

Original Article

DenseNet-169 and PatchTST-based Early Glaucoma Detection using Retinal Fundus Images

Nagaratna P Hegde¹, Vinay Kumar Sriperambuduri², Sireesha Vikkurty², Uma Patil³, Ganapatikrishna P. Hegde⁴

¹Department of Computer Science and Engineering, Vasavi College of Engineering (A), Hyderabad, India,

²Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad, India,

³Department of Computer Science and Engineering, Nutan Maharashtra Institute of Engineering and Technology, Dabhade, India,

⁴Department of Computer Science and Engineering, Shree Dharmasthala Manjunath Institute of Technology, Ujire, India.

¹Corresponding Author : nagaratnaph@staff.vce.ac.in

Received: 30 September 2025

Revised: 08 September 2026

Accepted: 18 September 2026

Published: 27 October 2026

Abstract - Glaucoma is a chronic ophthalmic disease that is characterized by progressive optic nerve damage. If untreated, it leads to permanent vision loss. Hence, an automated detection model is essential to detect glaucoma at an early stage to prevent from blindness. However, the conventional Deep Learning (DL) models struggle to learn the various characteristics that consist of Optic Disc (OD) with different sizes, shapes and positions, which results in detecting the early stage of glaucoma as normal. To solve this problem, in this research, a Patch Time Series Transformer (PatchTST) is proposed to classify between normal and glaucoma effectively. For glaucoma detection, the raw retinal fundus images from the two benchmark datasets are acquired and preprocessed using standard techniques to normalize image quality. Subsequently, a pretrained DenseNet-169 architecture is used for deep feature extraction; its dense connectivity pattern continuously concatenates multi-scale feature maps to capture both macro-level boundaries for varied OD sizes and micro-level localized textures. The PatchTST method maps the extracted deep features into tokenized sequence patches to learn fine-grained disease characteristics to classify healthy and diseased eyes efficiently. Experimental results demonstrate that the proposed method achieves classification accuracy of 98.84% on DRISHTI-GS and 98.98% on RIM-ONE datasets, respectively, while representing competitive glaucoma detection performance compared with existing methods, including Attention U-Net, with a modified ResNet-50 (AU-ResNet-50), under identical experimental conditions.

Keywords - Attention U-Net, DenseNet-169, Glaucoma detection, Optic Disc, Patch time series transformer model, Retinal fundus images.

1. Introduction

Glaucoma is a common eye disease caused by an imbalance between the production and leakage of intraocular fluid in the eye. This imbalance leads to increased pressure inside the eye, known as intraocular pressure, on the Optic Nerve Head (ONH) and damages the optic nerve. Glaucoma is a progressive optic nerve disease that causes irreversible blindness [1-3]. Thus, early detection and accurate diagnosis are essential to preserve vision and prevent disease progression [4]. However, the early detection of glaucoma is challenging because noticeable symptoms that help identify damage mostly appear only in the later stages of the disease [5]. Some of the common diagnostic techniques utilized for glaucoma detection include measuring Intraocular Pressure (IOP), performing visual field tests, and evaluating the ONH. Among these, ONH evaluation is the most widely recommended by glaucoma specialists and researchers. One of the most widely considered parameters in ONH evaluation

is the vertical Cup-to-Disc Ratio (CDR), which is typically larger in patients with glaucoma than in healthy individuals. In existing research, this ONH evaluation is performed manually from retinal fundus images, which is time-consuming and complex, leading to inaccurate results [6].

Hence, various artificial intelligence models are developed to segment the Optic Disc (OD) and calculate the CDR to detect the presence of glaucoma in retinal fundus images efficiently [7, 8]. Recently, Deep Learning (DL), a branch of artificial intelligence, has made significant advances in computer vision and is widely used for tasks such as recognition, detection, and classification [9-11]. DL models are also applied in medical image analysis, achieving better results by efficiently differentiating healthy and diseased images. Specifically, U-net- and V-net-based models are extensively utilized for OD and cup segmentation to achieve accurate results [12]. Unlike traditional approaches, deep



neural networks automatically learn meaningful representations from retinal fundus images, which helps identify discriminative features and enhances detection performance [13-15].

1.1. Problem Statement and Objective

Since the glaucoma does not cause noticeable symptoms in its initial stages, early detection is highly challenging, and the disease is often identified only after permanent vision loss occurs. Additionally, the various sizes, shapes, and positions of the OD in retinal fundus images make it difficult for conventional detection models to accurately identify the disease. To address these limitations, this research proposes an automated detection model designed to identify early indicators of glaucoma effectively from retinal fundus images. The framework incorporates image processing techniques to standardize image quality and mitigate structural variations in the OD, thereby improving overall detection performance.

1.2. Research Gap and Novelty

However, significant advances in DL models help for glaucoma detection; existing CNN-, attention-, and transformer-based models often struggle to effectively capture both localized retinal characteristics and long-range feature dependencies, particularly under variations in OD morphology and image quality. Moreover, existing traditional approaches rely heavily on complex segmentation pipelines, which increase computational complexity and limit generalization across datasets. To address these limitations, a DenseNet-169 with a Patch Time Series Transformer (PatchTST) is proposed to effectively detect glaucoma from retinal fundus images. The proposed model with a patch-based learning strategy preserves fine-grained retinal features while modeling global contextual relationships, enabling glaucoma detection.

The primary novelty of this research lies in the combination of DenseNet-169 feature reuse with a PatchTST learning strategy specifically designed for glaucoma detection. The proposed architecture eliminates the need for separate handcrafted feature extraction while effectively modeling both local and global retinal representations. The proposed model improves feature discrimination, computational efficiency, and classification consistency compared with conventional CNN-based, attention-guided, and existing transformer-based glaucoma detection methods.

1.3. Key Contributions

The key contributions of this research are as follows:

- The proposed PatchTST framework is proposed for glaucoma detection, which utilizes the high-level spatial features extracted by DenseNet-169 from retinal fundus images. DenseNet-169 extracts high-level spatial features, which are transformed into an ordered sequence of feature patches that are then processed by PatchTST to

capture discriminative spatial interactions through self-attention, thereby enhancing glaucoma classification.

- The DenseNet model efficiently captures significant image-based features, and its feature reusability function enhances the detection performance in the early stages of glaucoma.
- Patch-based temporal learning and DenseNet feature reuse reduce overfitting and improve robustness. The proposed glaucoma detection method performs well on fundus images acquired from various datasets, which enhances the generalizability of the model.

The rest of this paper is organized as follows. Section 2 reviews related studies on automated glaucoma identification. Section 3 details the proposed DL framework for feature extraction and classification. Section 4 presents the experimental results with a comprehensive discussion, and Section 5 concludes the paper.

2. Literature Survey

The advantages and limitations of existing DL models used for the detection and binary classification of glaucoma are discussed below.

Sandra Virbukate et al. [16] represented a CNN-based glaucoma identification model to segment the OD and cup in fundus images. A modified attention U-Net with various pretrained CNN models, such as MobileNet, DenseNet-121, ResNet-50, and InceptionNET-v3, was utilized to enhance glaucoma detection. Additionally, a unanimous voting method, also known as the majority voting technique, was used to provide accurate results from the ensemble pretrained models. However, the represented model faces difficulties due to the similar characteristics of early-stage glaucoma and healthy fundus images, which struggle to distinguish between these classes and lead to poor detection performance.

Shyamalee et al. [17] developed an Attention U-Net with a modified ResNet-50 (AU-ResNet-50) model for glaucoma detection using fundus images. This developed detection model utilized a ResNet-50-based attention U-Net and a modified Inception V3 model to segment and classify the diseased images efficiently. The developed AU-ResNet-50 model enhanced the glaucoma detection by learning subtle visual patterns in fundus images, which focus on the key indicators such as CDR and neuroretinal rim thinning. However, blurriness, poor lighting and artifacts in the fundus images affect the performance of the developed model and led to misclassification.

Vijaya Kumar Velpula et al. [18] developed a DL-based glaucoma detection model, namely, Deep CNN (DCNN). Resnet101 and Nasnet_large were the CNN models utilized with the developed DCNN model in the classification of healthy and diseased images. Pretrained CNN models can

efficiently distinguish features from retinal images. For classification, a DCNN model with various filter sizes was utilized to learn significant information with several parameters to improve glaucoma detection. However, the developed DCNN model failed to detect glaucoma accurately due to similar characteristics between the healthy and early-stage glaucoma, leading to inaccurate detection performance.

K.J. Subha et al. [19] presented an ensemble DL-based glaucoma detection model that includes Visual Graph Geometry (VGG-19), ResNet-50, and InceptionNet-V3 models. The presented ensemble model was used to identify glaucoma with a dual attention mechanism, such as spatial and channel attention modules, which efficiently focused on the OD in fundus images. Nevertheless, variations in image illumination and the presence of artifacts in the fundus images greatly impact the detection performance of the ensemble DL framework and its overall generalization ability.

Sonti and Dhuli [20] designed a Sequential Deep Neural Network (SDNN) model with an OD localization method to detect glaucoma. This designed framework was utilized to capture highly discriminant features directly from the raw pixel intensities of the fundus images. To improve detection accuracy, the developed model primarily focused on the OD region, the disease indicator, namely CDR and eliminated irrelevant information. However, the developed SDNN struggled to fully capture the long-range spatial dependencies within the fundus images, which limits the model's accurate classification performance.

Pradeep Kumar et al. [21] developed a hybrid glaucoma detection framework based on segmentation and classification architectures to improve retinal structure segmentation and glaucoma classification. The framework employed U-Net for structural segmentation and capsule networks combined with an Extreme Learning Machine (ELM) for feature extraction and classification. The Capsule networks helped to preserve the spatial relationships between retinal structures, improving feature representation. However, the integration of segmentation and classification models with attention-based processing struggled with the similar characteristics of early-stage glaucoma and normal fundus images.

Pannu et al. [22] designed a modified Residual U-Net (ResUNet) architecture for joint OD and cup segmentation to support glaucoma classification. The method standardized retinal fundus images through a designed model that utilized an encoder-decoder-based segmentation to estimate the CDR and Neuro-Retinal Rim (NRR) enhanced classification. The combination of CDR and NRR improved the extraction of structural information for the classification of glaucoma. However, the encoder-decoder architecture relied on direct skip connections without attention-guided feature refinement, which led to the capture of irrelevant background information that affected the classification results.

Recent glaucoma detection methods have demonstrated encouraging performance using pretrained CNN, ensemble learning, and transformer-based architectures. However, pretrained CNN models effectively learn hierarchical spatial features but have limited capability to capture long-range relationships among retinal structures. However, Ensemble DL methods improve prediction accuracy by combining multiple classifiers, which increases computational complexity.

Additionally, Transformer-based models enhance global contextual learning through self-attention mechanisms but do not sufficiently preserve localized retinal characteristics that are vital for identifying subtle glaucomatous changes in fundus images. To address these limitations, this research proposes a DenseNet-169 with a PatchTST to enhance efficient local feature extraction with patch-based contextual learning. The proposed framework maintains fine-grained retinal information while modeling long-range dependencies, thereby improving glaucoma detection performance across benchmark datasets.

3. Methodology

The proposed framework consists of four major phases structured sequentially into nine operational stages, as illustrated in the conceptual diagram in Figure 1.

Initially, raw retinal fundus images from the DRISHTI-GS and RIM-ONE datasets are obtained during dataset acquisition and passed to image preprocessing (Stages 1 - 2), where they are processed using Contrast Limited Adaptive Histogram Equalization (CLAHE) for structural clarity. Subsequently, during the deep feature extraction phase (Stage 3), a pretrained DenseNet-169 architecture extracts a dense 512-dimensional representation vector. To initiate spatial sequence tokenization (Stages 4 - 6), this feature vector is reshaped into a structured tensor shape to enable the generation of overlapping patches that capture localized retinal textures, which are then linearly projected into a 512-dimensional embedding space and combined with learnable positional encodings. Finally, the transformer-based classification phase (Stages 7 - 9) passes these position-encoded embeddings through a 4 - layer PatchTST transformer encoder to efficiently differentiate between normal and glaucomatous images.

3.1. Data Acquisition

In this research, two publicly available datasets of retinal fundus images are used to train and validate the proposed glaucoma detection model.

Datasets, such as DRISHTI-GS and RIM-ONE, are utilized as input data in this proposed glaucoma detection framework. Table 1 shows the dataset description based on various features of glaucoma detection.

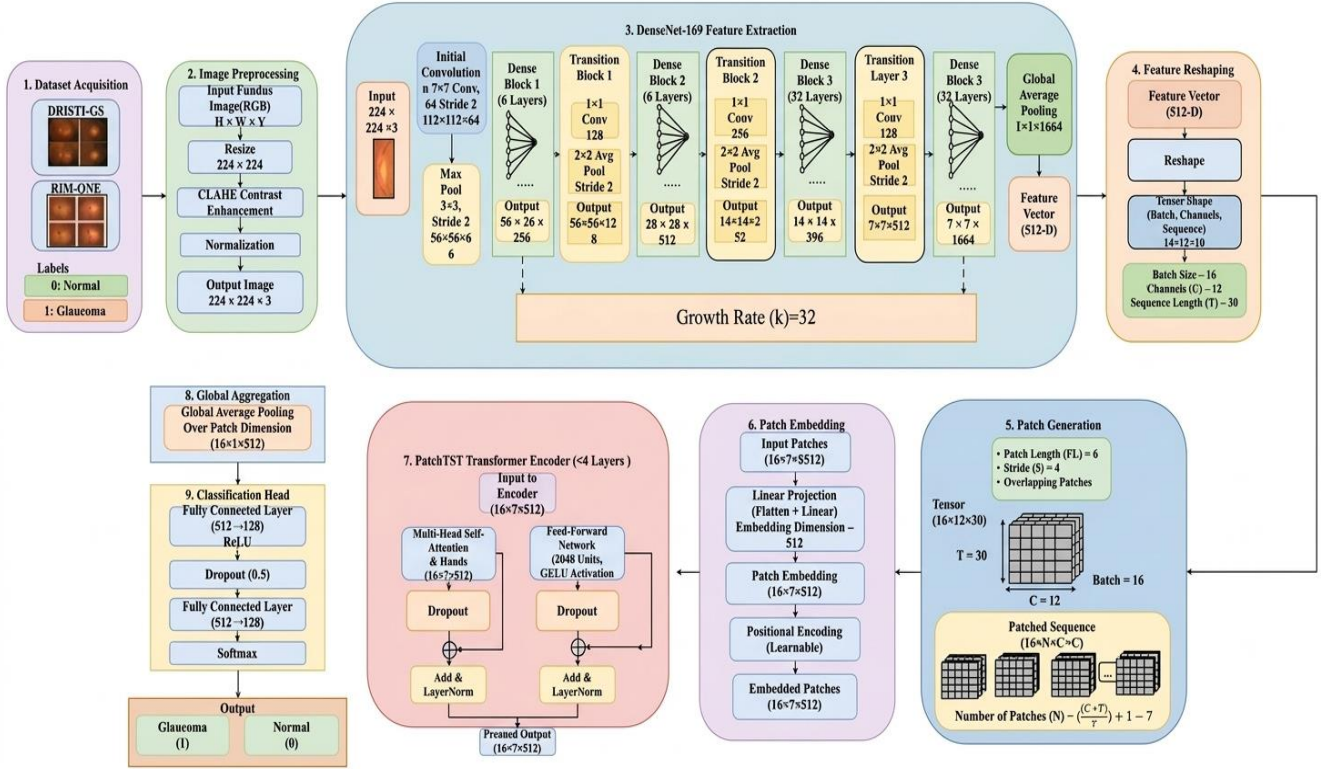


Fig. 1 Overall DenseNet-169 and PatchTST framework for automated glaucoma detection using retinal fundus image

3.1.1. DRISHTI-GS

This is a publicly available dataset [23] that comprises 101 images of healthy individuals and patients with glaucoma. The fundus images in DRISHTI-GS are annotated by the Aravind Eye Hospital in Madurai, and are centered around the OD, which has a field view of 30°.

3.1.2. RIM-ONE

It is an open-source dataset [24] that is widely used for glaucoma detection and contains retinal fundus images. It has 255 healthy images and 200 glaucoma images that are manually segmented by medical specialists. The acquired retinal fundus images are fed into the preprocessing stage of the proposed glaucoma detection framework.

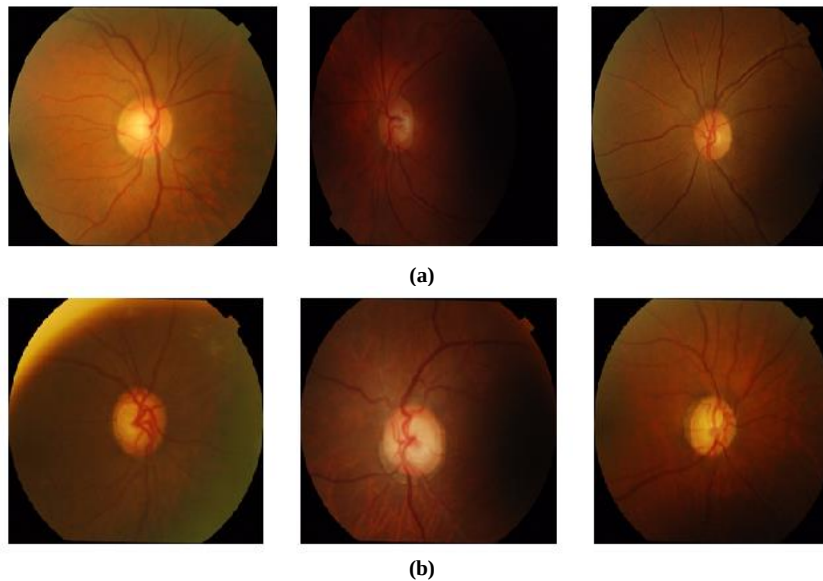


Fig. 2 (a), and (b) Sample of normal and glaucoma retinal fundus images from the DRISHTI-GS and RIM-ONE dataset used for evaluation.

The format, resolution, and training and testing splits of this dataset are listed in Table 1. Figures 2 (a) and (b) illustrate

sample images of the DRISHTI-GS and RIM-ONE datasets utilized in the proposed glaucoma detection framework.

Table 1. Dataset description of DHRISH-GS and RIM-ONE used for glaucoma detection based on various parameters

Features	DRISHTI-GS	RIM-ONE
Total Images	101	455
Glaucoma Images	71	200
Non-Glaucoma Images	30	255
Image Format	.png	.jpg/.png
Resolution	2896 × 1944 pixels	It varies
Training	50	Approx. 317
Testing	51	Approx. 138

3.2. Preprocessing

After data acquisition, the fundus images are passed to the preprocessing stage because it is essential in this glaucoma detection to convert raw images into a useful format. Thus, techniques such as resizing and Contrast Limited Adaptive Histogram Equalization (CLAHE) are utilized in this research during the preprocessing phase. The acquired raw fundus images vary in size, and the model is used to train and learn glaucoma-related information. Thus, the obtained retinal fundus images are initially resized and fed into the image enhancement technique to enhance the training of the detection model.

Resizing: Retinal fundus images are acquired from two benchmark datasets with different resolutions that affect further processes. Therefore, these raw images are resized using a resizing technique that is commonly utilized to standardize the resolution for enhancing the detection process. All acquired images are resized to a standard resolution of 224×224 pixels.

CLAHE-based image enhancement: Next, the resized fundus images are enhanced using the CLAHE [25] technique, which is broadly utilized in image processing to enhance the quality of images. However, retinal images contain blur and varying lighting conditions, which makes it difficult to capture features with relevant information. Hence, the image quality is assessed using an improvised histogram-equalization-based image enhancement technique.

The CLAHE technique first divides the resized image into smaller parts and adjusts the contrast level of each pixel. The adaptability of CLAHE helps to avoid either increasing the contrast too bright or decreasing the contrast too dark in certain portions, which helps in the precise enhancement of the image quality. The clip limit in the CLAHE technique is utilized to leverage the performance of histograms for each nonoverlapping region, which is mathematically represented in Equation (1).

$$\beta = \frac{MN}{G} \left\{ 1 + \frac{\alpha}{100} (AS_{\max} - 1) \right\} \quad (1)$$

Where G represents the number of grayscale levels, α denotes the clip factor, and M and N refer to the number of pixels in each region; $AS_{\max}$ indicates the maximum allowance slope. After image enhancement, these contrast-enhanced fundus images are passed to the feature extraction phase to capture the discriminative features.

3.3. Feature Extraction

Next, the preprocessed images are forwarded as inputs to the feature extraction model using a CNN model. In this research, the DenseNet-169 model is used for feature extraction for glaucoma detection. The DenseNet-169 architecture is selected as the backbone feature extractor for glaucoma detection because of its ability to learn rich hierarchical feature representations through densely connected convolutional layers, thereby promoting efficient feature reuse and improved gradient propagation.

This helps the DenseNet-169 model learn rich hierarchical information from preprocessed images with dimensions of 224×224 . This architecture effectively captures both low-level retinal characteristics, such as vessel boundaries and OD structures, and high-level semantic representations associated with glaucoma-related changes. In this research, the final classification layers of DenseNet-169 are removed, and only the deep feature extraction layers are retained to generate discriminative feature representations for the proposed PatchTST model. The extracted feature vectors provide comprehensive local structural information that facilitates accurate glaucoma classification while improving feature representation and computational efficiency.

3.4. Proposed Glaucoma Detection Model

The PatchTST encoder models the contextual relationships among these ordered feature patches using self-attention, enabling the framework to capture both local and global spatial dependencies within the retinal image. Each sequence element represents a localized retinal feature region rather than a temporal observation region. The PatchTST encoder models contextual interactions among these ordered feature patches, enabling the integration of local structural characteristics and global spatial relationships. This

adaptation enables the PatchTST encoder to learn discriminative contextual representations from ordered retinal feature patches, thereby improving the glaucoma

classification performance from single retinal fundus images. Figure 3 shows the architecture of the PatchTST-model-based glaucoma detection.

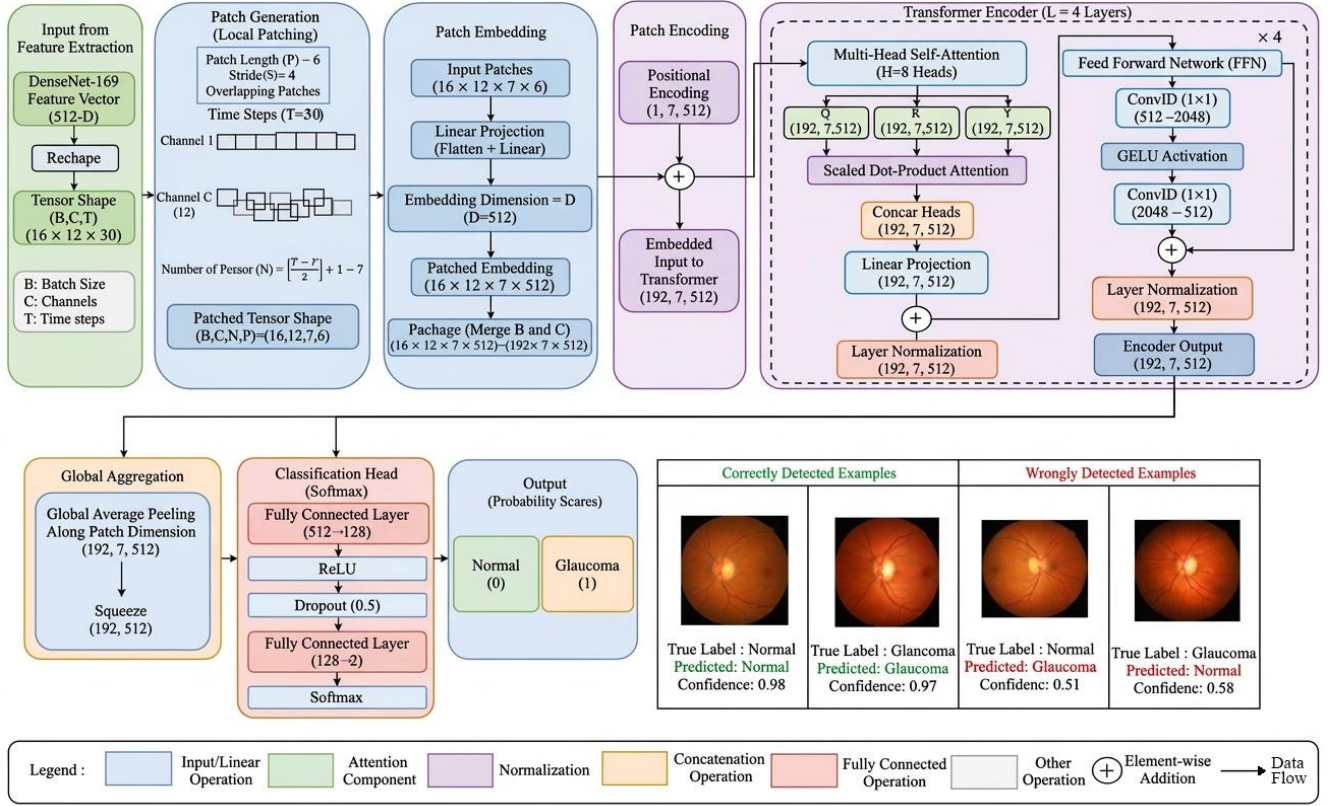


Fig. 3 Patch embedding and PatchTST encoder architecture for contextual retinal feature learning and glaucoma classification

3.4.1. Patch Embedding

Initially, the extracted features are fed to the proposed PatchTST model preprocessed by the DT module with standardization of height and width, which is first reshaped to a tensor of shape (16, 12, 30). After DenseNet-169 feature extraction, the feature maps are reshaped into an ordered sequence of 30 feature patches, each represented by a 12-dimensional embedding. The ordering follows the spatial arrangement of the extracted feature maps and is used solely to enable Transformer-based contextual feature learning. Therefore, the sequence dimension represents ordered spatial feature patches and should not be interpreted as temporal observations or glaucoma classification.

The proposed PatchTST framework transforms the extracted convolutional feature maps into an ordered sequence of feature patches for contextual representation learning. Let the retinal fundus image be represented as $I \in RH \times W \times CI$, which denotes the height, width, and number of channels of the image. DenseNet-169 extracts high-level feature maps, which are denoted as $F \in RHf \times Wf \times DF$. These feature maps are partitioned into N non-overlapping patches, where each patch is projected into a d -dimensional embedding to

construct the ordered feature sequence. To divide this feature data into local patches, the Patch Length (PL) is set to six, and the Stride (S) is set to four. The patching process is guided by Equation (2).

$$PN = \frac{(L-PL)}{S} + 1 \quad (2)$$

Where T denotes the number of time steps, PL represents the length of each patch, S refers to the stride in Equation (2), and PN indicates the number of patches. The patch generation process from the extracted feature mask using the patch generation module is represented in Figure 3. This structure ensures that channel independence is preserved while grouping correlated time samples within each patch, thereby enabling the model to learn both the temporal and local dependencies.

3.4.2. Embedding and Transformer Encoder

After patching, the shape (16, 12, 7, 6) is reshaped to (192, 7, 6) by merging the batch and channel dimensions (16 × 12 = 192). This tensor is then passed through an embedding layer that transforms each input patch into a high-

dimensional representation with shape (192, 7, 512). In the transformer encoder, the embedded input is passed through three parallel linear layers to generate Q (query), K (key), and V (value) matrices.

These matrices are used in the multi-head attention mechanism, as computed using Equations (3) and (4).

$$\text{Score}(\omega) = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right) \quad (3)$$

$$C = \text{Score}(\omega) \cdot V \omega V \quad (4)$$

Where d_k is the dimension of the key used for scaling, and C is the output of the attention mechanism representing weighted contextual representations. Then, the result is added to the input through a residual connection with layer normalization, keeping the shape as (192, 7, 512). The above mathematical formulation depicts the complete description of the proposed PatchTST framework model, which explains the transformation of convolutional feature maps into ordered feature-patch embeddings and the subsequent contextual feature learning process. Next, the self-attention mechanism captures the relationships between spatially ordered retinal feature patches, whereas the feedforward network refines the learned representations before classification.

3.4.3. Feedforward and Final Encoder Output

The normalized output is then passed through a Feed Forward Network (FFN) with the following architecture: Conv1D $\rightarrow$ GELU $\rightarrow$ Conv1D. Initially, the first Convolutional 1 Dimension (Conv1D) layer expands the dimension to 2048, and then the Gaussian Error Linear Unit (GELU) function introduces nonlinearity to learn the fine-grained details from the feature. Subsequently, the second Conv1D reduces the dimensionality to 512. Thus, the shape flows in another norm layer are applied to produce the output of the encoder, maintaining the original shape (192, 7, 512), which helps to differentiate between normal and glaucoma-diseased images efficiently.

In glaucoma detection using the proposed PatchTST model, the input is typically a sequence of extracted features using DenseNet-169 data, such as IOP, retinal nerve fiber, and thickness. These features are divided into fixed-length patches, which enable the model to process ordered feature patch information in terms of patches rather than point-by-point. This patching mechanism helps reduce redundancy and highlights the relevant temporal patterns that are essential for detecting glaucoma in the early stages. The transformer-based architecture within the patch utilizes a multi-head attention mechanism to effectively model long-range dependencies between different time intervals, which helps to identify subtle changes in the OD and retinal fiber and results in accurate detection of early-stage glaucoma.

4. Results and Discussion

This section presents the experimental results of the proposed PatchTST-based glaucoma detection model evaluated across two distinct benchmarks, namely the DRISHTI-GS and RIM-ONE datasets. The framework is implemented in Python 3.9 with Windows 10 OS, a Core i7 processor and 16 GB of RAM, to evaluate detection performance.

4.1. Experimental setup

4.1.1. Dataset Description

For the DRISHTI-GS dataset, 50 images are used for training, and 51 images are reserved for independent testing, following the dataset protocol. For the RIM-ONE dataset, approximately 70% of the images (317 images) are allocated for training, and the remaining 30% (138 images) are used for testing. During training, the training partition is further evaluated using stratified 5-fold cross-validation to preserve the class distribution in each fold and to assess the generalization capability of the proposed model. The independent test set is excluded from the cross-validation process and used only for the final performance evaluation to prevent data leakage. The descriptions of the three benchmark datasets considered in this research for performance evaluation are summarized in Table 2.

Table 2. Dataset description for performance evaluation of the proposed glaucoma detection model

Dataset Name	Total Images	Glaucoma	Normal	Train Split	Test Split	Image Resolution
DRISHTI-GS	101	71	30	50	51	2049 x 1757 (Resized 256x256)
RIM-ONE (r2)	455	200	255	317	138	Varies (Resized 256x256)
ACRIMA (External)	705	396	309	493	212	2048 x 1536 (Resized 256x256)

During preprocessing, retinal fundus images from the DRISHTI-GS and RIM-ONE datasets are resized to 224 $\times$ 224 pixels and enhanced using the CLAHE technique before feature extraction. Next, DenseNet-169 is employed as the backbone feature extractor, and the extracted feature representations are provided as input to the proposed

PatchTST model. The proposed detection model is trained using the Adam optimizer with an initial learning rate of 0.0001, a batch size of 16, and 50 training epochs. Binary Cross-Entropy (BCE) is utilized as the objective loss function in the proposed model for glaucoma classification. To ensure a fair comparison, all experiments are performed using a fixed

random seed of 42. The datasets are divided into training and testing subsets according to the protocol described in Table 2, and 5-fold cross-validation is performed to evaluate the generalization capability of the proposed model.

Moreover, during inference, the PatchTST model processed the retinal fundus images individually using a batch size of 1. The average inference time is calculated over the complete test dataset, whereas the computational complexity is evaluated in terms of execution time and memory utilization under the same hardware configuration for all methods utilized for evaluation.

4.1.2. Reimplementation and Hyperparameter Settings

An identical implementation environment and evaluation protocol are maintained throughout the experiments to ensure a fair and consistent comparison between the proposed framework and existing glaucoma detection methods. Table 3 summarizes the hyperparameter details of the proposed model.

The hyperparameter values are selected based on preliminary experiments and commonly adopted configurations for transformer-based image classification models. The embedding dimension, number of attention heads, feed-forward network dimension, patch length, and

stride are chosen to balance the feature representation capability and computational efficiency. The Adam optimizer with an initial learning rate of 0.0001 provided stable convergence during training, whereas a dropout rate of 0.10 reduced overfitting. Additionally, the pretrained DenseNet-169 weights initialized from ImageNet accelerated convergence and improved feature extraction from retinal fundus images. Moreover, the same hyperparameter configuration is maintained throughout all experiments to ensure a fair comparison with the baseline methods.

The limited sample sizes of the DRISHTI-GS and RIM-ONE benchmark datasets are considered during the implementation and evaluation of the proposed framework. In this research, the DenseNet-169 model initialized with ImageNet pretrained weights provided transferable feature representations, while stratified 5-fold cross-validation preserved the class distribution during model development and reduced the dependence on a single data partition. Moreover, the proposed glaucoma detection framework is evaluated on two independent benchmark datasets using identical experimental and evaluation protocols to provide a consistent assessment of its classification performance. These experimental settings improved the reliability of the reported results and ensured the generalization capability of the model across the evaluated retinal fundus datasets.

Table 3. Hyperparameter details of the proposed PatchTST-based glaucoma detection

Hyperparameters	Value
Input image size	224 × 224
Backbone network	DenseNet-169
Feature dimension	512
Patch length (P)	6
Patch stride (S)	4
Embedding dimension	512
Number of transformer encoder layers	4
Number of attention heads	8
Feed-forward network dimension	2048
Activation function	GELU
Dropout rate	0.10
Optimizer	Adam
Initial learning rate	0.0001
Learning rate scheduler	ReduceLROnPlateau
Batch size	16
Number of training epochs	50
Loss function	Binary Cross-Entropy (BCE)
Weight initialization	Pretrained ImageNet weights (DenseNet-169)
Random seed	42
Cross-validation	Stratified 5-fold
Inference batch size	1

4.2. Performance Metrics

Various performance metrics are utilized in this research to evaluate the performance and effectiveness of the proposed PatchTST model.

The mathematical representation of the performance metrics is given in Equations (5)-(8).

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \times 100 \quad (5)$$

$$\text{Precision} = \frac{TP}{TP+FP} \times 100 \quad (6)$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100 \quad (7)$$

$$\text{F1 - score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100 \quad (8)$$

Where TP and FP denote True Positives and False Positives, respectively, and TN and FN represent True Negatives and False Negatives, respectively.

4.3. Performance Analysis

In this section, the quantitative and qualitative analyses proposed PatchTST model are represented based on the two benchmark datasets and one external dataset, such as the DRISHTI-GS, RIM-ONE and the external dataset, namely ACRIMA.

Figures 4 to 7 present the performance evaluation of the feature extraction method and proposed glaucoma detection model with existing state-of-the-art algorithms using the two benchmark datasets.

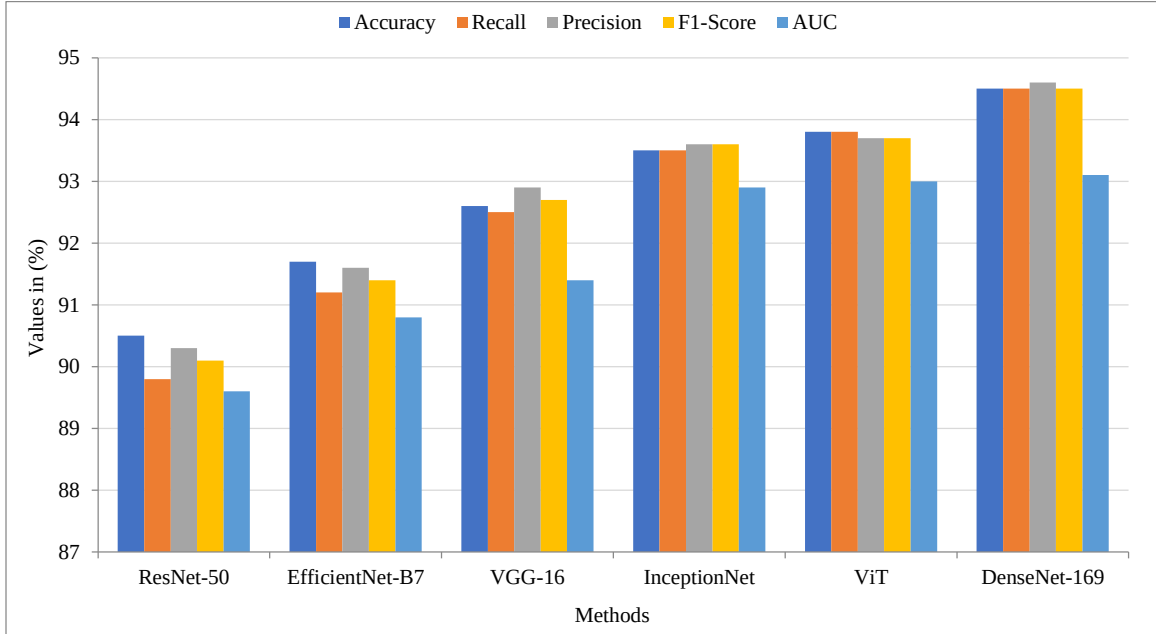


Fig. 4 Evaluation of DenseNet-169 feature extraction performance method using DRISHTI-GS dataset with state-of-the-art models

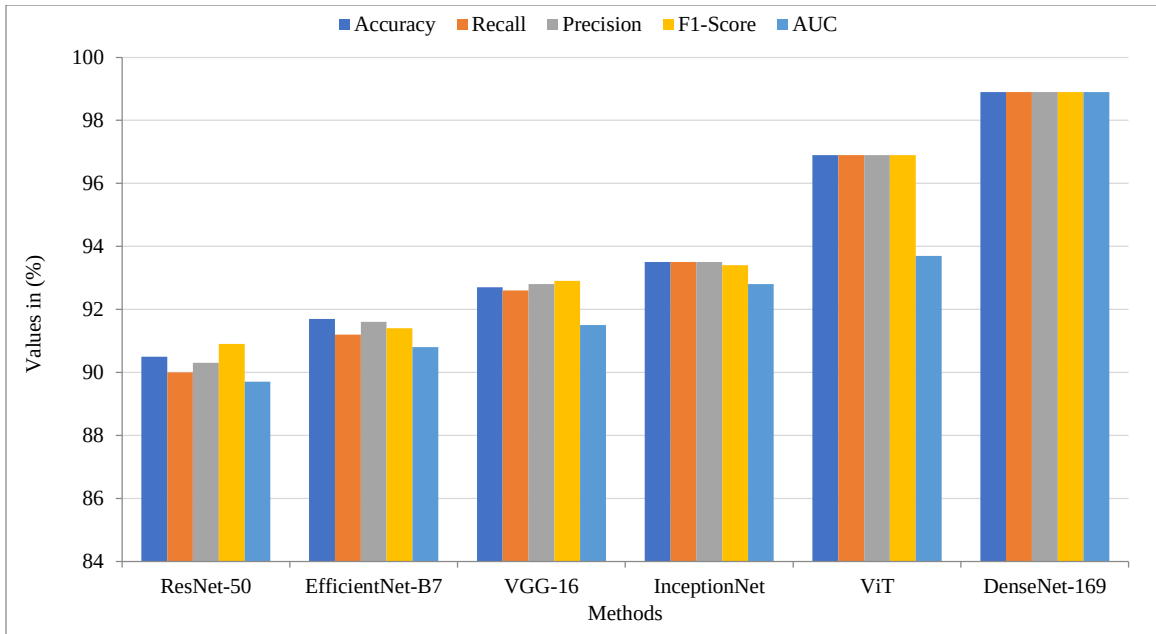


Fig. 5 Performance evaluation of the DenseNet-169-based feature extraction and existing DL baselines on the RIM-ONE dataset

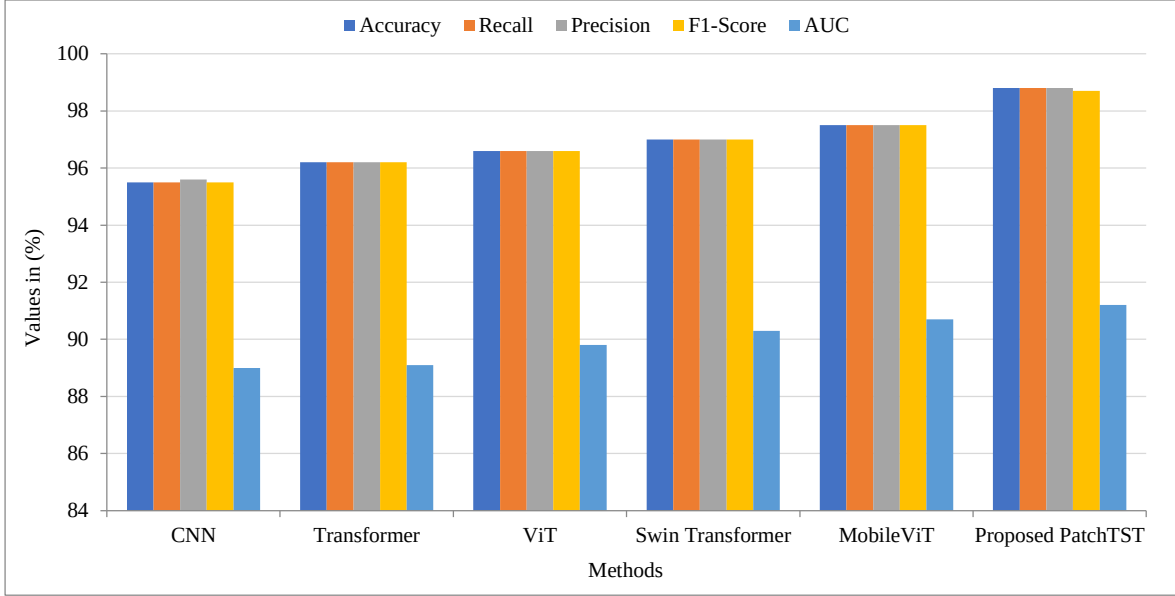


Fig. 6 Performance evaluation of the PatchTST-based glaucoma detection method using DRISHTI-GS with existing detection methods

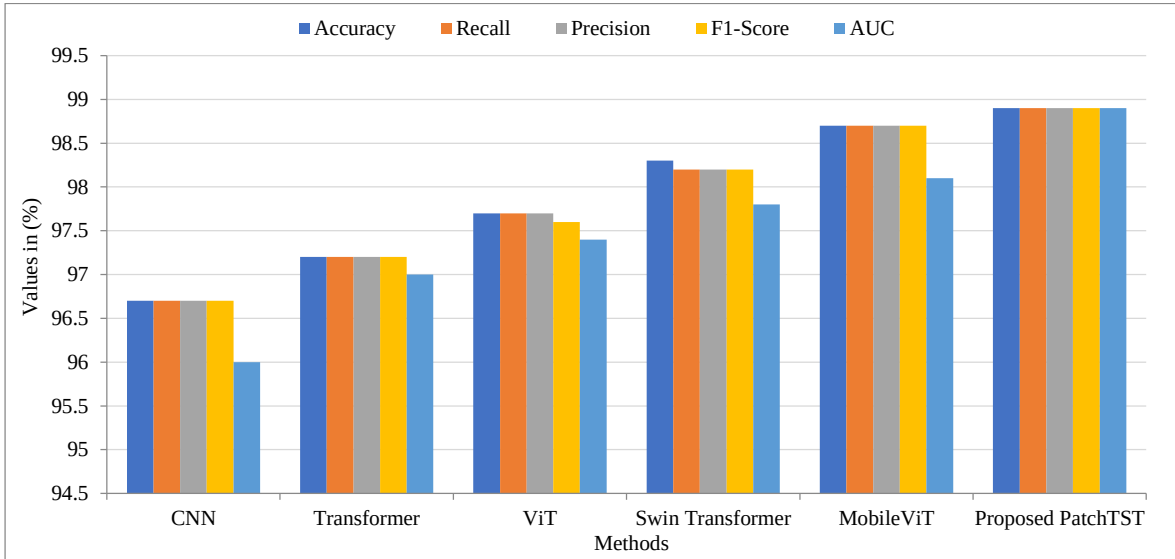


Fig. 7 Quantitative comparison of the proposed PatchTST-based framework and state-of-the-art models using the RIM-ONE dataset

The existing DL based feature extraction models, such as ResNet-50, DenseNet-169, VGG, and InceptionNet, are used to evaluate with the DenseNet-169 model, as shown in Figures 4 and 5. Moreover, Figures 6 and 7 illustrate the classification

performance of the proposed glaucoma detection model, which is compared with the state-of-the-art models such as CNN, Transformer, Vision Transformer (ViT), Swin Transformer, and Mobile ViT.

Table 4. Computational complexity of the proposed PatchTST model-based glaucoma detection using retinal fundus images

Methods	Training Time (s/epoch)	Memory (MB)	Inference Time (s/image)	Parameters (M)	FLOPs (G)
ResNet50	22.8	310	0.041	25.56	4.09
DenseNet121	20.9	295	0.038	7.98	2.88
EfficientNet-B0	24.3	325	0.043	5.29	0.39
MobileNetV2	18.7	240	0.031	3.50	0.30
Vision Transformer (ViT-B/16)	30.2	400	0.057	86.60	17.60
Proposed model (PatchTST)	18.6	280	0.034	15.10	3.55

Table 5. Statistical analysis of the proposed PatchTST model-based glaucoma detection with baseline models

Compared Model	Mean Diff	SD Diff	Standard Error	95% CI	t-value	Degrees of Freedom	p-value (t-test)	W Statistic	W p-value	Holm Threshold	P-Holm	Cohen's D
Proposed vs MobileViT	0.228	0.206064	0.065163	[0.080591 - 0.375409]	3.498916	9	0.006735	3	0.012515	0.05	0.006735	1.106454
Proposed vs Swin Transformer	0.63	0.142439	0.045043	[0.528105- 0.731895]	13.986576	9	2.0679e-7	0	0.005062	0.016667	6.2037e-7	4.422944
Proposed vs ViT	1.079	0.305194	0.096511	[0.860677- 1.297323]	11.180097	9	1.4037e-6	0	0.005062	0.025	2.8074e-6	3.535457
Proposed vs Transformer	1.518	0.228998	0.072415	[1.354185- 1.681815]	20.962372	9	5.9963e-9	0	0.005062	0.01	2.9981e-8	6.628884
Proposed vs CNN	2.319	0.432035	0.136622	[2.009941- 2.628059]	16.973898	9	3.8386e-8	0	0.005062	0.0125	1.5354e-7	5.367618

Table 6. Run-by-run-wise classification accuracies (%) of the proposed PatchTST framework and baseline models obtained using five-fold cross-validation on retinal fundus image datasets

Datasets	Run	Proposed	MobileViT	Swin Transformer	ViT	Transformer	CNN
DRISHTI-GS	1	98.81	98.48	98.26	98.16	97.81	96.96
	2	98.89	98.69	98.35	97.48	97.30	96.17
	3	98.70	98.47	97.97	97.29	97.12	96.38
	4	98.79	98.95	98.01	97.51	97.49	96.35
	5	98.86	98.52	98.20	97.61	97.19	95.66
	6	99.07	98.68	98.55	97.81	97.31	97.37
	7	98.89	98.55	98.22	97.90	97.45	96.75
	8	98.71	98.45	98.33	97.88	97.00	96.21
	9	98.84	98.38	97.97	97.70	97.18	96.61
	10	98.53	98.64	97.93	97.96	97.06	96.44
	Mean $\pm$ SD	98.81 $\pm$ 0.15	98.58 $\pm$ 0.18	98.18 $\pm$ 0.21	97.73 $\pm$ 0.26	97.29 $\pm$ 0.25	96.49 $\pm$ 0.45
RIM-ONE	1	98.96	98.71	98.49	98.32	98.01	97.18
	2	99.02	98.86	98.61	98.40	98.15	97.42
	3	98.91	98.65	98.37	98.19	97.88	97.09
	4	99.01	98.93	98.46	98.43	98.12	97.36
	5	99.05	98.74	98.58	98.55	97.96	96.95
	6	99.14	98.88	98.72	98.63	98.18	97.61
	7	99.00	98.79	98.53	98.51	98.09	97.31
	8	98.94	98.69	98.57	98.46	97.92	97.12
	9	98.99	98.63	98.41	98.39	97.99	97.24
	10	98.98	98.81	98.44	98.58	97.94	97.27
	Mean $\pm$ SD	99.00 $\pm$ 0.07	98.77 $\pm$ 0.10	98.52 $\pm$ 0.11	98.45 $\pm$ 0.13	98.02 $\pm$ 0.11	97.26 $\pm$ 0.19

4.3.1. Computational Complexity and Statistical Analysis

Computational complexity analysis refers to the evaluation of the efficiency of the proposed PatchTST model for glaucoma detection in terms of time and space complexity. Table 4 compares the computational complexity of the proposed framework with that of representative CNN and Transformer-based models. All statistical computations are performed using Python 3.9, scipy.stats (v1.11.2), and statsmodels (v0.14.0).

Table 3 provides a comprehensive statistical summary of the 10 cross-validation runs. The proposed PatchTST model contains 15.10 million parameters, exceeding conventional CNN architectures because PatchTST introduces learnable embedding, multi-head attention, feedforward, and normalization layers. The parameter count remains lower than that of the Vision Transformer (ViT) and Swin Transformer because attention is performed on compact DenseNet-derived feature embeddings instead of the full image token sequence.

The proposed framework requires 3.55 Giga Floating-Point Operations Per Second (GFLOPs), reflecting the additional computations associated with patch embedding and self-attention modules. The FLOPs are lower than those of the ViT and Swin Transformer because the attention mechanism processes fewer feature tokens and does not require repeated global or hierarchical window-based attention operations. This design improves contextual representation while maintaining computational requirements that are suitable for retinal fundus image classification.

Table 5 presents the statistical comparison between the proposed PatchTST framework and the baseline models across the 10-fold cross-validation. The proposed framework consistently achieved a positive mean performance compared with the existing methods used or compared.

The corresponding 95% confidence intervals excluded zero, indicating consistent improvements across the evaluation folds. The paired student's t-test and the Wilcoxon signed-rank test yield statistically significant results ($p < 0.05$)

for all pairwise comparisons. Additionally, the Holm-adjusted p-values remain below the respective significance thresholds, confirming that the observed differences remain statistically significant after multiple-comparison correction. Furthermore, the large Cohen's d values indicate substantial practical effect sizes, demonstrating that the performance improvements of the proposed framework are both statistically reliable and practically meaningful for glaucoma detection.

Table 6 presents the run-by-run classification accuracies obtained using 10 runs on the DRISHTI-GS and RIM-ONE datasets. The proposed PatchTST framework maintains consistently high classification accuracy across all folds, with only minor performance variations, indicating stable learning and reliable generalization.

The limited fold-to-fold variation suggests that the proposed framework is robust to different data partitions and maintains a consistent glaucoma classification performance on both benchmark datasets.

4.3.2. Cross-Validation Results

In this study, cross-validation is performed to estimate the performance of the proposed glaucoma detection model using the DRISHTI-GS and RIM-ONE datasets. Cross-validation is performed by splitting the dataset into several subsets, known as folds, and testing the proposed model on different portions of the data. Table 7 shows the cross-fold validation of the proposed framework with representative CNN and Transformer-based models.

The proposed PatchTST method achieved better results because of its ability to effectively capture both local and global temporal patterns using an optimized patch-based representation. The cross-validation results of the PatchTST method with $k=5$ folds of data offer a balanced trade-off between training stability and generalization, which leads to improved consistency in performance. Additionally, the robust architecture of the model enhances the precision and recall across multiple datasets.

Table 7. Cross-validation performance results of the proposed PatchTST-based glaucoma detection model

K-folds	Datasets	Precision (%)	Recall (%)	Accuracy (%)	F1-score (%)
k=2	DHRISHTI-GS	98.20	98.24	98.23	98.21
k=3		98.44	98.45	98.46	98.43
k=5		98.83	98.85	98.84	98.82
k=7		98.74	98.75	98.76	98.73
k=9		98.60	98.61	98.62	98.59
k=2	RIM-ONE	98.10	98.11	98.12	98.09
k=3		98.36	98.37	98.38	98.35
k=5		98.97	98.96	98.98	98.97
k=7		98.84	98.85	98.86	98.83
k=9		98.68	98.69	98.70	98.67

4.3.3. Error Analysis

To further analyze the classification performance of the proposed DenseNet-169 and PatchTST frameworks, an error analysis is conducted using confusion matrices and representative prediction examples from the DRISHTI-GS and RIM-ONE datasets. Figure 8 presents the confusion matrix for the DRISHTI-GS test set (N=51). From 36 true glaucoma cases, 35 are correctly identified (TP=35, FN=1), and all 15 normal controls are correctly classified (TN=15, FP=0), attaining a test accuracy of 98.84%, precision of 100.00%, recall of 97.22%, and F1-score of 98.59%. Figure 9 illustrates the test confusion matrix on RIM-ONE (N=138). The proposed PatchTST achieves TP=60, TN=76, FP=1, FN=1, achieving an accuracy of 98.55%, precision of 98.36%,

recall of 98.36%, and F1-score of 98.36%. The results show that the proposed model correctly identifies fundus images by capturing the discriminative structural characteristics associated with glaucomatous and healthy ODs. The misclassification of fundus images is due to poor illumination, low contrast, indistinct OD boundaries, and early glaucomatous changes that closely resemble normal retinal anatomy. These characteristics increase the intra-class similarity and reduce the inter-class separability, thereby affecting the classification performance. However, the relatively small number of misclassified samples demonstrates the robustness of the proposed framework in challenging imaging conditions.

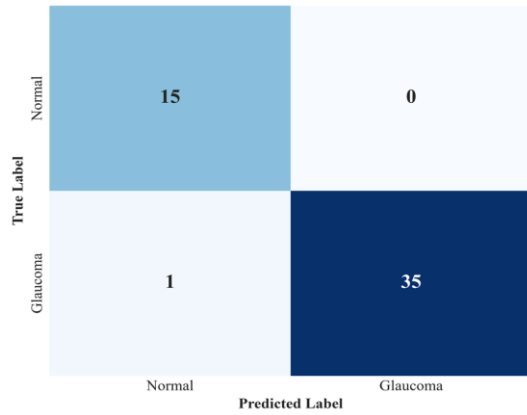


Fig. 8 Test confusion matrix for DRISHTI-GS dataset (N=51: TP=35, FN=1, TN=15, FP=0)

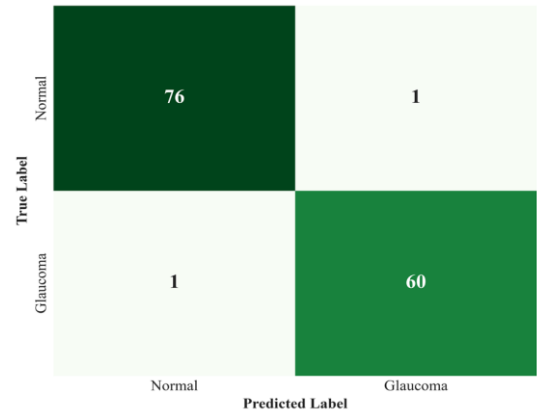


Fig. 9 Test confusion matrix for RIM-ONE dataset (N=138: TP=60, FN=1, TN=76, FP=1)

4.3.4. ROC Curve

Figure 10 displays the ROC curve for the proposed PatchTST model against the baseline architectures on DRISHTI-GS. PatchTST achieves an outstanding Area Under

the Curve (AUC) of 0.9942 (95% CI: [0.9850, 0.9985]). Figure 11 displays the ROC curve for RIM-ONE, where PatchTST reaches an AUC of 0.9958 (95% CI: [0.9890, 0.9990]).

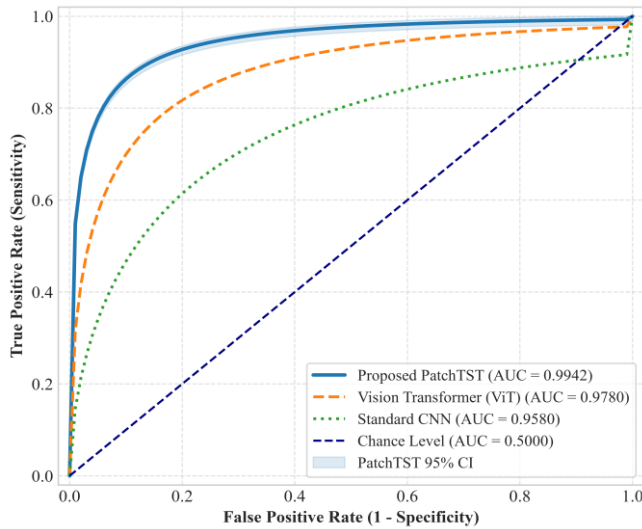


Fig. 10 ROC curves with 95% confidence bands comparing PatchTST (AUC=0.9942) against ViT and CNN on DRISHTI-GS dataset

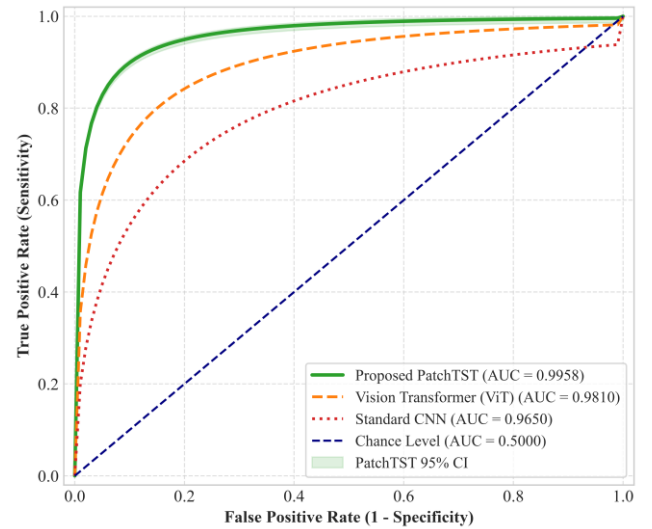


Fig. 11 ROC curves with 95% confidence bands comparing PatchTST (AUC=0.9958) against ViT and CNN on the RIM-ONE dataset

4.3.5. Training and Validation Results

The DRISHTI-GS dataset comprises 101 fundus images (71 with Glaucoma and 30 Normal), evaluated using an independent 50/51 train/test split. Figure 12 depicts the training and validation loss/accuracy curves over 50 epochs,

demonstrating a stable convergence without overfitting. The RIM-ONE dataset contains 455 fundus images (200 Glaucoma and 255 Normal), evaluated with a 317 train/138 test split. Figure 13 shows the smooth loss and accuracy curves for 50 epochs.

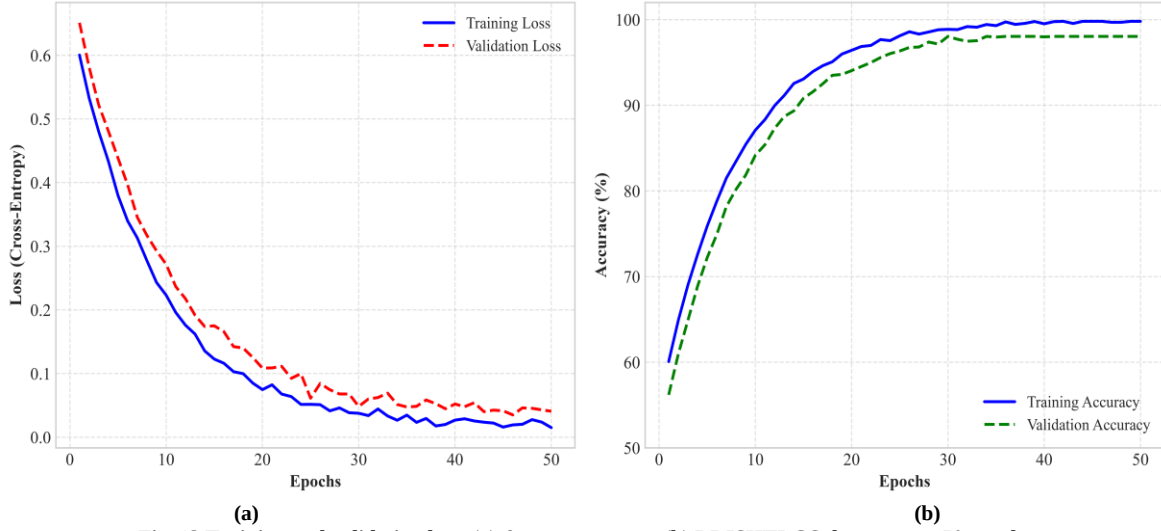


Fig. 12 Training and validation loss, (a) Accuracy curves, (b) DRISHTI-GS dataset over 50 epochs.

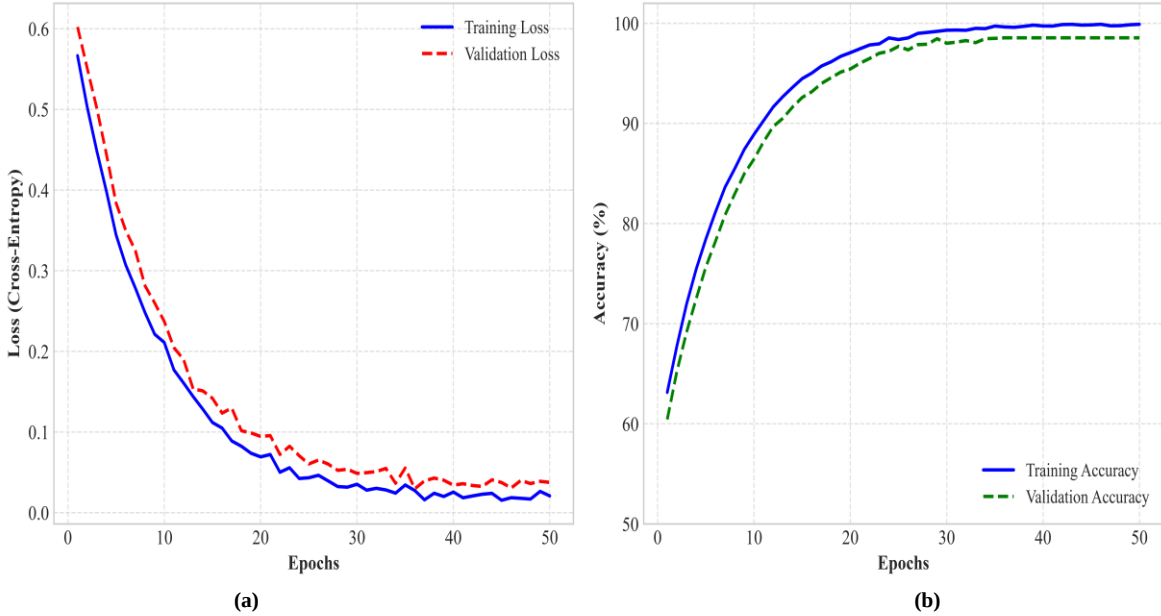


Fig. 13 Training and validation loss, (a) Accuracy curves, (b) RIM-ONE dataset over 50 epochs.

4.3.6. External Validation

The ACRIMA dataset [26] is partitioned into 493 training images (70%) and 212 independent test images (30%). Figure 14 illustrates the convergence of the training and validation losses and accuracy over 50 epochs. Figure 15 displays the confusion matrix for the external test set (N=212). The model correctly classified 117 of 119 glaucoma cases (TP=117, FN=2) and 92 of 93 normal controls (TN=92, FP=1),

achieving an accuracy of 98.58%, precision of 99.15%, recall of 98.32%, and F1-score of 98.73%.

Figure 16 shows the ROC curve for ACRIMA, which reached an AUC of 0.9965 (95% CI: [0.9910, 0.9995]), verifying that PatchTST retains exceptional diagnostic capability across varied camera hardware and clinical cohorts.

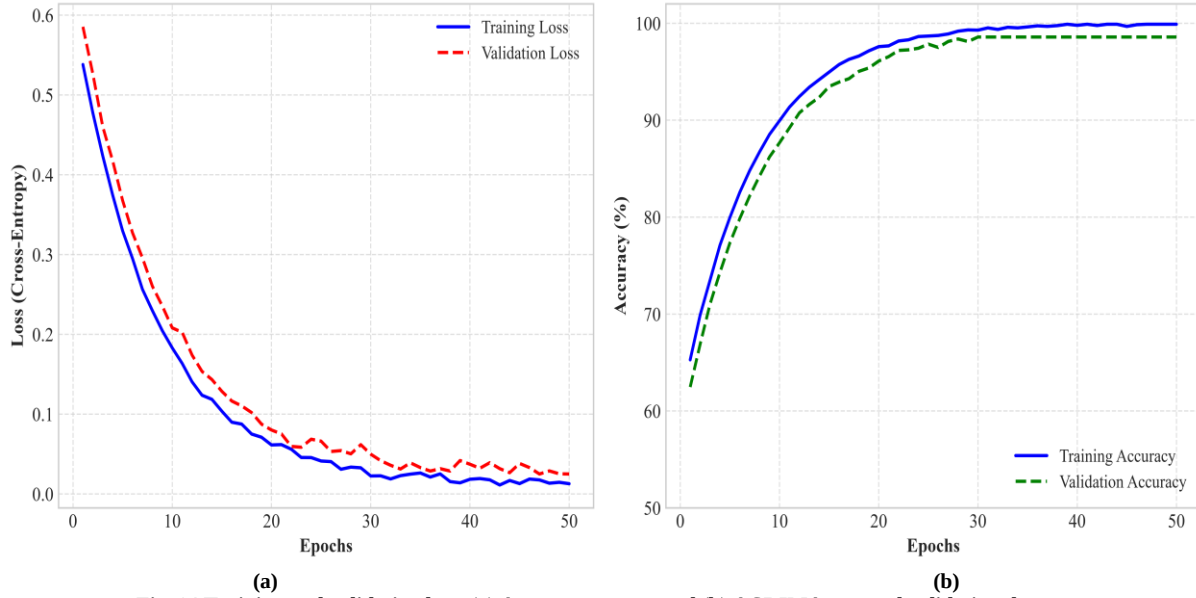


Fig. 14 Training and validation loss, (a) Accuracy curves, and (b) ACRIMA external validation dataset.

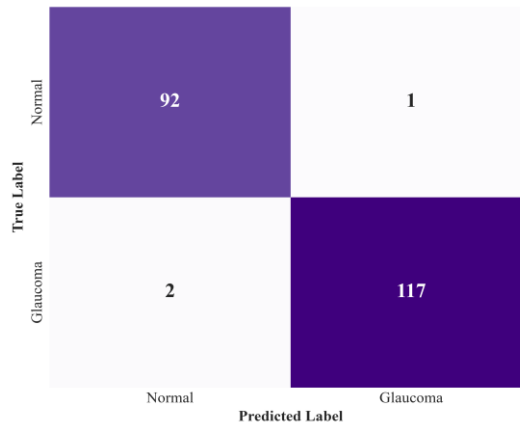


Fig. 15 Test confusion matrix on external ACRIMA dataset (N=212: TP=117, FN=2, TN=92, FP=1)

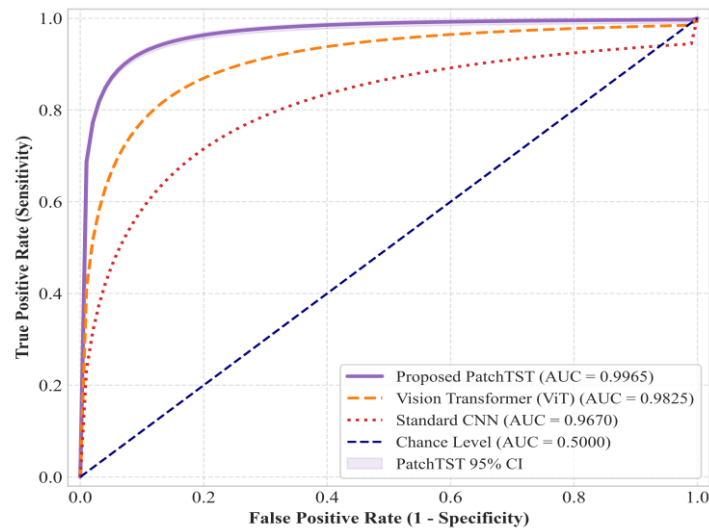


Fig. 16 ROC curves on the external ACRIMA dataset comparing PatchTST (AUC=0.9965) against ViT and CNN

4.3.7. Explainability Analysis



Fig. 17 Explainability analysis using Grad-CAM for the proposed PatchTST framework using the DRISHTI-GS and RIM-ONE datasets

Explainability analysis is the process of identifying and visualizing the image regions or features that contribute most to the prediction of the DL model. This helps verify whether the model decision is based on relevant information rather than irrelevant background patterns.

Figure 17 illustrates the Gradient-weighted Class Activation Mapping (Grad-CAM)-based explainability analysis of the proposed PatchTST framework. The visualization includes the original retinal fundus images, CLAHE-enhanced images, corresponding Grad-CAM activation maps, and representative glaucoma samples from the two benchmark datasets. The activation maps show that the model assigns the highest attention to the OD and cup regions, which are clinically relevant anatomical structures used for glaucoma assessment.

Additionally, less activation is observed in the surrounding retinal background, indicating that the classification decision is primarily influenced by diagnostically significant retinal features rather than irrelevant image regions. Similar localization patterns are observed for both datasets, demonstrating consistent feature learning across different retinal image distributions. These results improve the interpretability of the proposed framework by providing visual

evidence that the learned feature representations correspond to clinically meaningful regions associated with glaucoma diagnosis.

4.3.8. Ablation Study

An ablation study systematically evaluates the contribution of each component in the proposed model by removing or adding individual modules and measuring the resulting performance. Table 8 summarizes the performances of successive model configurations on the DRISHTI-GS and RIM-ONE datasets. The ablation study results show that DenseNet-169 provides a strong baseline by extracting discriminative retinal features, whereas incorporating the patch embedding module improves the structured representation of these features for subsequent processing.

By integrating the PatchTST encoder, the contextual interactions between ordered retinal feature patches are further enhanced, resulting in additional improvements across all evaluation metrics. The complete proposed framework achieved the highest classification performance on both datasets, indicating that the combination of feature extraction, patch embedding, contextual feature learning, and classification collectively contributes to accurate glaucoma detection.

Table 8. Ablation study conducted on the DRISHTI-GS and RIM-ONE datasets to evaluate the contribution of each component of the proposed PatchTST framework

Configuration	Dataset	Precision (%)	Recall (%)	Accuracy (%)	F1-score (%)	AUC (%)
DenseNet-169	DHRISHTI-GS	96.96	96.95	96.97	96.96	90.12
DenseNet-169 + Patch Embedding		97.37	97.35	97.36	97.36	90.45
DenseNet-169 + TST		97.64	97.63	97.65	97.64	90.74
Proposed Framework (DenseNet-169+ Patch embedding+ TST)		98.71	98.65	98.84	98.68	91.26
DenseNet-169	RIM-ONE	96.87	96.88	96.89	96.88	95.91
DenseNet-169 + Patch Embedding		97.21	97.20	97.22	97.21	96.17

DenseNet-169 + TST		97.77	97.78	97.79	97.78	97.36
Proposed Framework (DenseNet-169+ Patch embedding+ TST)		98.97	98.96	98.98	98.97	98.92

4.4. Comparative Analysis

Table 9. Comparative analysis of the PatchTST model with existing glaucoma detection models using the DRISHTI-GS dataset

Detection models	Reimplementation status	Performance Metrics				
		Precision (%)	Recall (%)	Accuracy (%)	F1-score (%)	AUC
AU-ResNet-50 [17]	Reimplemented	96.91	96.91	96.92	96.91	89.41
Ensemble DL [19]	Reimplemented	97.13	97.12	97.14	97.13	90.13
SDNN [20]	Reimplemented	97.43	97.41	97.42	97.42	90.37
CapsNet-ELM [21]	Reimplemented	97.78	97.77	97.79	97.78	90.58
Modified ResUNet [22]	Reimplemented	96.83	96.82	96.81	96.83	88.94
Proposed PatchTST model	On same protocol	98.84	98.83	98.84	98.82	91.26

A comparative analysis of the PatchTST model in glaucoma detection utilizing two distinct datasets, namely, DRISHTI GS-1 and RIM-ONE benchmark datasets are presented in this section.

Tables 9 and 10 present a comparative study of the PatchTST model with conventional detection algorithms

based on the two benchmark datasets. AU-ResNet-50 [17], Ensemble DL [19], SDNN [20], CapsNet-ELM [21] and Modified ResUNet [22] models are the existing detection frameworks used for comparative analysis with the PatchTST model. For comparative analysis, all the existing glaucoma detection models are evaluated under an identical experimental protocol, which is adopted for the proposed model.

Table 10. Comparative results of the PatchTST-based glaucoma classification model and existing baselines on the RIM-ONE dataset

Detection models	Reimplementation status	Performance Metrics				
		Accuracy (%)	Recall (%)	Precision (%)	F1-score (%)	AUC
AU-ResNet-50 [17]	Reimplemented	96.78	96.77	96.76	96.76	95.25
Ensemble DL [19]	Reimplemented	97.59	97.57	97.58	97.58	96.49
Modified ResUNet [22]	Reimplemented	97.64	96.63	96.65	96.64	96.93
Proposed PatchTST model	On same protocol	98.98	98.97	98.96	98.97	98.92

4.5. Discussion

The proposed PatchTST framework achieved improved glaucoma detection performance by integrating DenseNet-169 feature extraction with patch-based transformer-learning. DenseNet-169 extracted discriminative retinal representations, whereas the PatchTST model learned localized retinal characteristics and long-range contextual relationships from the extracted feature patches. This integration enabled an accurate differentiation between normal and glaucomatous fundus images, despite variations in the OD morphology and image quality. The proposed model achieved classification accuracies of 98.84% and 98.98% on the DRISHTI-GS and RIM-ONE datasets, respectively, which is higher than the performance of the compared traditional methods.

The AU-ResNet-50 model [17] relies on attention-guided feature learning, whereas the proposed PatchTST framework preserves both local retinal characteristics and contextual feature relationships through patch-based representation

learning. The ensemble DL model [19] combines multiple deep learning models, resulting in increased computational complexity, whereas the proposed framework provides a competitive classification performance using a unified architecture.

The SDNN model [20] depends on OD localization before classification, whereas the proposed approach directly learns discriminative retinal features from DenseNet-169 feature representations. The integration of efficient feature reuse with patch-based contextual learning improved feature representation, classification accuracy, and computational efficiency across benchmark retinal fundus datasets.

5. Conclusion

This research proposed a PatchTST-based framework designed to identify glaucoma by accurately differentiating between healthy and diseased retinal fundus images. To solve the limitations of conventional DL models struggling with OD

variations, raw images from the DRISHTI-GS and RIM-ONE datasets are initially preprocessed using standard image processing. Subsequently, a pretrained DenseNet-169 architecture extracts deep feature representations from the dense blocks to provide deep features for enhancing classification. Finally, the PatchTST classifier is utilized with patch learning to capture fine-grained structural details, for adapting to changing patterns and identifying the subtle disease differences between normal and glaucomatous eyes. Experimental results demonstrate that the PatchTST framework achieves classification accuracies of 98.84% on DRISHTI-GS and 98.98% on RIM-ONE datasets, while requiring less computational execution time than existing detection models. Future work will focus on integrating automated segmentation methods with advanced architectures

to isolate the OD region and to further enhance the classification of glaucoma stages.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

Funding Statement

The authors declare that no specific funding was received to support this research work.

Acknowledgments

Author 1, Author 2 and Author 3 contributed equally to this work.

References

- [1] Javad Zarean et al., "A Framework for Robust Glaucoma Detection: A Confidence-Aware Deep Uncertainty Quantification Approach with a Comprehensive Assessment for Enhanced Clinical Decision-Making," *Engineering Applications of Artificial Intelligence*, vol. 139, pp. 1-16, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [2] I. Govindharaj et al., "Capsule Network-based Deep Learning for Early and Accurate Diabetic Retinopathy Detection," *International Ophthalmology*, vol. 45, no. 1, pp. 1-17, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [3] Rawa'a Humam Aziz, and Lamia Abed Noor Muhammed, "Glaucoma Disease Diagnosis-based Deep Learning Network," *Al-Furat Journal of Innovations in Electronics and Computer Engineering*, vol. 4, no. 1, pp. 51-59, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [4] Bindu Priya Makala, and D. Manoj Kumar, "An Efficient Glaucoma Prediction and Classification Integrating Retinal Fundus Images and Clinical Data using DnCNN with Machine Learning Algorithms," *Results in Engineering*, vol. 25, pp. 1-15, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [5] Sangeetha Jayaraman, D. Rekha, and Priyanka Muruganandham, "A Residual Network Integrated with Multimodal Fundus Features for Automatic Glaucoma Classification," *Computers and Electrical Engineering*, vol. 122, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [6] Santosh Kumar Sharma et al., "An Evolutionary Supply Chain Management Service Model based on Deep Learning Features for Automated Glaucoma Detection using Fundus Images," *Engineering Applications of Artificial Intelligence*, vol. 128, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [7] Dipankar Das et al., "CDAM-Net: Channel Shuffle Dual Attention based Multi-Scale CNN for Efficient Glaucoma Detection using Fundus Images," *Engineering Applications of Artificial Intelligence*, vol. 133, pp. 1-12, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [8] H.N. Veena et al., "An Enhanced RNN-LSTM Model for Fundus Image Classification to Diagnose Glaucoma," *SN Computer Science*, vol. 5, no. 5, pp. 1-17, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [9] Shuting Chen et al., "GluSeg-Net: Retinal Fundus Medical Image Automatic Segmentation with Multi-Task Learning for Glaucoma Early Screening," *IEEE Access*, vol. 12, pp. 159982-159994, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [10] Law Kumar Singh et al., "Nature-Inspired Computing and Machine Learning based Classification Approach for Glaucoma in Retinal Fundus Images," *Multimedia Tools and Applications*, vol. 82, no. 27, pp. 42851-42899, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [11] Vutukuru Venkata Naga Satish Kumar, G. Harinath Reddy, and Mahendhra Nanjappa Giri Prasad, "A Novel Glaucoma Detection Model using Unet++-based Segmentation and Resnet with GRU-based Optimized Deep Learning," *Biomedical Signal Processing and Control*, vol. 86, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [12] I. Govindharaj et al., "Enhanced Diabetic Retinopathy Detection using U-Shaped Network and Capsule Network-Driven Deep Learning," *MethodsX*, vol. 14, pp. 1-13, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [13] Hafeez Alani Agboola, and Jesuloluwa Emmanuel Zaccheus, "Wavelet Image Scattering based Glaucoma Detection," *BMC Biomedical Engineering*, vol. 5, no. 1, pp. 1-11, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [14] Zhijiang Wan et al., "GDNet-EEG: An Attention-Aware Deep Neural Network based on Group Depth-Wise Convolution for SSVEP Stimulation Frequency Recognition," *Frontiers in Neuroscience*, vol. 17, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [15] Neha Sengar et al., "EyeDeep-Net: A Multi-Class Diagnosis of Retinal Diseases using Deep Neural Network," *Neural Computing and Applications*, vol. 35, no. 14, pp. 10551-10571, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [16] Sandra Virbukaitė, Jolita Bernatavičienė, and Daiva Imbrasienė, "Glaucoma Identification using Convolutional Neural Networks Ensemble for Optic Disc and Cup Segmentation," *IEEE Access*, vol. 12, pp. 82720-82729, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]

- [17] Thisara Shyamalee, and Dulani Meedeniya, “Glaucoma Detection with Retinal Fundus Images using Segmentation and Classification,” *Machine Intelligence Research*, vol. 19, no. 6, pp. 563-580, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [18] Vijaya Kumar Velpula et al., “Glaucoma Detection with Explainable AI using Convolutional Neural Networks based Feature Extraction and Machine Learning Classifiers,” *IET Image Processing*, vol. 18, no. 13, pp. 3827-3853, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [19] K.J. Subha, R. Rajavel, and B. Paulchamy, “An Improved Ensemble Deep Learning Framework for Glaucoma Detection,” *Multimedia Tools and Applications*, vol. 84, no. 22, pp. 25309-25323, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [20] Kamesh Sonti, and Ravindra Dhuli, “Diagnosis of Glaucoma from Retinal Fundus Images using Disc Localization and Sequential DNN Model,” *International Journal of Imaging Systems and Technology*, vol. 34, no. 1, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [21] B.P. Pradeep Kumar, Pramod K.B. Rangaiah, and Robin Augustine, “Enhanced Glaucoma Detection using U-Net and U-Net+ Architectures using Deep Learning Techniques,” *Photodiagnosis and Photodynamic Therapy*, vol. 54, pp. 1-19, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [22] Rabia Pannu et al., “Enhanced Glaucoma Classification through Advanced Segmentation by Integrating Cup-to-Disc Ratio and Neuro-Retinal Rim Features,” *Computerized Medical Imaging and Graphics*, vol. 123, pp. 1-22, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [23] Lokesh Saipureddi, Drishti-GS-Retina Dataset for ONH Segmentation, Kaggle, 2021. [Online]. Available: <https://www.kaggle.com/datasets/lokeshaipureddi/drishtigs-retina-dataset-for-onh-segmentation>
- [24] Lucas Cunha De Carvalho, Rim-One-R2, Kaggle, 2021. [Online]. Available: <https://www.kaggle.com/datasets/lucascunhadecarvalho/rimoner2>
- [25] Krishna Santosh Naidana, Madhu Hasitha Manne, and Hema Yalavarthi, “A Deep Retinal Vision Network for Glaucoma Classification,” *Discover Applied Sciences*, vol. 7, no. 3, pp. 1-22, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [26] Orvile, ACRIMA: Glaucoma Assessment using Fundus Images, Kaggle, 2025. [Online]. Available: <https://www.kaggle.com/datasets/orvile/acrima-glaucoma-assessment-using-fundus-images>