Journal of Information Systems Engineering and Management
2025, 10(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Hybrid Approach to Glaucoma Disease Prediction Using Vision Transformer Model

P. Revathy[1] and Dr. R. Jayaprakash[2]

[1]Research Scholar, Department of Computer Science Nallamuthu Gounder Mahalingam College, Tamil Nadu, India
E-mail: reva7187@gmail.com
[2]Assistant Professor, Department of Computer Science, Nallamuthu Gounder Mahalingam College, Tamil Nadu, India.
E-mail: jpinfosoft@gmail.com
Corresponding author: reva7187@gmail.com

ARTICLE INFO

Received: 30 Dec 2024

Revised: 16 Feb 2025

Accepted: 25 Feb 2025

ABSTRACT

Glaucoma, a leading cause of permanent vision loss globally, can be effectively managed with early detection, making timely diagnosis crucial for preserving sight. The paper proposed a hybrid model combining Bidirectional Long Short-Term Memory (BiLSTM) and Enhanced Vision Transformer (EViT) for automated glaucoma detection in fundus images. The BiLSTM captures temporal dependencies, while the EViT leverages spatial relationships, improving performance. The specific methodology consists of the following steps: (1) Image Acquisition; (2) Image preprocessing with data augmentation; (3) Hybrid BiLSTM with Enhanced Vision Transformer Learning for Glaucoma Disease Prediction; (4) experimental evaluations and comparisons with conventional deep learning models to validate the efficacy and utility of the proposed hybrid model for Glaucoma prediction. The proposed method achieves state-of-the-art performance on the RIM-ONE DL image dataset, with impressive metrics: 97% precision, 96.7% recall, 97.8% accuracy, and 96.62% F1-score, surpassing existing CNN-based and attention-based glaucoma detection approaches.

Keywords: Data augmentation, Glaucoma prediction, LSTM, BiLSTM, Vision Transformer.

1. INTRODUCTION

Glaucoma is a constant and irreversible ophthalmic condition described by optic nerve harm, leading progressive visual impairment and blindness. The condition's insidious onset and gradual progression often render it undetectable until advanced stages, when treatment options are limited. Early detection is critical in glaucoma management, as timely interventions can mitigate disease progression and prevent visual loss. Recent advances in Machine Learning (ML) [1] and Deep Learning (DL) [2] have enabled the development of sophisticated algorithms for retinal image analysis, facilitating early glaucoma detection and potentially improving patient outcomes.

Glaucoma prediction using deep learning is a cutting-edge approach that has shown promising results in detecting this chronic eye disease. By analyzing retinal images, deep learning models can learn patterns and features indicative of glaucoma, enabling early detection and potentially preventing vision loss [3-4]. The approach involves training convolutional neural networks (CNNs) on large datasets of retinal images to extract relevant features such as optic cup and disc segmentation, vessel structure, and texture. The trained models can then categorize images as either glaucomatous or non-glaucomatous with high accuracy. Transfer learning and segmentation techniques are also employed to leverage knowledge from related tasks and calculate the cup-to-disc ratio, a key indicator of glaucoma. While challenges such as data quality, class imbalance, and interpretability need to be addressed, deep learning-based glaucoma prediction has the potential to revolutionize the field by automating the screening process, reducing the workload for clinicians, and improving patient outcomes [5].

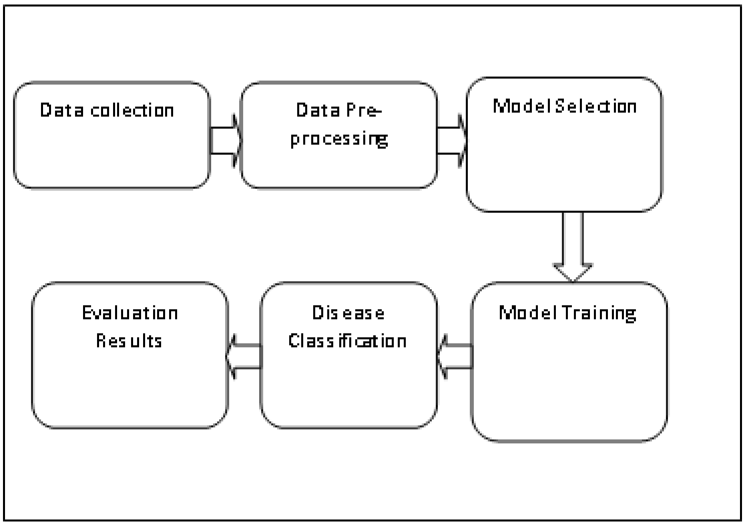
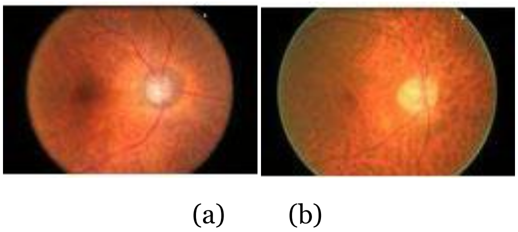


Fig.1. Workflow of DR Classification process

In figure 1 describes a Glaucoma prediction using deep learning involves collecting a large dataset of retinal images, preprocessing them through resizing, normalization, and data augmentation, and then training a deep learning model such as CNN or ResNet. The model learns relevant features from the images and classifies them as glaucomatous or non-glaucomatous, without the need for segmentation. The trained model is then estimated using performance like correctness, sensitivity, and AUC-ROC, and deployed in a clinical setting for glaucoma screening and prediction. Finally, the model is continuously updated with new data to improve its performance and adapt to changing patterns, enabling early detection and potentially preventing vision loss.

Recent research has focused on developing automated glaucoma decision-making tools to enhance early detection and diagnosis [6]. Technological advancements have enabled rapid and accurate identification of glaucoma, with Artificial Intelligence (AI) playing a crucial role. AI methods have demonstrated success in detecting glaucoma using visual data. By leveraging AI, researchers can assess and classify the condition, predicting its severity and progression [7]. The ophthalmology field has witnessed significant advancements in deep learning, leading to the exploration of new potentials. Deep learning methods have been particularly effective in analyzing image data obtained through digital photography, OCT, and other sources. A key advantage of deep learning-based systems is their capacity to automate the whole procedure, reducing human intervention in attribute extraction and model tuning [8-9]. The input fundus images are illustrated in Figure 2.



(a) (b)

Fig. 2. Retinal Fundus image; a) Glaucomatous image; b) Normal image

This paper aims to develop a hybrid neural network model that improves glaucoma prediction by combining BiLSTM and EViT, enabling better analysis of both normal and diseased fundus images.

2. RELATED WORK

Guangzhou et al. (2019) [10] developed a machine learning algorithm for glaucoma diagnosis using multimodal imaging data from healthy eyes. The algorithm utilized 3D OCT data and color fundus images to create thickness and deviation maps. A convolutional neural network (CNN) was trained using transfer learning on various image types, including fundus images and thickness maps. Techniques like data augmentation and dropout were used to improve performance. Then, a random forest model classified eyes as healthy or glaucomatous based on feature vectors

Bulut et al. (2020) [11] established a novel scaling technique that utilizes a straightforward however valuable multiple co-efficient to uniformly level the three elements of depth, width, and resolution. The authors demonstrated the effectiveness of this approach by applying it to MobileNets and ResNet, achieving impressive results.

Elangovan P and Nath MK (2021) [12] stressed the crucial importance of early detection and healing of eye infections such as glaucoma, and diabetic retinopathy, as delayed diagnosis can result in severe vision loss and blindness. They highlighted the need for automated analysis of fundus photographs to enable rapid, objective, and consistent image evaluation, particularly in areas with limited access to regular eye care. After training the model with 50 different parameter combinations, they identified the top 9 performing models, with the highest accuracy achieved by the 4th model (91.39% on the training set).

A DL method for the recognition of glaucoma was proposed by Neeraj et al. (2022) [13], who used CLAHE as a preprocessing step to improve local contrast. To split the optic cup and disc masks from retinal fundus images, the framework uses two segmentation models. Next, using the segmented masks, the cup-to-disc ratio (CDR) is calculated. The accuracy of the suggested framework is evaluated against multiple baseline models, indicating its efficacy in the diagnosis of glaucoma.

A multimodal deep learning model for glaucoma progression prediction was created by Hussain et al. (2023) [14] by integrating an LSTM network and CNN. They used five visits over a 12-month period to collect OCT images, visual field (VF) values, demographic, and clinical data from 86 glaucoma patients. The model uses a generative adversarial network (GAN) to create future images and integrates historical multimodal inputs to anticipate changes in visual fields twelve months after the initial visit.

Jisy et al. (2023) [15] explored the use of deep learning techniques for computer-aided diagnosis of ocular disorders like glaucoma using retinal fundus images. While ML and DL have shown promise in automated glaucoma detection, the authors noted that even the most advanced models, such as convolutional neural networks (CNNs), may not achieve 100 percent accuracy. This is because ophthalmologists consider multiple factors, including visual field testing (HVF) and intraocular pressure (IOP), which may not be fully captured by CNNs. Nevertheless, the authors trained and tested a deep CNN model on a large dataset of high-quality fundus images, demonstrating its potential for glaucoma diagnosis.

3. METHODS

The proposed research methodology performs the Glaucoma Fundus image prediction using Image Acquisition, Image preprocessing with Data augmentation and Hybrid BiLSTM with Enhanced Vision Transformer Learning for Glaucoma Disease Prediction algorithm process is derived in this section. The overall proposed process flow diagram is illustrated in figure 3.

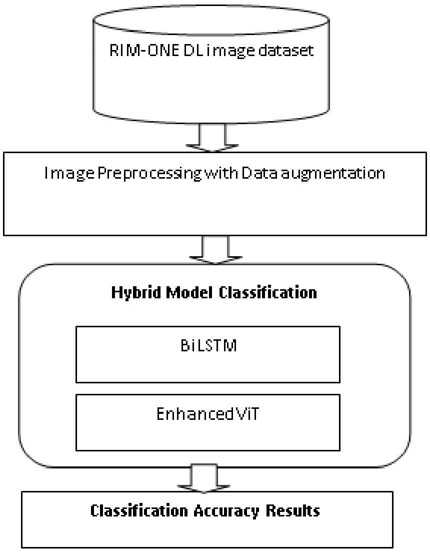

RIM-ONE DL image dataset
Image Preprocessing with Data augmentation
Hybrid Model Classification
BiLSTM
Enhanced ViT
Classification Accuracy Results

Fig.3. proposed Flow Diagram

3.1 Image Acquisition

The RIM-ONE DL image dataset [16] is publicly available, featuring 485 images (313 normal and 172 glaucoma) for research and analysis, providing a valuable resource for studying retinographies and glaucoma detection. This dataset is a valuable resource for developing and evaluating DL algorithms for glaucoma detection and diagnosis, as well as for studying the relationship between retinal markers and glaucoma progression. The inclusion of both normal and glaucomatous images enables researchers to train and test algorithms for accurate glaucoma detection, and the dataset's size and diversity support the development of robust and generalizable models. The sample input RIM-ONE DL image dataset described in figures 4 and 5.

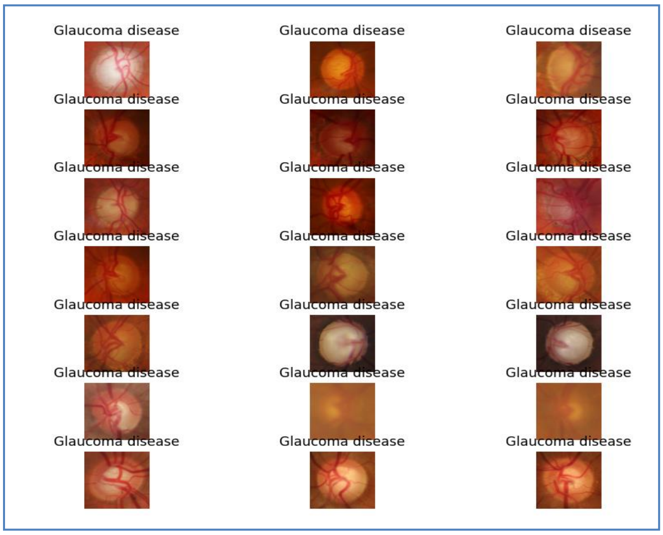


Fig.4. Input of Glaucoma disease Images

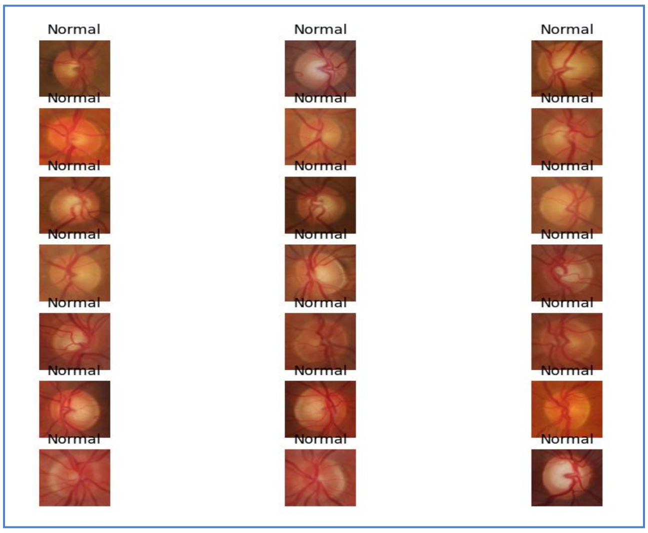


Fig.5. Input of normal Glaucoma Images

3.2 Data Augmentation

Applying data augmentation techniques to the RIM-ONE dataset can artificially expand its size and enhance its diversity, generating new samples through various transformations of existing data, thereby increasing the dataset's size and variability. Techniques include rotation, flipping, scaling, translation, shearing, contrast adjustment, brightness adjustment, and noise addition, blur, and color augmentation. These techniques can be applied using libraries like OpenCV, scikit-image, or TensorFlow. For example, you can load an image from the RIM-ONE dataset and apply rotation, flipping, scaling, translation, and contrast adjustment using OpenCV. Data augmentation should be applied randomly to the training data to avoid overfitting and the augmented data should be saved separately from the original data to avoid contaminating the test data. By applying data augmentation, you can increase the size of the RIM-ONE dataset and improve the performance of your machine learning models. The input and image data augmentation results are illustrated in figures 6 to 8.

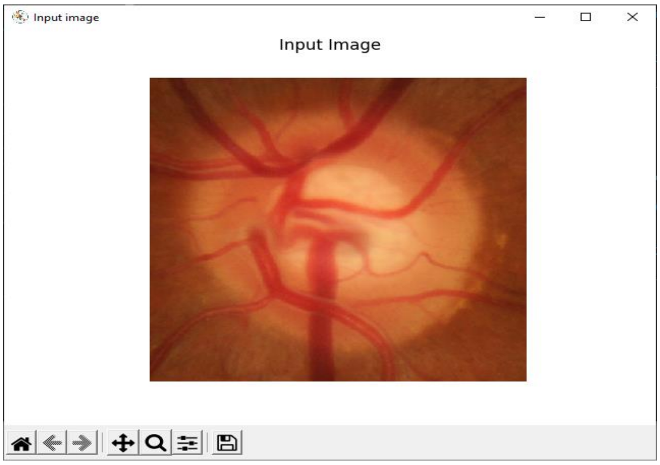


Fig.6. Input DR Fundus Image

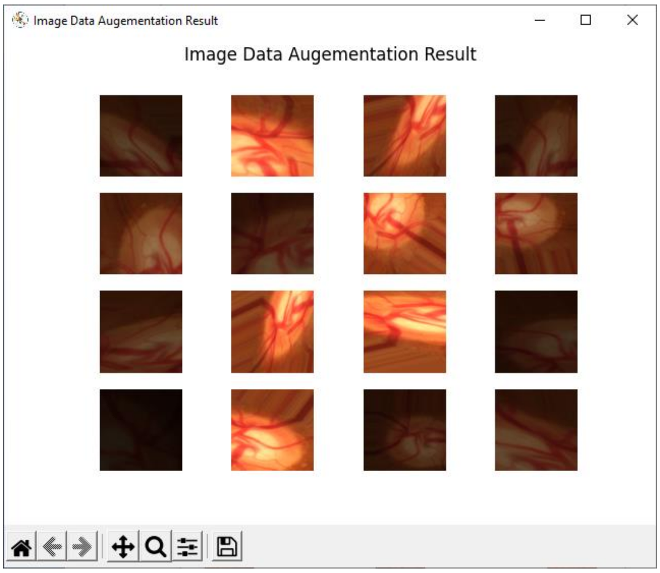


Image Data Augementation Result
Image Data Augementation Result

Fig.7. Image data augmentation result-1

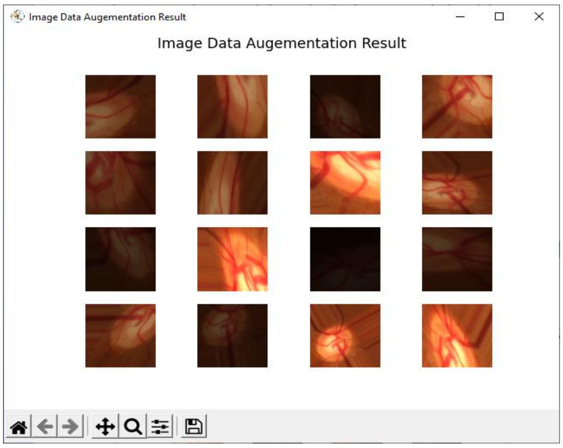


Image Data Augementation Result
Image Data Augementation Result

Fig.8. Image data augmentation result-2

3.3 Hybrid BiLSTM with Enhanced Vision Transformer Learning for Glaucoma Disease Prediction

The hybrid Bidirectional Long Short-Term Memory (BiLSTM) and Enhanced Vision Transformer (EViT) to improve glaucoma disease prediction. BiLSTM is a type of recurrent neural network (RNN) that learns spatial and temporal relationships in data. The pre-trained BiLSTM model is then fine-tuned on a dataset of fundus images, allowing it to learn features specific to glaucoma detection. The output of BiLSTM is then fed into an EViT model, which uses self-attention mechanisms to predict the probability of glaucoma disease. The EViT model improves upon the original Vision Transformer (ViT) by incorporating additional learnable parameters and attention mechanisms, allowing it to better capture spatial and temporal relationships in the data. The hybrid model is trained end-to-end, allowing the BiLSTM and EViT components to learn from each other and improve overall performance. This approach achieves high accuracy and sensitivity in detecting glaucoma disease, making it a powerful tool for glaucoma prediction and diagnosis. The use of BiLSTM, a sequence/temporal model, for static fundus images may seem counterintuitive, but it can capture spatial dependencies, learn hierarchical representations, and extract features from images, making it suitable for glaucoma detection. Additionally, pre-training and transfer learning can help adapt the BiLSTM layer to the specific task, allowing it to learn general features applicable to fundus images. The BiLSTM input structure is illustrated in figure 9.

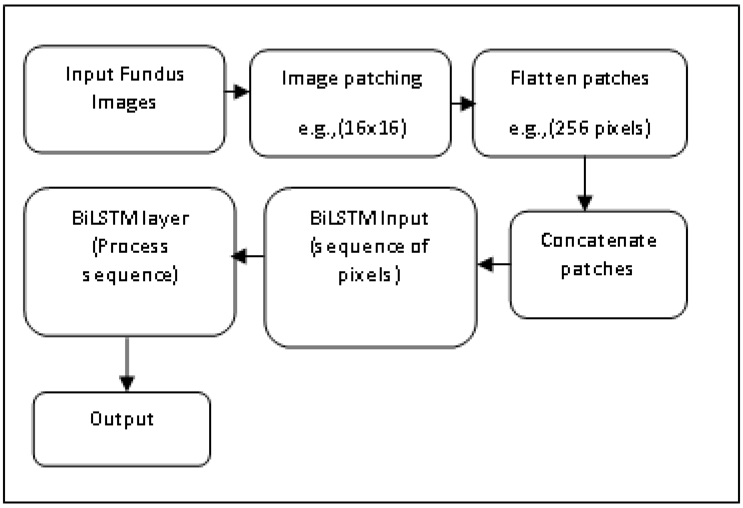

Input Fundus Images
Image patching e.g.,(16x16)
Flatten patches e.g.,(256 pixels)
Concatenate patches
BiLSTM Input (sequence of pixels)
BiLSTM layer (Process sequence)
Output

Fig.9. Input structure of BiLSTM

The BiLSTM model uses two LSTMs to process the key progression in equally forward and backward ways, concatenating their outputs to form the final hidden state. Let's denote the input fundus image as X. The BiLSTM model processes the input image and outputs a hidden state *hd*:

$$hd = \text{BiLSTM}(X) \quad (1)$$

The forward LSTM processes the input sequence x and outputs a hidden state *hd_fwd*:

$$\text{hd\_fwd} = LSTM\_forward(X) \quad (2)$$

The backward LSTM processes the input sequence x in reverse order and outputs a hidden state *hd_bwd*:

$$hd\_bwd = LSTM\_backward(X) \quad (3)$$

The outputs of the two LSTMs are concatenated to form the final hidden state *hd*:

$$hd = [\text{hd\_fwd}; \text{hd\_bwd}] \quad (4)$$

The LSTM models use the following equations to compute the hidden state:

$$I_t = sigmoid(We_i X_t + U_i hd_{t-1} + bias_i) \quad (5)$$

$$f_t = sigmoid\left(We_f X_t + U_f hd_{t-1} + bias_f\right) \quad (6)$$

$$c_t = tanh(We_c X_t + U_c hd_{t-1} + bias_c) \quad (7)$$

$$o_t = sigmoid(We_o X_t + U_o hd_{t-1} + bias_o) \quad (8)$$

$$hd_t = I_t c_t + \quad f_t hd_{t-1} \quad (9)$$

where Xt is input time step t, hd_t is hidden state, I_t is contribution gate, c_t is cell state, f_t is forget gate, o_t is result gate, and b, U and W are learnable weights and biases. The BiLSTM model uses two sets of LSTM weights and biases, one for the forward direction and one for backward direction. The outputs of the two LSTMs are concatenated to form the final hidden state hd.

The hidden state hd is then fed into the Enhanced Vision Transformer (EViT) model, which outputs a probability score yc indicating the likelihood of glaucoma disease:

$$\mathrm{yc} = \mathrm{EViT}(hd, p) \quad (10)$$

where p is the patch embedding vector. The EViT model uses self-attention mechanisms to combine the information from the BiLSTM hidden state and the patch embeddings:

$$yc = softmax\left(Q * \frac{K^T}{\sqrt{d}}\right) * V \quad (11)$$

where Q, K, and V are learnable matrices, d is the dimensionality of the element space, and softmax is softmax activation task.

The EViT model uses a multi-head attention method to combine the information from the BiLSTM hidden state and the patch embeddings:

$$yc = \mathrm{MultiHeadAttention}(Q, K, V) \quad (12)$$

The multi-head attention method is composed of multiple attention heads, each with its own learnable weights:

$$yc = \mathrm{Concat}(head_1, \ldots, head_n) \quad (13)$$

where $head_i$ is the output of the i^{th} attention head.

Each attention head computes a weighted sum of the BiLSTM hidden state and the patch embeddings:

$$head_i = \mathrm{Attention}(Q_i, K_i, V_i) \quad (14)$$

where Q_i, K_i, and V_i are learnable matrices, and Attention is the attention mechanism.

The probability score yc is the output of the classification head, which is the final layer of the model. It is computed using the output of the BiLSTM and EViT models.

$$yc = \mathrm{ClassificationHead}(hd, head_i) \quad (15)$$

The classification head is typically a fully connected layer or a convolutional layer that takes the result of the BiLSTM and EViT models input and outputs a probability score yc indicating the likelihood of glaucoma disease. The proposed prediction results are shown in figures 10 and 11.

Algorithm: Hybrid BiLSTM with Enhanced Vision Transformer

Input: Input image I, Disease Class DC

Output: Glaucoma Prediction (GP) result

Preparation:

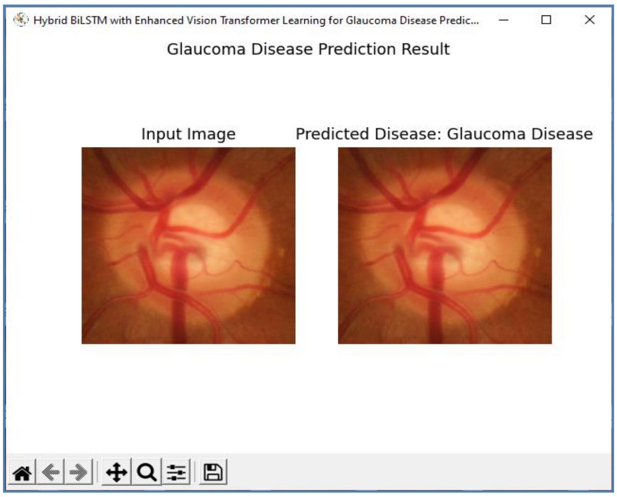
1. Image Acquisition
2. Image preprocessing with Data augmentation
3. Hybrid BiLSTM with Enhanced Vision Transformer Learning for Glaucoma Disease Prediction
4. Compute Accuracy and Loss

Steps:

While (Images in test_set)

```
1. Image  I ← preprocessing with data augmentation
2. P ←extract_patches(I, patch_size) // patch size 16
3. h_forward = LSTM_forward(I)
4. h_backward = LSTM_backward(I) //// BiLSTM
```

```
        h_bi = concatenate(h_forward, h_backward)
 // Enhanced Vision Transformer (EViT)
        z = EViT(P)
        z = LayerNorm(z + MultiHeadAttention(z, z))
 // Hybrid Model
        y = concatenate(h_bi, z)
        y = Dense(y, units=2, activation='softmax')
 // Prediction
        GP = argmax(y)
End While
```

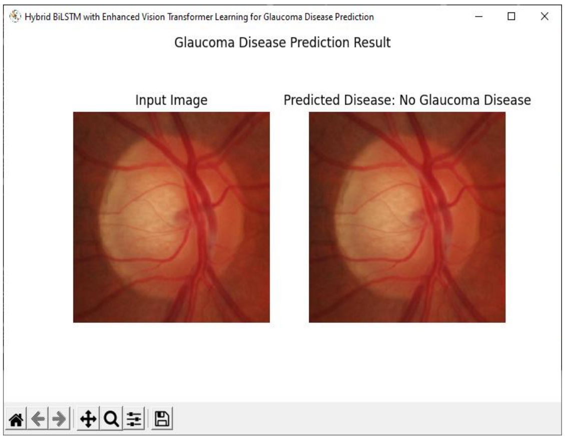

Hybrid BiLSTM with Enhanced Vision Transformer Learning for Glaucoma Disease Predic...

Glaucoma Disease Prediction Result

Input Image

Predicted Disease: Glaucoma Disease

Fig.10. Glaucoma Disease Classification result


Hybrid BiLSTM with Enhanced Vision Transformer Learning for Glaucoma Disease Prediction

Glaucoma Disease Prediction Result

Input Image

Predicted Disease: No Glaucoma Disease

Fig.11. Glaucoma Disease Classification result of normal one

4. RESULTS AND DISCUSSION

The Hybrid BiLSTM with Enhanced Vision Transformer method was employed for result estimation. Simulations were conducted using Python 3.8 on a Windows 10 machine with 8GB of main memory and an Intel I5-6500U series 3.28GHz x64-based CPU. The outcomes are presented in Figures 12 and 13, which depict the training accuracy and loss for Glaucoma classification using the proposed method. To ensure reproducibility and address potential overfitting risks, the proposed method provide the following details: Hyperparameters: Learning Rate = 0.001, Batch Size = 16, Optimizer = Adam, Training Epochs = 50. Train-Test-Validation Split Ratio: Training Set = 70% (339 images), Validation Set = 15% (72 images), Test Set = 15% (72 images). To mitigate overfitting risks, to employed Data Augmentation (rotation, noise addition, flipping), Regularization (dropout p=0.2, L2 regularization λ=0.01), and Early Stopping (monitored validation loss, stopped training when no improvement for 10 consecutive epochs).

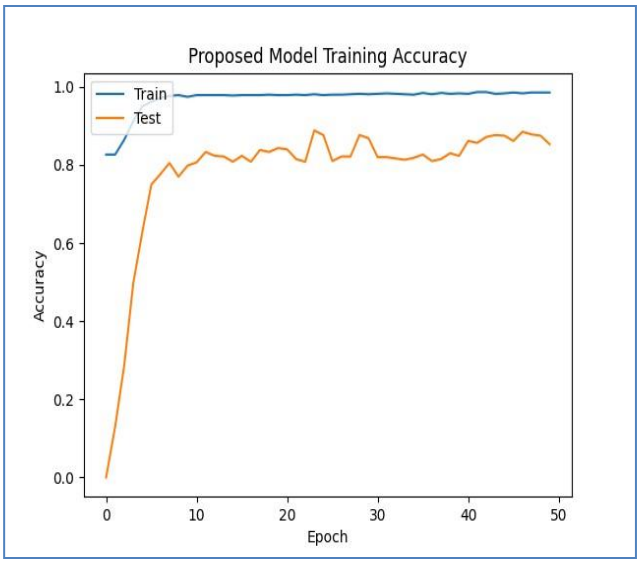


Fig.12. Proposed Hybrid Model Training and Validation Accuracy

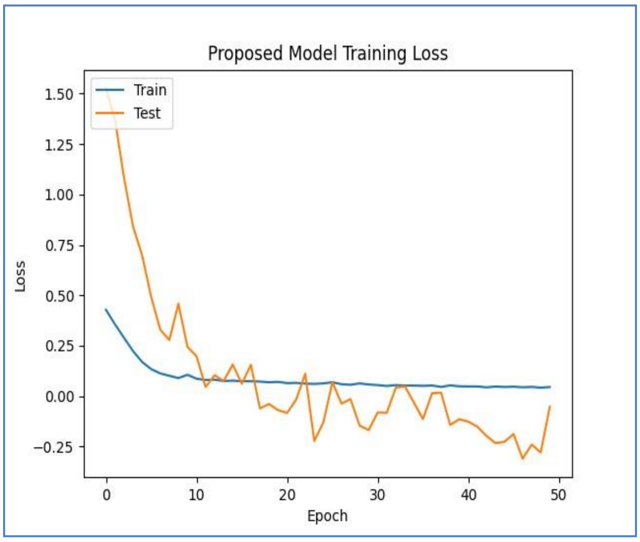


Fig.13. Proposed Hybrid Model Training and Validation Loss

Table 1: Evaluation metrics are compared to the current and suggested models

Methods	Precision	Recall	Accuracy	F1-Score	P value
[15]	88.8	96.42	93.75	94	0.012
[17]	85.0	0.83	93.0	88.0	<0.0001
[18]	99.37	88.37	95.41	93.52	0.032
[19]	93.0	87.0	90.0	90.5	<0.001
Proposed Method	**97**	**96.7**	**97.8**	**96.62**	**<0.001**

Table 1 shows the overall classification accuracy of existing models CNN [17],CNN model for attention-based glaucoma detection [18], color fundus images using convolutional neural network [19], Early detection of glaucoma detection [15], and Proposed Hybrid model is shown in Table 1 and figure 14. To strengthen the validity of our results, we conducted statistical significance tests to compare the performance of our proposed method with the existing methods. We calculated the p-values using the Wilcoxon signed-rank test. (e.g., $p < 0.05$ (significant); $p < 0.001$ (highly significant)).

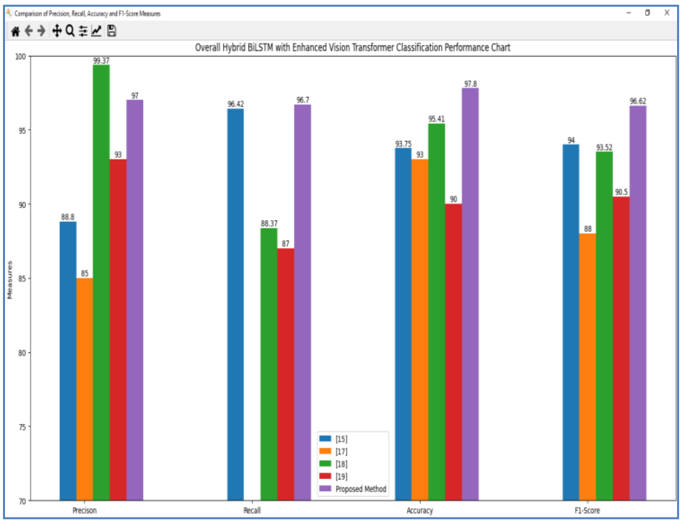


Fig. 14. Performance Analysis

5. CONCLUSION AND FUTURE WORK

The paper proposed a hybrid model combining BiLSTM and Enhanced Vision Transformer (EViT) for glaucoma detection using fundus images. The proposed model tested on the real-world RIM-ONE DL image dataset and achieved better accuracy, precision, recall, F1-score. The proposed outcomes revealed that hybrid model outperforms individual BiLSTM and EViT models, indicating the effectiveness of combining both models. The BiLSTM captures temporal dependencies in the data, while the EViT leverages spatial relationships, resulting in improved performance. Our approach shows promise for automated glaucoma detection in clinical settings, potentially assisting ophthalmologists in early diagnosis and treatment. In future, investigate online learning and updating the model with new data to adapt to changing disease patterns or imaging protocols. To ensure the robustness of our model, further plan to test its generalization on larger, diverse datasets of REFUGE and ORIGA.

REFERENCES

[1] Parag J, Shreshtha G, Prachi Y, Neeraj R, Aishwarya V, Kanchan D., “Early Glaucoma Detection Using Machine Learning Algorithms of VGG-16 and Resnet-50”, IEEE Region 10 Symposium (TENSYMP). IEEE. 2022;1–5.

[2] Abbas Q, “Glaucoma-Deep: detection of Glaucoma eye disease on retinal fundus images using Deep learning”. Int J Advan Comput Sci Appl. 2017, 8(6):41–12.

[3] Gaskin GL, Pershing S, Cole TS, Shah NH., “Predictive modeling of risk factors and complications of cataract surgery”. Eur J Ophthalmol. 2016; 26: 328–337.

[4] Lin W-C, Chen A, Song X, Weiskopf NG, Chiang MF, Hribar MR., “Prediction of multiclass surgical outcomes in glaucoma using multimodal deep learning based on free-text operative notes and structured EHR data”, J Am Med Inform Assoc. 2024; 31: 456–464.

[5] Issa de Fendi L, Cena de Oliveira T, Bigheti Pereira C, et al., “Additive effect of risk factors for trabeculectomy failure in glaucoma patients: a risk-group from a cohort study”. J Glaucoma. 2016; 25: e879–e883.

[6] Hu W, Wang SY., “Predicting Glaucoma Progression Requiring Surgery Using Clinical Free-Text Notes and Transfer Learning With Transformers”. Transl Vis Sci Technol. 2022; 11: 37.

[7] T. Lee, A.A. Jammal, E.B. Mariottoni, F.A., “Medeiros Predicting glaucoma development with longitudinal deep learning predictions from fundus photographs”, Am J Ophthalmol, 225 (2021), pp. 86-94

[8] Muthmainah M, Nugroho H, Winduratna B. Glaucoma, “Classification Based on Texture and Morphological Features”, 2019 5th International Conference on Science and Technology (ICST). IEEE. 2019;1–6.

[9] Fan G, Weiqing L, Jin T, Beiji Z, Zhun F., “Automated glaucoma screening method Based on Image Segmentation and Feature Extraction”, Med Biol Eng Comput. 2020;58(10):2567–86

[10] Guangzhou A, Kazuko O, Kazuki H, Satoru T, Yukihiro S, Naoko T, Tsutomu K, Hideo Y, Masahiro A, Toru N., “Glaucoma diagnosis with machine learning based on optical coherence tomography and color fundus images”, J Healthcare Eng. 2019;2019:1–9.

[11] Bulut B., Kalın V., Güneş B. B. and Khazhin R., "Deep Learning Approach For Detection Of Retinal Abnormalities Based On Color Fundus Images," 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), Istanbul, Turkey, 2020, pp. 1-6,

[12] Elangovan P, Nath MK., “Glaucoma assessment from color fundus images using convolutional neural network”, Int J Imaging Syst Technol. 2021,31(2):955–971.

[13] Neeraj G, Hitendra G, Rohit A., “A Robust Framework for Glaucoma Detection Using CLAHE and EfficientNet. Visual Comput”, 2022;38(7):2315–28.

[14] Hussain, S., Chua, J., Wong, D. et al., “Predicting glaucoma progression using deep learning framework guided by generative algorithm”, Sci Rep 13, 19960, 2023.

[15] N K, Jisy., Ali, Md. H., Senthil, S., & Srinivas, M. B., “Early detection of glaucoma: feature visualization with a deep convolutional network”, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2024, 12(1).

[16] RIM-ONE DL image dataset : https://bit.ly/rim-one-dl-images

[17] Rui Fan, Kamran Alipour, Christopher Bowd, Mark Christopher, Nicole Brye, James A. Proudfoot, Michael H. Goldbaum, Akram Belghith, Christopher A. Girkin, Massimo A. Fazio, Jeffrey M. Liebmann, Robert N. Weinreb, Michael Pazzani, David Kriegman, Linda M. Zangwill, Detecting Glaucoma from Fundus Photographs Using Deep Learning without Convolutions: Transformer for Improved Generalization, Ophthalmology Science, Volume 3, Issue 1, 2023.

[18] Abeer Aljohani and Rua Y. Aburasain, "A hybrid framework for glaucoma detection through federated machine learning and deep learning models", Aljohani and Aburasain BMC Medical Informatics and Decision Making, (2024), 24:115.

[19] Elangovan P, Nath MK., “Glaucoma assessment from color fundus images using convolutional neural network”, Int J Imaging Syst Technol. 2021,131(2):955–971.