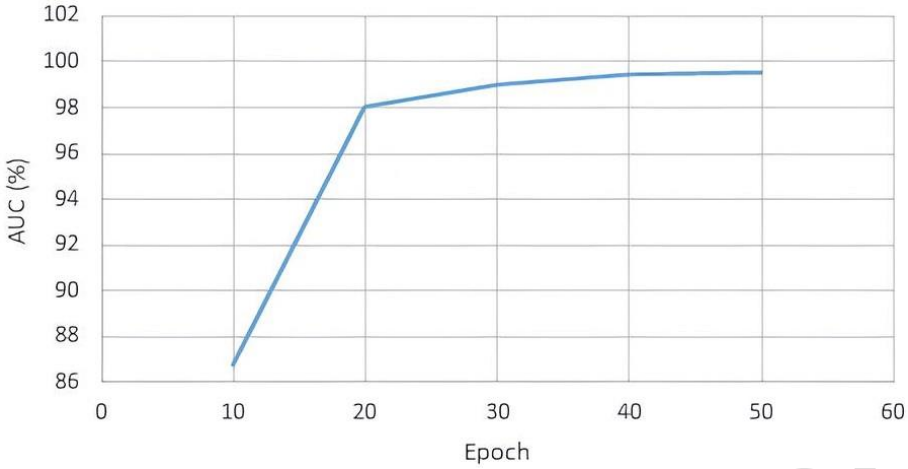

Fig. 7. Epoch and AUC graph of the proposed method.

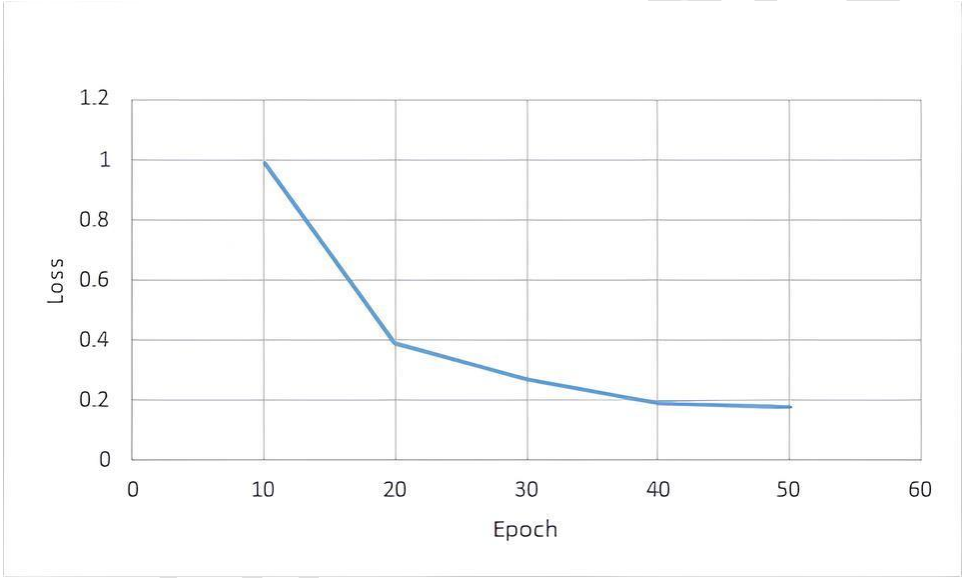

Fig. 8. Epoch and loss graph of the proposed method.

Finally, Fig. 9 provides a comparison of our research method's results with other methods, highlighting the competitive performance of the PVT-Tiny model.

10.22060/EEJ.2024.23500.5621

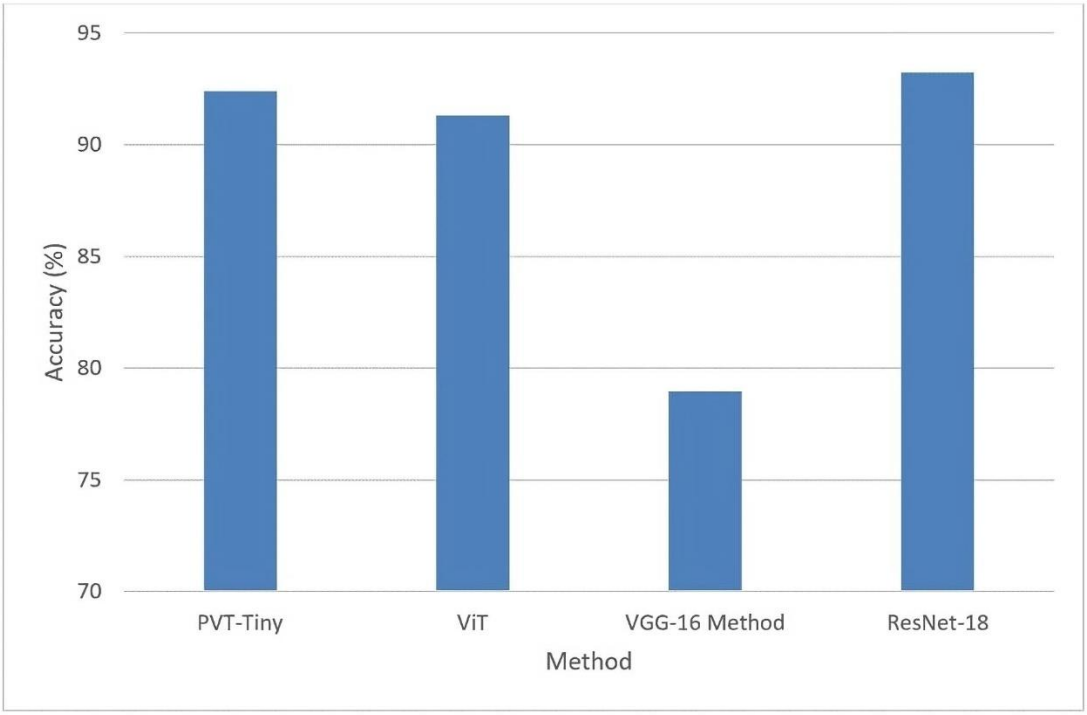

Fig. 9. Comparison chart of our research method results with other methods

The ACC and AUC metrics were calculated for the dataset to evaluate the proposed method. The evaluation results using images from the The Asia Pacific Tele-Ophthalmology Society 2019 (APTOS) dataset showed that our proposed algorithm outperformed the ViT and VGG-16 models. For the APTOS dataset, the ACC and AUC of the proposed method were 92.38% and 99.58%, respectively. These results indicate improvement in performance compared to ViT and VGG-16. Specifically, the proposed method achieved better results across all evaluation metrics. The Table 3 below clearly demonstrates the performance improvements of the PVT method compared to ViT and VGG-16.

The PVT offers a distinct advantage over traditional CNNs by using a hierarchical feature extraction method, which effectively captures both local and global image context. In CNNs, pooling layers progressively reduce spatial resolution, potentially losing critical spatial details necessary for detecting small lesions. PVT's pyramid-based structure retains both fine and coarse details at different scales, enabling enhanced detection of subtle DR indicators.

While the ViT set a foundation for Transformer models in vision tasks, its structure presents challenges for dense prediction applications, such as low resolution and high computational demand. PVT's pyramid-based architecture effectively overcomes these issues by maintaining high-resolution outputs and efficiently managing memory through spatial reduction. Our study demonstrates that PVT not only achieves superior accuracy but also offers a more computationally viable solution for high-resolution medical imaging.

Table 3. Comparison of results

Method	Accuracy	AUC
ViT	91.25	99.17
PVT-Tiny	92.38	99.58
VGG-16	78.94	85.11
ResNet-18	93.51	99.81

5. Discussion, Conclusion, and Future Work

In this study, we examined and evaluated the performance of three advanced DL models: PVT, ViT, VGG-16, and ResNet-18 for detecting DR from retinal images. Our primary goal was to assess the ability of the PVT model in comparison to the well-known ViT, VGG-16 and ResNet-18 models to deliver higher accuracy and efficiency in this critical medical task. The PVT model, with its progressively shrinking pyramid structure and SRA layers, demonstrates remarkable capability in generating high-resolution, multi-scale feature maps. These features allow the model to perform effectively in environments with limited computational and memory resources. Experiments showed that PVT, due to its superior ability to model long-range pixel relationships—especially in complex retinal images outperformed the other models in the precise detection and classification of DR. VGG-16, a well-established CNN architecture, utilizes a deep structure to extract local features from images. However, due to its inability to model long-range pixel relationships, its performance in detecting DR was more limited compared to PVT. This limitation was particularly evident in cases where small and scattered lesions were present in retinal images. The ResNet-18 model, with its use of residual blocks to address

learning challenges in deep networks, showed acceptable performance in extracting local features from images. However, VGG-16 was less effective than PVT in modeling long-range relationships between pixels. On the other hand, a significant advantage of PVT over VGG-16 was its ability to reduce noise and improve the quality of input images, which was achieved through image preprocessing techniques, leading to increased diagnostic accuracy. Overall, the results demonstrated that PVT, with its suitable architecture and superior ability to model complex relationships in retinal images, has a significant advantage over ViT and VGG-16 in detecting diabetic retinopathy. These findings highlight the high potential of using transformers in medical applications and early disease detection, opening new avenues for future research.

In the future, our plans include further development of the PVT model and its evaluation on larger and more diverse datasets. Although this study focuses on DR detection, the potential for testing on images with complex conditions, such as combined glaucoma cases or low-contrast images, is acknowledged. Future work will include a comprehensive evaluation under challenging imaging conditions to strengthen the model's robustness and reliability for clinical use. Additionally, exploring and integrating new technologies such as graph convolutional networks and various optimizations could bring further improvements in the model's accuracy and efficiency. Given that transformer-based models in computer vision are still in the early stages of development, we believe there are many potential technologies and applications to explore in the future. We hope that PVT can serve as a strong starting point for this journey. Our ultimate goal is to provide more accurate and reliable diagnostic systems for use in clinical settings. Finally, the PVT's hierarchical structure and multi-scale feature extraction capability suggest its applicability to other medical imaging tasks beyond diabetic retinopathy. Potential areas include skin lesion classification and tumor detection, where capturing subtle and large-scale features is essential. Future research will explore the application of PVT on diverse medical datasets to validate its versatility.

6. References

[1] S. Sunkari, A. Sangam, M. Suchetha, R. Raman, R. Rajalakshmi, S. Tamilselvi, A refined ResNet18 architecture with Swish activation function for Diabetic Retinopathy classification, Biomedical Signal Processing and Control, 88 (2024) 105630.
[2] A. Biran, Automatic detection and classification of diabetic retinopathy from retinal fundus images, Toronto Metropolitan University.
[3] K. Shankar, A.R.W. Sait, D. Gupta, S.K. Lakshmanaprabu, A. Khanna, H.M. Pandey, Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model, Pattern Recognition Letters, 133 (2020) 210-216.
[4] M. Canayaz, Classification of diabetic retinopathy with feature selection over deep features using nature-inspired wrapper methods, Applied Soft Computing, 128 (2022) 109462.
[5] V.P.C. Reddy, K.K. Gurrala, OHGCNet: optimal feature selection-based hybrid graph convolutional network model for joint DR-DME classification, Biomedical Signal Processing and Control, 78 (2022) 103952.
[6] S.Z. Beevi, Multi-Level severity classification for diabetic retinopathy based on hybrid optimization enabled deep learning, Biomedical Signal Processing and Control, 84 (2023) 104736.
[7] R.C. Joshi, A.K. Sharma, M.K. Dutta, VisionDeep-AI: Deep learning-based retinal blood vessels segmentation and multi-class classification framework for eye diagnosis, Biomedical Signal Processing and Control, 94 (2024) 106273.
[8] G.T. Zago, R.V. Andreão, B. Dorizzi, E.O.T. Salles, Diabetic retinopathy detection using red lesion localization and convolutional neural networks, Computers in biology and medicine, 116 (2020) 103537.
[9] F.J.M. Shamrat, R. Shakil, B. Akter, M.Z. Ahmed, K. Ahmed, F.M. Bui, M.A. Moni, An advanced deep neural network for fundus image analysis and enhancing diabetic retinopathy detection, Healthcare Analytics, 5 (2024) 100303.
[10] M. Phridviraj, R. Bhukya, S. Madugula, A. Manjula, S. Vodithala, M.S. Waseem, A bi-directional Long Short-Term Memory-based Diabetic Retinopathy detection model using retinal fundus images, Healthcare Analytics, 3 (2023) 100174.
[11] D.A. Da Rocha, F.M.F. Ferreira, Z.M.A. Peixoto, Diabetic retinopathy classification using VGG16 neural network, Research on Biomedical Engineering, 38(2) (2022) 761-772.
[12] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, L. Shao, Pyramid vision transformer: A versatile backbone for dense prediction without convolutions, in: Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 568-578.
[13] M. Karthik, S. Dane, Aptos 2019 blindness detection, Kaggle https://kaggle. com/competitions/aptos2019-blindness-detection Go to reference in, (2019) 5.
[14] M. Xu, S. Yoon, A. Fuentes, D.S. Park, A comprehensive survey of image augmentation techniques for deep learning, Pattern Recognition, 137 (2023) 109347.
[15] S. Wu, P. Flach, A scored AUC metric for classifier evaluation and selection, in: Second workshop on ROC analysis in ML, bonn, Germany, Citeseer, 2005.
[16] A. Gunawardana, G. Shani, A survey of accuracy evaluation metrics of recommendation tasks, Journal of Machine Learning Research, 10(12) (2009).
[17] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, An image is worth 16x16 words: Transformers for image recognition at scale, arXiv preprint arXiv:2010.11929, (2020).