DeepRetinaNet: An Automated AI-Based Framework for Retinal Disease Diagnosis

Akshya Kumar Sahoo, Priyadarsan Parida, Manoj Kumar Panda, Chittaranjan Nayak, *Member, IEEE,* and N. Mohankumar, *Senior Member, IEEE*

***Abstract*—Automated retinal disease diagnosis leveraging cutting-edge computer vision methodologies supports clinicians in the early identification of pathological conditions. This investigation delivers a novel framework, DeepRetinaNet for automating retinal disease diagnosis. The developed DeepRetinaNet model has two stages of novelties, including vessel extraction and disease identification. In the vessel extraction stage, the green channel, known for its heightened sensitivity to retinal vascular structures, is extracted from the source images. Subsequently, the vessel extraction network: RetiSegNet, processes these green channel images to extract retinal vessels, generating binary vessel maps. During the fusion phase, the original fundus images are combined with the extracted vessel maps to produce fused representations, encapsulating enriched spatial details from both sources. In the identification stage, these fused images are utilized to train the proposed classification framework: STDeepNet, which incorporates Modified Identity (MI), Modified Convolution (MCONV) blocks, and Long Short-Term Memory (LSTM) layers to effectively identify the diseases. The efficacy of the developed technique is corroborated using visual illustration and objective analysis. Also, the efficiency of the designed framework is verified on six benchmark datasets. The proposed framework demonstrates superior performance compared to 49 state-of-the-art methods, achieving notable accuracy in retinal disease diagnosis.**

Link to graphical and video abstracts, and to code: https://latamt.ieeer9.org/index.php/transactions/article/view/9534

***Index Terms*—Retinal image, LSTM, feature fusion, diabetic retinopathy, Glaucoma.**

I. Introduction

RETINAL diseases, including Glaucoma, Diabetic Retinopathy (DR), and Age-Related Macular Degeneration (AMD), are among the causes of vision loss and blindness globally [1]. Early detection is crucial to prevent irreversible damage, but traditional diagnostic methods often rely on specialized equipment and expertise that may not be available at the time of need. Artificial intelligence can ease it using computer vision-assisted approaches. In the last decade, numerous approaches including both supervised and unsupervised approaches have been suggested by various researchers using fundus images to detect various ocular diseases. The major hindrances in using these approaches for automated disease diagnosis from fundus images include image acquisition with diverse appearances, complex branching patterns of vessels with varying diameters, obscure vessel boundaries, and subtle changes in retinal anatomy, making the detection process more challenging. Moreover, in the fundus image, the foreground information (*in this case it is the vascular structure*) compared to the background is more sparse, which poses an important aspect to consider while developing a comprehensive assisted technology [2]. This in turn provides a class imbalance issue, which leads to inefficient feature extraction and an inability to capture fine-grained detailed structure precisely.

To address these challenges, an innovative deep-learning framework entitled DeepRetinaNet was introduced to automate retinal disease detection through fundus images. The primary objective of the proposed algorithm is to extract the blood vessels from the fundus images with enriched vessel features. Further, the extracted vessels are utilized for the disease classification task with reduced misclassification. To attain it, a DeepRetinaNet framework is proposed that utilizes the Retina Segmentation Network (RetiSegNet) model for vessel extraction and the Spatio-Temporal Deep Neural Network (STDeepNet) model for different disease classifications. Here, the RetiSegNet model consists of a parallel stream Convolutional Neural Network (CNN)-induced feature fusion mechanism to produce the vessel masks with reduced background exudates. Again, the proposed STDeepNet comprises custom CNN with various LSTM blocks to precisely predict the various diseases in the fundus images.

DeepRetinaNet begins by extracting intricate blood vessel structures from these images, vital for spotting early disease indicators. By examining vascular patterns, the system can identify abnormalities related to various conditions, such as optic nerve damage that suggests Glaucoma, microvascular irregularities associated with Diabetic Retinopathy, and the blood vessel changes characteristic of Age-Related Macular Degeneration. Once these blood vessel structures are extracted and analyzed, DeepRetinaNet categorizes each image into one of several groups: normal, Glaucoma, DR, or AMD. This automated approach not only boosts diagnostic accuracy but

The associate editor coordinating the review of this manuscript and approving it for publication was Anabel Martin (*Corresponding author: Manoj Kumar Panda*).

A. K. Sahoo is with the Department of Electrical and Electronics Engineering, GIET University, Gunupur, Odisha, India (e-mail: akshyasahoo@giet.edu).

P. Parida, and Manoj Kumar Panda are with the Department of Electronics and Communication Engineering, GIET University, Gunupur, Odisha, India (e-mails: priyadarsanparida@giet.edu, and manojkumarpanda@giet.edu).

C. Nayak is with the Department of Communication Engineering, School of Electronics Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India (e-mail: 83chittaranjan@gmail.com).

N. Mohankumar is with the Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Pune, India (e-mail: mohankumar.n@sitnagpur.siu.edu.in).

also enhances access to early detection, especially in areas with limited resources.

The major contributions of this work are as follows:

1) A unique DeepRetinaNet framework is proposed to simultaneously extract in-depth features from fundus images, capable of retaining the vessels along with the disease category identification.
2) The developed DeepRetinaNet framework comprises RetiSegNet and STDeepNet networks for vessel extraction and disease categorization, respectively.
3) The designed RetiSegNet framework provides better performance for unseen data setup.
4) For disease classification, the proposed STDeepNet is capable of retaining subtle details and the usage of skip connections makes the model more robust.
5) In the STDeepNet network, a multi-stage feature fusion mechanism is proposed to further enhance the disease identification performance.

The rest of the article is organized as: The current state-of-the-art (SOTA) approaches developed so far are discussed in Section II. The designed network and its description are provided in Section III. Network performance via empirical analysis is done in Section IV. Various ablation study of the developed network is done in Section V. Finally, the article is concluded with future directions in Section VI.

II. State-of-the-Art-Techniques

This section discusses various SOTA methods related to retinal image diagnosis. The methods are divided into three parts: extraction of retinal blood vessels, classification of retinal diseases, and simultaneous extraction as well as classification of retinal diseases.

Retinal vessel segmentation is crucial in ophthalmology, as it facilitates the early detection and diagnosis of ocular diseases such as Glaucoma, AMD, and DR. However, this task is challenging due to the complexity of retinal images, which contain variations in vessel width, tortuosity, and contrast. Additionally, artifacts, low-resolution areas, and overlapping anatomical structures further complicate accurate segmentation, demanding advanced algorithms to ensure precise extraction of vascular features [3]. Deep neural networks (DNNs) demonstrate significant potential in addressing the limitations of traditional methods for visual feature extraction from fundus images [4].

Recent advancements, such as architectures employing dilated convolutional filters, have significantly improved retinal vessel segmentation [5]. An automated algorithm employing a trainable filtering mechanism and vessel-linking procedures effectively detected fine vascular structures and extracted relevant clinical features [6]. The authors [7] introduced the learnable ophthalmology Segment Anything Model (SAM), a learnable prompt layer for ophthalmology multi-modal image segmentation. A multidimensional DNN model employing cross-dimensional transformations and self-attention mechanisms for the segmentation of vessels from fundus images [8]. An enhanced U-Net architecture [9] is utilized in segmenting complex, low-contrast blood vessels. The authors in [10] implemented the dual-channel asymmetric CNN in the U-Net model for retinal vessel segmentation by capturing the global and fine-grained features.

Intelligent systems utilize DNN and machine learning algorithms for the classification of fundus images, significantly improving the detection of glaucomatous retinal changes. These advanced models enhance the ability to accurately distinguish between healthy and diseased retinal images by learning complex patterns and features relevant to Glaucoma, leading to improved diagnostic accuracy and early intervention [11]. The authors [12] utilized a hybrid graph convolutional network integrated with a self-attention mechanism and an ensemble of machine learning classifiers for the multi-grade classification of retinal disorders. ActiveLearn, a deep learning framework based on ActiveLearn Transformer architecture [13] integrates transfer learning techniques to enhance the low-level features from the fundus images for retinal disease classification. In [14] the authors presented an automated, DNN-based non-invasive framework EyeDeep-Net for the diagnosis of multiple eye diseases using color fundus images. The authors in [15] introduced an automated algorithm for early Glaucoma detection by extracting critical diagnostic parameters. Authors in [16] developed a multiple improved Inception-v4 ensemble model for detecting DR and diabetic macular edema. However, vessel extraction along with disease diagnosis has its own merits over the existing literature.

Hence, recent studies have combined segmentation and classification tasks to enhance fundus image diagnosis. A hybrid deep learning approach is utilized, incorporating an Enhanced Fuzzy C-Means clustering scheme for the segmentation of retinal vessels. Further, a hybridized model using DenseNet and ShuffleNet is used to perform the classification task [17]. In [18] a Dagum probability distribution function integrated with a matched filter is introduced for retinal blood vessel segmentation. This approach combines various extracted features and utilizes an ensemble classifier for the classification task. The authors [19] introduced VisionDeep-AI, a segmentation and classification framework designed for retinal vessel extraction with multi-class classification from fundus images. The framework [20] utilized DNN-based de-noising and Wiener filtering followed by blood vessel extraction using Otsu thresholding and mathematical morphology. Later, modified ResNet 101 is utilized for the feature extraction and classification of retinal images.

From the above literature, it is found that various SOTA methods are developed for simultaneous extraction of vessels followed by disease classification. However, these techniques are incapable of extracting subtle details from the fundus images which may degrade the performance of the network and may not generalize to unseen setups due to diverse variations in disease categories. Therefore, in this article, the DeepRetinaNet framework was introduced for the effective identification of retinal diseases.

III. Proposed Method

The process of retinal image analysis is carried out in two stages shown in Fig. 1: in the first stage, the retinal

vessels are extracted using RetiSegNet and in the second stage, the extracted vessels are classified into various classes using STDeepNet. Initially, a pre-processing strategy is adopted where the green channels are extracted from the input fundus images indicated in Step-1. Later, the green channels along with the ground truth masks are utilized to train the proposed RetiSegNet framework that provides the extracted vessel masks denoted as Step-2. Further, these masks are combined with the green channel fundus images depicted in Step-3 which are utilized as testing data for STDeepNet framework. Again, the green channel fundus image and their corresponding ground truths are combined to generate the training data for STDeepNet denoted as Step-4. In Step-5, the testing data obtained from Step-4 are then applied to the STDeepNet framework for classification. The detailed architecture of a unified network representing both segmentation and classification tasks is shown in Fig.2. The green channel of the fundus image is selected as it provides better vessel contrast, captures sharper details, reduces noise, and enhances the performance of the developed network. These factors make the green channel ideal for more accurate and reliable retinal blood vessel extraction.

A. Proposed RetiSegNet model

The detailed architecture of the RetiSegNet model is shown in Fig.2 (a) consisting of 2 parallel networks. Here, the Residual Quattro Network (Residual QuattroNet) model and Atrous Encoder-Decoder Network (AEDNet) model are used in parallel for the RetiSegNet model. The use of a parallel stream strategy makes the proposed RetiSegNet model computationally efficient. Again, the parallel stream network can extract significant details that are highly essential for vessel extraction. The input image (I_m) is passed through these networks during the segmentation stage: the local feature (thick vessels) extraction using AEDNet and Residual QuattroNet for global feature (thin vessels) extraction.

1) AEDNet: The proposed AEDNet network consists of three encoder blocks (EBs) (Fig. 2 (c)) and three decoder blocks (DBs) (Fig. 2 (d)-(e)) via a bridge path (Fig. 2 (i)). The three EBs are used to downsampling images from a size $224 \times 224 \times 1$ to a feature size $28 \times 28 \times 256$. Each EB consists of a convolution layer followed by a Rectified linear unit (ReLU), atrous convolution, ReLU, and max-pooling layer. The EB1 and EB2 have the same configuration, while the EB3 additionally uses a dropout layer with rate of 0.5. The dilation factor for atrous convolution is set to 2 throughout AEDNet to preserve the spatial resolution of the input features without increasing the computational cost. The number of filters selected for EB1, EB2, and EB3 are 64, 128, and 256 respectively with a filter size of 3×3. The outcomes of the EBs with dimension $28 \times 28 \times 256$ is sent to the bridge path, which sandwiches convolution, ReLU, and atrous convolution with a sampling rate of 2 followed by ReLU and dropout with a rate of 0.5.

The bridge module preserves the spatial features with a dimension of $28 \times 28 \times 512$. Similarly, the three DBs are used to increase the size of the features $28 \times 28 \times 512$ to a feature size $224 \times 224 \times 64$. The first 2 decoder blocks consist of a transposed convolution layer followed by ReLU, convolution followed by ReLU, and atrous convolution followed by ReLU. However, the third decoder block contains an atrous convolution with ReLU. The proposed decoder is capable of retaining the spatial resolution during upsampling. The number of filters selected for DB1, DB2, and DB3 are 256, 128, and 64 respectively with a uniform kernel size of 3×3.

Let e_1, e_2, e_3 be the outputs from the encoder blocks EB1, EB2, and EB3 respectively such that $e_1 \in \Re^{112\times112\times64}$, $e_2 \in \Re^{56\times56\times128}$, and $e_3 \in \Re^{28\times28\times256}$. Similarly, d_1, d_2, d_3 represent the outputs from the decoder blocks DB1, DB2, and DB3 respectively where, $d_1 \in \Re^{56\times56\times256}$, $d_2 \in \Re^{112\times112\times128}$, and $d_3 \in \Re^{224\times224\times64}$.

2) Residual QuattroNet: The Residual QuattroNet module in the RetiSegNet consists of an input convolution layer followed by batch normalization, ReLU, and max pooling layers. Later, two cascaded identity blocks (IBs) (Fig. 2 (f)) followed by six alternate convolution blocks (CBs) (Fig. 2 (h)) and identity blocks. Both the IBs and CBs use residual connections. The IB makes the network deeper and mitigates the vanishing gradient problem. Similarly, the CB is used to extract the sparse vessel features from the fundus images. Further, as the network deepens, multiple CBs and IBs are used to extract multi-scale and multi-resolution in-depth features with accurate thin vessel feature extraction.

Let x be the input to IB and $i(x)$ be the output from the set of convolution layers. Then, output from the IB is represented in equation 1:

$$y = i(x) + x \tag{1}$$

Similarly, if a is the input to a CB, $c(a)$ is the output after passing through multiple convolution layers. $W_i \times a$ representing the transformed input through skip connection using convolution and batch normalization layer. Then, the output from the CB is given in equation 2:

$$o = c(a) + (W_i \times a) \tag{2}$$

Let the outputs from IB1, IB2, IB3, IB4 and IB5 are represented as y_1, y_2, y_3, y_4, and y_5 respectively. Similarly, the output from CB blocks are denoted as o_1, o_2, and o_3 respectively. The first convolution layer with 3×3 max-pooling uses 64 filters to produce a feature size of $113 \times 113 \times 64$. The feature sizes in IB1 and IB2 are not changed. CB1 changes the feature size from $113 \times 113 \times 64$ to, $57 \times 57 \times 128$ using 128 filters and a stride of 2. Similarly, IB3 does not change the input feature size. Further, CB2 changes the feature size from $57 \times 57 \times 128$ to $29 \times 29 \times 256$ using 256 filters and a stride of 2. IB4 keeps the feature map fixed at $29 \times 29 \times 256$. Furthermore, CB3 changes the feature size from $29 \times 29 \times 256$ to $29 \times 29 \times 512$ using 512 filters and an atrous convolution with dilation factor 2. Further, the IB5 block does not change the feature size.

Following the IB5 block, the quattroNet module is used, which consists of four convolution blocks each having one convolutional layer, followed by batch normalization and ReLU activation. The first convolution block in the quattroNet module performs standard convolution, whereas the remaining

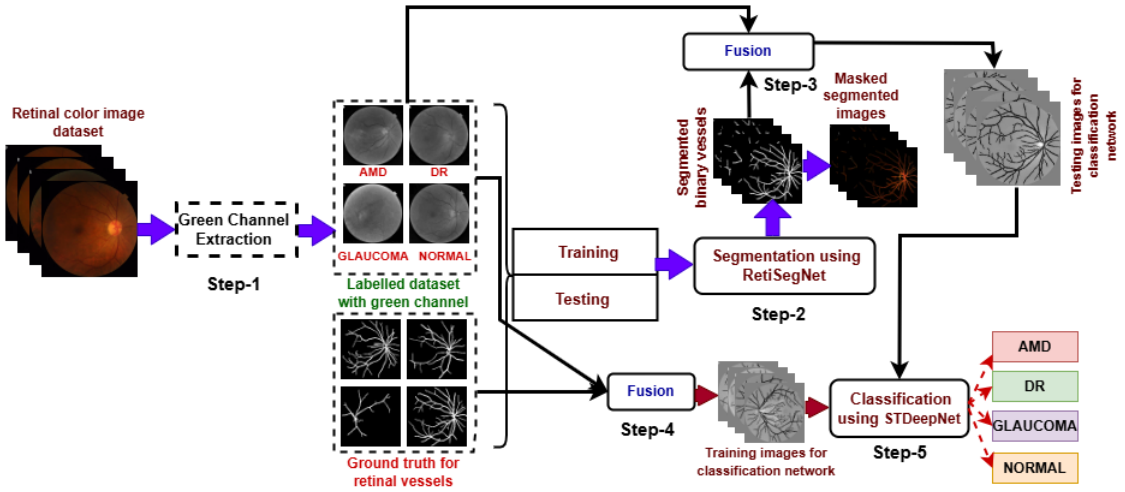

Fig. 1. Flowchart of the proposed algorithm.

three blocks use atrous convolution with dilation rates of 18, 12, and 6 respectively. It is used for capturing multi-scale contextual information like small and fine blood vessels along with vessel bifurcations. The input feature of the quattroNet module (Fig. 2 (g)) can be expressed as $y_5 \in \Re^{29\times29\times512}$. Using the standard convolution block with 3×3 filter and 256 channels, the output of the convolution layer is $K_1 \in \Re^{3\times3\times512\times256}$. Through batch normalization and ReLU activation, the output of the block is $q_1 \in \Re^{29\times29\times256}$. Similarly, the dilated convolution blocks with dilation rates of 18,12, and 6 produce outputs q_2, q_3, q_4, $\in \Re^{29\times29\times256}$. The net output from the quattroNet module is expressed in the equation 3 through depth concatenation of the outputs q_1, q_2, q_3, and q_4:

$$Q = (q_1 \otimes q_2 \otimes q_3 \otimes q_4) \in \Re^{29\times29\times1024} \tag{3}$$

Where $\otimes$ represents the depth concatenation operation. The output Q is convoluted using 256 filters to get a feature size of $29 \times 29 \times 256$. Further, a transposed convolution layer with a filter size of 8×8 and stride of 4 is used to upsample the feature size to $116 \times 116 \times 256$. Again, a cropping layer produces the output $Q_1 \in \Re^{113\times113\times256}$. On the other hand, y_2 is convoluted using 48 filters to produce output $y_{21} \in \Re^{113\times113\times48}$. Now, y_{21} is concatenated with Q_1 to produce the output Q_2 using equation 4:

$$Q_2 = (y_{21} \otimes Q_1) \in \Re^{113\times113\times304} \tag{4}$$

Further, by using a set of convolutional layers with 256 filters and 2 filters respectively, the feature size is reduced to $113 \times 113\times2$ from $113\times113\times256$. Next to the convolutional layers, a transposed convolutional layer of filter size 8×8, stride of 4, and cropping of 2 is used to find the final output of the Residual QuattroNet unit $Q_o \in \Re^{452\times452\times2}$.

Again, the output from the Residual QuattroNet module is concatenated with the output from DB3, the output from EB1, and the image input I_m by using equation 5:

$$O = (Q_o \otimes d_3 \otimes e_1 \otimes Im) \in \Re^{224\times224\times130} \tag{5}$$

The output convolutional layer with 64 filters is used to reduce the feature to $224 \times 224 \times 64$ and another convolutional layer with 2 filters is used to bring the feature size to $224 \times 224 \times 2$. Finally, a softmax classifier is used to extract the blood vessels. These 2 convolution layers followed by softmax layer is termed as Vessel Extractor (Fig. 2 (j)).

B. Proposed STDeepNet Model

The classification network developed in this article is shown in Fig. 2 (b) utilizes cascading of two modified identity (MI) blocks (Fig. 2 (k)) followed by four alternate MI blocks and modified convolution (MCONV) blocks (Fig. 2 (l)). Both the MI and MCONV blocks use instant normalization as the task. Retinal blood vessel classification is more precise and complex in terms of subtle feature variations like thin vessels against thick vessels, texture changes between healthy and diseased regions, and gradual intensity variations along the vessel paths. The instant normalization layer also provides better performance on small batches with lesser computational cost. The expression normalized feature from the instant normalization layer $\hat{n_i}$ is given in equation 6:

$$\hat{n_i} = \frac{n_i - \mu_i}{\sigma_i + \epsilon} \tag{6}$$

Where n_i is the input feature of the current sample at index i, $\mu_i = \frac{1}{HW}\sum_{h-1}^{H}\sum_{w-1}^{W} n_i$ represents the standard deviation across the spatial dimensions $H \times W$, and $\sigma_i = \frac{1}{HW}\sum_{h-1}^{H}\sum_{w-1}^{W}(n_i - \mu_i)^2$ represents the mean across the spatial dimensions $H \times W$.

Further, the network utilizes the parametric ReLU (PReLU) activation function as it improves the model flexibility by learning the negative slope, making it effective for complex feature extraction tasks. The PReLU activation function can be defined in equation 7 where p is the input and α is the learnable parameter:

$$f(p) = \begin{cases} p & p \geq 0 \\ \alpha p & p < 0 \end{cases} \tag{7}$$

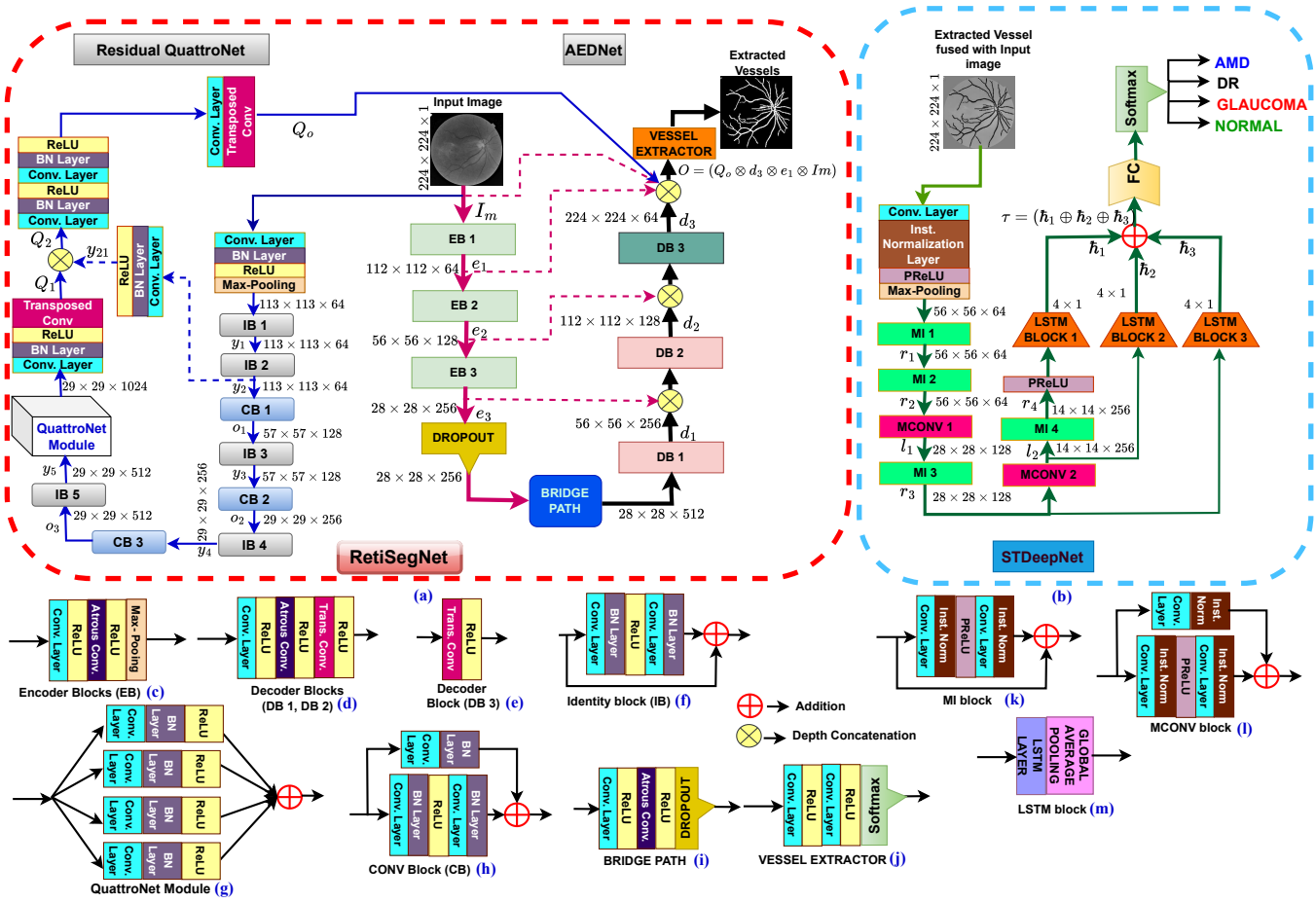

Fig. 2. Detailed framework of developed segmentation and classification model.

The MI block consists of a set of convolutional layers followed by instant normalization and PReLU activation. The output of the MI block can be expressed in the equation 8:

$$r = \eta_2(W_2 * \phi(\eta_1(W_1 * t))) + t \tag{8}$$

Where t is the input to the MI block, ϕ is the PReLU activation function, η is the instant normalization function, W_1 is the kernel size of the first convolutional layer and W_2 is the kernel size for the second convolutional layer.

The MCONV block consists of a set of convolutional layers followed by instant normalization and PReLU activation, along with the transformed input through a skip connection using a convolutional layer and instant normalization. The output of the MCONV block can be expressed in equation 9:

$$l = \eta_2(W_2 * \phi(\eta_1(W_1 * g))) + \eta_{skip}(W_{skip} * g) \tag{9}$$

where g is the input to the MCONV block and l is the output. W_{skip} is the kernel size of the convolutional layer in the skip connection path.

In the developed STDeepNet, multi-stage feature fusion is performed through LSTM block (Fig. 2 (m)) consists of LSTM layer followed by global average pooling. The LSTM layers enhance model performance by capturing sequential dependencies, integrating low-level and high-level features, and improving gradient flow. It also offers dynamic feature representation, adapting to complex transformations.

TABLE I
DATASET SUMMARY

Dataset	Number of Images	Resolution (Pixels)	Use Case
FIVES	800	2,048 × 2,048	Vessel segmentation, Multi-disease classification
STARE	20	700 × 605	Vessel segmentation, disease diagnosis
DRIVE	40	768 × 584	Vessel segmentation
CHASE-DB1	28	1,280 × 960	Vessel segmentation
RFMiD	3200	Varies (e.g., 4,288 × 2,848, 2,048 × 1,536, 2,144 × 1,424)	Multi-disease classification
ODIR	5000	Varies (width: 250 to 5,184; height: 188 to 3,456)	Disease classification

Let the outputs from MI1, MI2, MI3, and MI4 be r_1, r_2, r_3, and r_4 respectively. Similarly, the output from MCONV1 and MCONV2 are l_1 and, l_2 respectively.

The green channels of segmented vessels from RetiSegNet fused with the original fundus images having size $224 \times 224 \times 1$ are passed through the input convolutional layer of 64 filters having a kernel size of 7×7 followed by an instant normalization, a PReLU activation, and a 3×3 max-pooling layer to obtain a feature size of $56 \times 56 \times 64$. The same feature size is maintained through the blocks MI1 and MI2 using 64 filters: $r_1 \in \Re^{56\times56\times64}$ and $r_2 \in \Re^{56\times56\times64}$.

Similarly, the block MCONV1 through 128 filters of kernel size 3×3 down-samples r_2 to a feature size $l_1 \in \Re^{28\times28\times128}$. Further, the block MI3 preserves the same feature size at $r_3 \in \Re^{28\times28\times128}$. After MI3, the MCONV2 down-samples the features to $l_2 \in \Re^{14\times14\times256}$ through 256 filters of size 3×3. Further, MI4 preserves the features $r_4 \in \Re^{14\times14\times256}$. The output r_4 is passed through the LSTM layer with 4-hidden unit to produce an output $\hbar_1 \in \Re^{4\times1}$. Similarly, the output l_2 is passed through the second LSTM layer with 4-hidden unit to produce an output $\hbar_2 \in \Re^{4\times1}$. Further, the output r_3 is passed through the third LSTM layer with 4-hidden unit to produce an output $\hbar_3 \in \Re^{4\times1}$. The outputs $\hbar_1$, $\hbar_2$, and $\hbar_3$ are combined by using equation 10 to produce the output τ:

$$\tau = (\hbar_1 \oplus \hbar_2 \oplus \hbar_3) \in \Re^{4\times1} \tag{10}$$

Finally, the output τ is passed through a 4×4 fully connected layer and a softmax classifier to obtain 4 output classes: AMD, DR, GLAUCOMA, and NORMAL.

IV. Empirical Analysis

A. Datasets Utilized for Experimentation

The proposed framework is initially trained using the FIVES dataset [21]. Further, STARE[1], DRIVE[2], and CHASE-DB1 [22] were utilized for the validation of the designed RetiSegNet model. Similarly, the developed STDeepNet network utilized FIVES [21], RFMiD [23], and ODIR [24] datasets for training as well as validation. The details of the dataset utilized in this paper are given in Table I.

B. Parameter Settings

The proposed framework is designed for the vessel extraction and classification of retinal diseases. The training and testing of the model is conducted on a system with a Core *i*5 processor, 16GB RAM, and an *NVIDIA GeForce RTX 3050* GPU. MATLAB is used for the implementation, running on a Windows 11 environment. The model is trained for a maximum of 50 epochs with a batch size of 4, employing the Adaptive Moment Estimation (*ADAM*) optimizer with a learning rate of 0.0001. This setup ensures a balanced approach to training to avoid overfitting or underfitting. The *ADAM* optimizer facilitates adaptive and efficient learning by making small, controlled adjustments to the model parameters, promoting stable convergence across the epochs. This configuration ensures that the developed model learns effectively while maintaining computational efficiency.

The designed RetiSegNet framework is initially trained and tested on the FIVES dataset. Out of all retinal images, 66.66 % are used for training, while the remaining 33.33 % are allocated for testing. The test set includes the FIVES dataset (Seen Setup) and STARE, DRIVE as well as CHASE-DB1 (Unseen Setup). After the segmentation stage, the segmented blood vessels from the segmented mask are combined with the corresponding input retinal images. These fused images are then classified using a designed STDeepNet network. This strategy ensures robust feature extraction by leveraging original and segmented data for disease classification. The developed STDeepNet classification network is evaluated only for the seen setup where the samples are split into 66.66% and 33.33% for training and testing respectively.

[1]https://cecas.clemson.edu/ ahoover/stare/
[2]http://www.isi.uu.nl/Research/Databases/DRIVE/

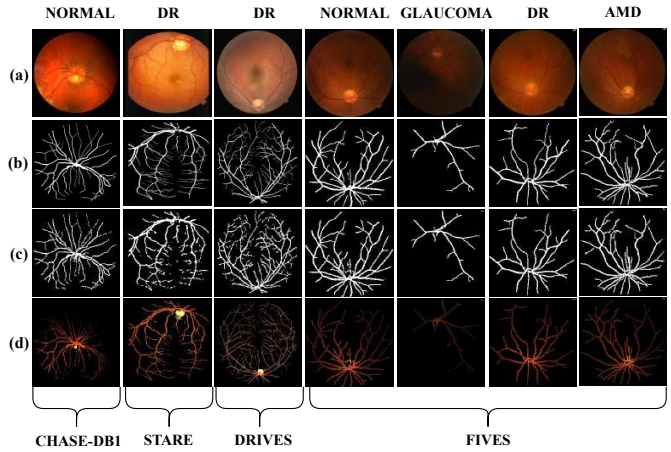

Fig. 3. Samples of vessel extraction outcomes for the proposed RetiSegNet model (a) Original image, (b) Ground truth, (c) Extracted vessels, (d) Masked image.

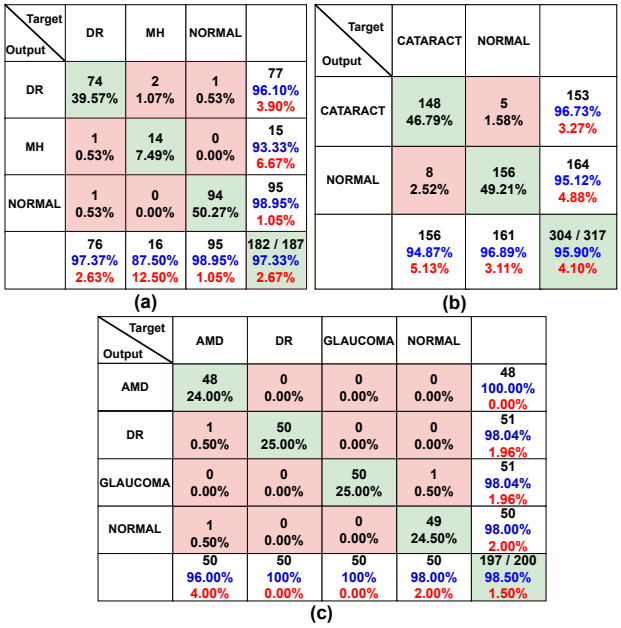

Fig. 4. Confusion matrix for validation of the classification performance of the developed model using: (a) the RFMid dataset, (b) the ODIR dataset, and (c) the FIVES dataset.

C. Subjective Analysis

The visual demonstration of vessel extraction performance is shown in Fig. 3. Fig. 3 (a) indicates the source images for the testing datasets. The corresponding vessel ground truths provided in the datasets are visualized in Fig. 3 (b). The vessels extracted by the RetiSegNet framework are demonstrated in Fig. 3 (c). It can be observed from Figs. 3 (c) that the developed RetisegNet model effectively extracts both the thick and

thin vessels. The corresponding masked image that contains the information from both vessels is shown in Fig. 3 (d).

Similarly, the developed STDeeepNet framework's disease classification capability can be demonstrated via confusion matrices. The confusion matrices of the STDeepNet framework for the FIVES, RMFiD, and ODIR datasets are demonstrated in Fig. 4. The RMFiD dataset consists of 3 disease classes of fundus images. The confusion matrix for STDeepNet on RFMiD, ODIR, and FIVES datasets are demonstrated in Fig. 4 (a), (b), and (c) respectively. From these figures, it may be observed that the designed STDeepNet is capable of identifying various fundus diseases from diverse datasets.

D. Objective Evaluation

The objective assessment of the developed framework RetiSegNet is carried out using the performance metrics [25] including Accuracy (Acc), Sensitivity (Sen), Specificity (Spe), and Dice coefficient (DSC) for the vessel extraction task. Similarly, the classification performance of the designed framework considers the Accuracy (Acc), F1-measure (F1), and area under the curve (AUC) [26]. The performance measures for vessel extraction tasks are reported in Table II. For the vessel extraction task, we have considered 4 benchmark datasets, as all these datasets include the vessel mask as ground truth. From Table II, it may be observed that the proposed RetiSegNet model attains higher Sen and DSC values against all SOTA methods for the FIVES dataset (Seen Setup). However, the RetiSegNet framework achieves competitive values of ACC and Spe against SOTA methods. Also, from Table II, it is found that the RetiSegNet model attains better performance over unseen data from the STARE, DRIVE, and CHASE-DB1 datasets against SOTA approaches.

To test the efficacy of the STDeepNet model, it is subjected to various disease class identification utilizing FIVES, RFMiD, and ODIR datasets. Table III indicates the performance of the STDeepNet against SOTA methods for the aforementioned datasets. It can be observed that the proposed framework achieves the higher values of Acc, F1-measures against all SOTA methods for all the datasets. Also, the STDeepNet attains a higher value of AUC for FIVES and RFMiD against all the SOTA methods. However, the STDeepNet provides the competitive value of AUC for ODIR dataset. From these Tables II - III it may be deduced that, the developed DeepRetinaNet is capable of generalizing diverse datasets and may be suitable for real-time applications. In the Tables II, III "-" indicates the unavailability of source code or data and the bold entries indicate the best value for the reported performance. The graphical analysis of the proposed RetiSegNet model against 25 SOTA approaches is presented in Fig. 5. Similarly, the graphical analysis of the STDeepNet is depicted in Fig. 6. Both graphs demonstrate the robustness of the proposed framework against 49 SOTA approaches.

TABLE II
OBJECTIVE COMPARISON OF VESSEL EXTRACTION PERFORMANCES OF THE PROPOSED RETISEGNET TECHNIQUE AGAINST SOTA METHODS

Models	Dataset	Acc(%)	Sen(%)	Spe(%)	DSC(%)
Multi-GlaucNet [25]	FIVES (Seen Setup)	97.98	89.36	**99.43**	85.62
SGAT-Net [27]		**98.86**	91.62	99.33	90.51
MCDAU-Net [28]		98.81	91	99.3	90.78
GT-DLA-dsHF [29]		98.76	90.79	99.26	90.48
O-SAM [7]		-	-	79.53	80.9
SCOPE [30]		-	-	85	85
VisionDeep-AI [19]		97.4	-	96.99	89.9
RetiSegNet (Proposed)		97.82	**94.71**	98.12	**91.2**
TUnet-LBF [31]	STARE (Unseen Setup)	96.81	80.04	98.5	81.98
SegR-Net [32]		-	82.12	97.41	80.75
ARSA-Unet [33]		96.63	87.93	97.72	85.35
SGL-Unet [34]		96.38	77.39	**98.67**	82.19
UCTransnet [35]		96.47	77.22	98.4	79.97
SwinUnet [35]		96.44	79.1	98.18	83.23
CSAU [36]		96.73	84.65	-	84.35
Octave-Unet [37]		97.13	86.64	97.98	81.91
RV-GAN [38]		**97.54**	83.56	98.64	83.23
RetiSegNet (Proposed)		97.04	**92.47**	97.52	**85.63**
CSAU [36]	DRIVE (Unseen Setup)	95.63	**83.49**	-	82.49
Octave-Unet [37]		96.64	83.74	97.9	81.27
Modified-Unet [35]		95.59	75.01	**98.59**	81.24
SGL-Unet [34]		**97.05**	83.8	98.34	83.16
RetiSegNet (Proposed)		95.48	82.83	97.7	**84.53**
Octave-Unet [37]	CHASE-DB1 (Unseen Setup)	97.59	86.7	98.4	83.13
Unet [39]		96.12	88.18	96.73	75.65
SGL-Unet [34]		97.71	86.9	**98.43**	82.71
UCTransnet [35]		96.47	77.22	98.4	79.97
SwinNet [35]		96.44	79.1	98.18	80.23
RetiSegNet (Proposed)		**97.92**	**98.4**	97.88	**86.76**

V. ABLATION STUDY

In this section, an ablation analysis was conducted for vessel extraction and classification to validate the efficacy of the developed DeepRetinaNet architecture. In the vessel extraction stage, the two components of the developed network: Residual QuattroNet and AEDNet are separately trained using the FIVES dataset. During training, the validation accuracy found from the Residual QuattroNet module is 61.96% with a space complexity of 20.5M. Similarly, the validation accuracy found from the AEDNet module is 97.74% with a space complexity of 7.6M. On the other hand, by combining the two models, the validation accuracy is increased to 99.66% with a space complexity of 28.2M illustrated in Table IV. Therefore, in the developed RetiSegNet, we combined both AEDNet and Residual QuattroNet to attain better performance which may be suitable for real-time applications.

In the classification stage, the classification is performed through developed STDeepNet with various network modifications illustrated in Table V. Initially, the LSTM layers used in the network are removed as a result, the classification accuracy is reduced to 81.31 % with a space complexity of 2.1M. Further, the network is trained with identity and convolution blocks without the introduction of MI and MCONV blocks, resulting in a classification accuracy of 74.24 %. On the use of the ReLU activation function instead of PReLU activation, the classification accuracy is reduced to 76.26 %. Hence, from the above analysis, it can be seen that the proposed STDeepNet achieves better performance as compared to any other modifications in the network.

An experimental analysis is performed to choose the number of epochs during the training of the proposed model. We have trained the proposed RetiSegNet and STDeepNet models with different numbers of epochs including 40, 50, and 60 while fixing the learning rate = 0.0001 and batch size = 4 which

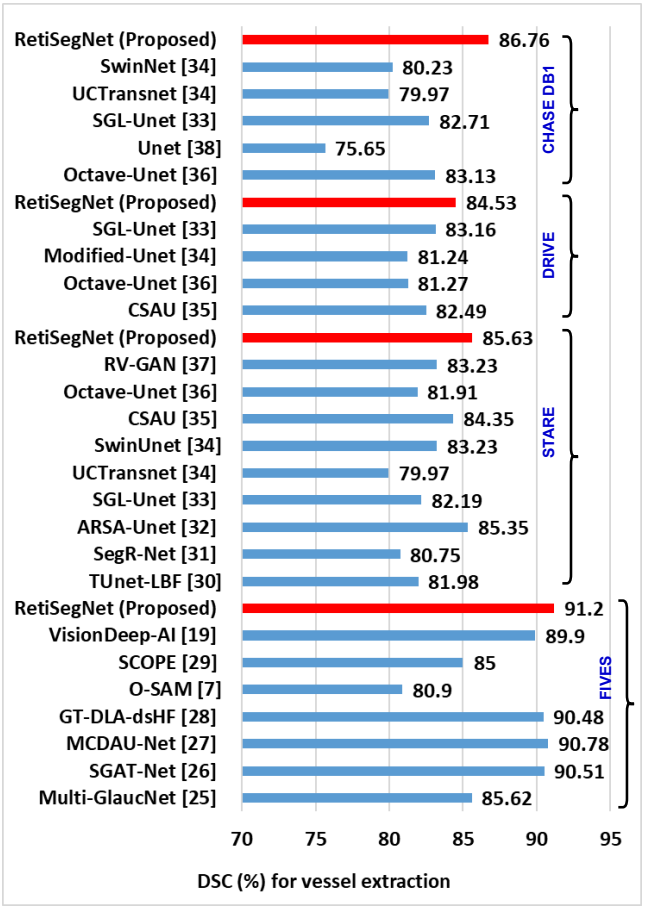

Fig. 5. Graphical comparison of the proposed RetiSegNet technique against 25 SOTA approaches for vessel extraction using DSC measure.

TABLE III
OBJECTIVE COMPARISON OF PROPOSED STDEEPNET MODEL AGAINST SOTA METHODS

Methods	Dataset	Acc(%)	F1(%)	AUC(%)
ResNet101 [40]		94.17	94.18	-
DenseNet169 [40]		93.33	93.33	-
Xception [40]		92.5	92.52	-
InceptionV3 [40]		91.67	91.66	-
DenseNet121 [40]		90.83	90.75	-
InceptionResNetV2 [40]	FIVES	90	90.08	-
ResNet50 [40]		89.17	89.26	-
EfficientNetB0 [40]		88.33	88.37	-
VisionDeep-AI [19]		81.5	81.15	82
STDeepNet (Proposed)		**99.25**	**98.49**	**99.01**
EfficientNetV2-S [41]		-	82.78	96.63
AlterNet-T [42]		-	83.92	96.67
EfficientNet B2 [43]		-	77.76	94.87
Inception-v3 [44]		-	82.45	96.13
Densenet-121 [45]	RFMiD	-	83.51	96.04
ResNet18 [46]		-	81.94	96.47
ResNeXt-50 [47]		-	81.3	95.54
Tj-CNN [48]		-	84.82	96.88
DNN [49]		92.34	95.19	-
STDeepNet (Proposed)		**98.18**	**95.33**	**98.34**
VGG16 + SGD [50]		71.28	85.57	84.93
Inception-v4 [51]		75.16	87.93	86.91
Vgg-16 [51]		73.02	87.3	86.81
BFENet [52]	ODIR	77.97	89.2	91.2
ResNet [53]		76.97	88.6	90.3
Fundus-DeepNet [26]		92.41	88.56	**99.76**
STDeepNet (Proposed)		**95.89**	**95.89**	95.33

TABLE IV
ABLATION ANALYSIS (IN TERMS OF SPACE COMPLEXITY) OF RETISEGNET WITH RESIDUAL QUATTRONET, AEDNET, AND COMBINATION OF BOTH

Network Type	Epochs	Layers	Trainable parameters (in million (M))	Segmentation Accuracy (%)
Residual QuattroNet	50	99	20.5M	61.96
AEDNet	50	45	7.6M	97.74
RetiSegNet	50	142	28.2M	99.66

are shown in Table VI. From Table VI it is observed that the proposed model attains better validation accuracy with a maximum 50 numbers of epochs as it balances learning and overfitting, allowing the model enough time to learn meaningful patterns without excessive risk of memorizing the data. Also, it balances the training time and computational resources.

Again, batch size is also an important hyperparameter that needs to be optimized for better performance of the developed algorithm shown in Table VII. From Table VII it is found that the proposed model with a batch size of 4 can enhance model generalization by introducing more noise into gradient updates, which helps the model avoid local minima and overfitting. It allows more frequent updates, potentially speeding up convergence in terms of iterations. Additionally, it reduces memory requirements, making it suitable for training on limited hardware. Further, an empirical analysis is conducted to choose the learning rate depicted in Table VII. This table indicates that training the proposed models with a learning rate = 0.0001 achieves higher validation accuracy against the learning rates of 0.01 and 0.00001. A small learning rate of 0.0001 helps ensure the training process is stable and avoids drastic jumps in the parameter values that could approach convergence. Furthermore, we have utilized an ADAM optimizer to adjust the learning rate for each parameter based on its past gradients, which can significantly speed up training and improve convergence. Further, ADAM combines the advantages of two popular optimization algorithms: AdaGrad and RMSprop. In Table VI and Table VII the proposed experiment is executed multiple times with different epochs. In each experiment, the training samples are chosen randomly. Therefore, the inference time for different images in the same dataset may vary due to differences in image complexity, such as texture, edges, and object density, affecting the inference time. The average inference times after training of each model are reported in Table VI. Similarly, for Table VII the proposed algorithms were tested by varying the learning rate and batch size by keeping the maximum number of epochs as 50. The inference time varies in multiple experiments based on the complexity of the image. The inference time of each model is mentioned in Table VII. From Table VII, it may be inferred that the RetiSegNet model with a batch size of 4 and a learning rate of 0.0001 attains a better validation accuracy of 99.66 %. Similarly, the STDeepNet attains a better validation accuracy of 98.18 % at a batch size of 4 and a learning rate of 0.0001.

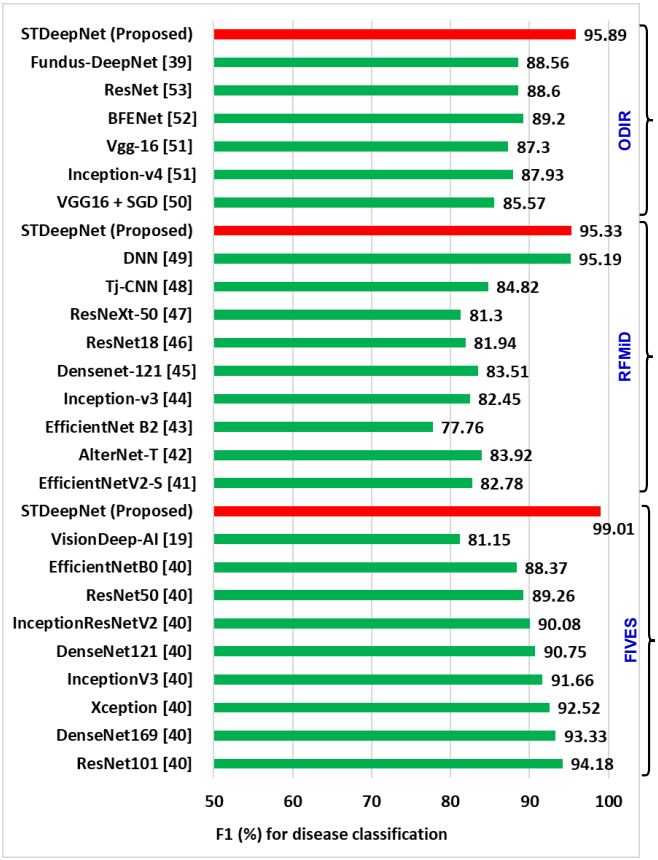

Fig. 6. Graphical comparison of the proposed STDeepNet framework against 24 SOTA approaches for disease classification using F1-measure.

TABLE V
ABLATION STUDY (IN TERMS OF SPACE COMPLEXITY)ON THE EFFECT ON THE PERFORMANCE OF VARIOUS COMPONENTS OF THE DEVELOPED STDEEPNET FRAMEWORK

Network Type	Layers	Trainable parameters (in million (M))	Validation Accuracy (%)
Without LSTM Layers	55	2.1M	81.31
Without MI and MCONV blocks	60	2.7M	74.24
With ReLU activation	60	2.7M	76.26
STDeepNet	60	2.7M	99.35

TABLE VI
EFFECT OF EPOCHS ON THE PERFORMANCE OF THE PROPOSED FRAMEWORKS

Proposed Frameworks	Epochs	Inference time	Validation Accuracy (%)
RetiSegNet	50	0.69 sec	99.66
	40	0.82 sec	97.51
	60	0.84 sec	94.25
STDeepNet	50	1.02 sec	98.18
	40	1.03 sec	93.58
	60	1.05 sec	91.44

VI. CONCLUSION

In this work, an effective DeepRetinaNet framework is developed for automated retinal disease diagnosis. The designed DeepRetinaNet model has two stages of novelties: retinal vessel extraction and disease classification. Here, the vessel extraction is performed using the developed RetiSegNet framework comprised of AEDNet and quattroNet networks. The developed RetiSegNet can effectively extract the vessels, as the AEDNet network is efficient for detecting the vessel boundary and the quattroNet network is capable of retaining global features. Subsequently, the disease classification task is carried out via the designed STDeepNet network consisting of MI, MCONV blocks, and the LSTM layers. The MI and MCONV blocks can accurately preserve the subtle detail variations in the fundus image. Also, the LSTM layers can effectively capture the images' long-term dependencies and contextual relationships.

TABLE VII
EFFECT OF LEARNING RATE AND BATCH SIZE ON THE PERFORMANCE OF THE PROPOSED FRAMEWORKS

Proposed Frameworks	Learning Rate	Batch Size	Inference time	Validation Accuracy (%)
RetiSegNet	0.0001	4	0.69 sec	99.66
	0.01	4	0.79 sec	91.21
	0.00001	4	0.57 sec	93.66
	0.0001	8	0.61 sec	96.54
STDeepNet	0.0001	4	1.02 sec	98.18
	0.01	4	0.82 sec	50.8
	0.00001	4	1.14 sec	82.89
	0.0001	8	0.81 sec	93.05

Therefore, the developed DeepRetinaNet framework is robust for vessel extraction as well as disease classification in challenging fundus images. Further, the designed model is capable of handling various complex datasets including FIVES, STARE, DRIVE, CHASE-DB1, RFMiD, and ODIR efficiently. From various experiments, it is observed that the proposed DeepRetinaNet surpasses 49 recently developed SOTA. However, the developed RetiSegNet model is computationally expensive. Again, the developed STDeepNet model needs to be validated for unseen data setup. Hence, in the future, the computational complexity of the framework would be reduced along with improving the network for more disease category identification with unseen setups.

REFERENCES

[1] M. J. Burton, J. Ramke, A. P. Marques, R. R. Bourne, N. Congdon, I. Jones, B. A. A. Tong, S. Arunga, D. Bachani, C. Bascaran *et al.*, "The lancet global health commission on global eye health: vision beyond 2020," *The Lancet Global Health*, vol. 9, no. 4, pp. 489–551, 2021, doi:10.1016/S2214-109X(20)30488-5.

[2] Z. Qiu, Y. Hu, X. Chen, D. Zeng, Q. Hu, and J. Liu, "Rethinking dual-stream super-resolution semantic learning in medical image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 1, pp. 451–464, 2024, doi: 10.1109/TPAMI.2023.3322735.

[3] E. Özbay, "An active deep learning method for diabetic retinopathy detection in segmented fundus images using artificial bee colony algorithm," *Artificial Intelligence Review*, vol. 56, no. 4, pp. 3291–3318, 2023, doi: 10.1007/s10462-022-10231-3.

[4] A. D. Vairamani, "Detection and diagnosis of diseases by feature extraction and analysis on fundus images using deep learning techniques," in *Computational Methods and Deep Learning for Ophthalmology*, 2023, doi: 10.1016/B978-0-323-95415-0.00009-7, pp. 211–227.

[5] R. Biswas, A. Vasan, and S. S. Roy, "Dilated deep neural network for segmentation of retinal blood vessels in fundus images," *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, vol. 44, no. 1, pp. 505–518, 2020, doi: 10.1007/s40998-019-00213-7.

[6] A. A. Abdulsahib, M. A. Mahmoud, H. Aris, S. S. Gunasekaran, and M. A. Mohammed, "An automated image segmentation and useful feature extraction algorithm for retinal blood vessels in fundus images," *Electronics*, vol. 11, no. 9, p. 1295, 2022, doi: 10.3390/electronics11091295.

[7] Z. Qiu, Y. Hu, H. Li, and J. Liu, "Learnable ophthalmology sam," *arXiv preprint arXiv:2304.13425*, 2023.

[8] A. Jayachandran, S. R. Kumar, and T. S. R. Perumal, "Multi-dimensional cascades neural network models for the segmentation of retinal vessels in colour fundus images," *Multimedia Tools and Applications*, vol. 82, no. 27, pp. 42 927–42 943, 2023, doi: 10.1007/s11042-023-15133-2.

[9] S. A. David, C. Mahesh, V. D. Kumar, K. Polat, A. Alhudhaif, and M. Nour, "Retinal Blood Vessels and Optic Disc Segmentation Using U-Net," *Mathematical Problems in Engineering*, vol. 2022, no. 1, p. 8030954, 2022, doi: 10.1155/2022/8030954.

[10] Y. Xu and Y. Fan, "Dual-channel asymmetric convolutional neural network for an efficient retinal blood vessel segmentation in eye fundus images," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 2, pp. 695–706, 2022, doi: 10.1016/j.bbe.2022.05.003.

[11] R. Thanki, "A deep neural network and machine learning approach for retinal fundus image classification," *Healthcare Analytics*, vol. 3, p. 100140, 2023, doi: 10.1016/j.health.2023.100140.

[12] I. K. Gupta, A. Choubey, and S. Choubey, "Mayfly optimization with deep learning enabled retinal fundus image classification model," *Computers and Electrical Engineering*, vol. 102, p. 108176, 2022, doi: 10.1016/j.compeleceng.2022.108176.

[13] S. Srinivasan, R. Nagarnaidu Rajaperumal, S. K. Mathivanan, P. Jayagopal, S. Krishnamoorthy, and S. Kardy, "Detection and grade classification of diabetic retinopathy and adult vitelliform macular dystrophy based on ophthalmoscopy images," *Electronics*, vol. 12, no. 4, p. 862, 2023, doi: 10.3390/electronics12040862.

[14] N. Sengar, R. C. Joshi, M. K. Dutta, and R. Burget, "EyeDeep-Net: a multi-class diagnosis of retinal diseases using deep neural network," *Neural Computing and Applications*, vol. 35, no. 14, pp. 10 551–10 571, 2023, doi: 10.1007/s00521-023-08249-x.

[15] S. Kumar and B. Kumar, "Automatic early glaucoma detection by extracting parapapillary atrophy and optic disc from fundus image using svm," *Multimedia Tools and Applications*, vol. 81, no. 10, pp. 13 513–13 535, 2022, doi: 10.1007/s11042-021-11023-7.

[16] F. Li, Y. Wang, T. Xu, L. Dong, L. Yan, M. Jiang, X. Zhang, H. Jiang, Z. Wu, and H. Zou, "Deep learning-based automated detection for diabetic retinopathy and diabetic macular oedema in retinal fundus photographs," *Eye*, vol. 36, no. 7, pp. 1433–1441, 2022, doi: 10.1038/s41433-021-01552-8.

[17] Y. Kumar and B. Gupta, "Retinal image blood vessel classification using hybrid deep learning in cataract diseased fundus images," *Biomedical Signal Processing and Control*, vol. 84, p. 104776, 2023, doi: 10.1016/j.bspc.2023.104776.

[18] K. Susheel Kumar and N. Pratap Singh, "Identification of retinal diseases based on retinal blood vessel segmentation using dagum pdf and feature-based machine learning," *The Imaging Science Journal*, vol. 71, no. 5, pp. 425–445, 2023, doi: 10.1080/13682199.2023.2183319.

[19] R. C. Joshi, A. K. Sharma, and 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*, vol. 94, p. 106273, 2024, doi: 10.1016/j.bspc.2024.106273.

[20] D. Nagpal, N. Alsubaie, B. O. Soufiene, M. S. Alqahtani, M. Abbas, and H. M. Almohiy, "Automatic detection of diabetic hypertensive retinopathy in fundus images using transfer learning," *Applied Sciences*, vol. 13, no. 8, p. 4695, 2023, doi: 10.3390/app13084695.

[21] K. Jin, X. Huang, J. Zhou, Y. Li, Y. Yan, Y. Sun, Q. Zhang, Y. Wang, and J. Ye, "FIVES: A fundus image dataset for artificial Intelligence based vessel segmentation," *Scientific data*, vol. 9, no. 1, p. 475, 2022, doi:10.1038/s41597-022-01564-3.

[22] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman, "An ensemble classification-based approach applied to retinal blood vessel segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 9, pp. 2538–2548, 2012, doi: 10.1109/TBME.2012.2205687.

[23] S. Pachade, P. Porwal, D. Thulkar, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, L. Giancardo, G. Quellec, and F. Mériaudeau, "Retinal fundus multi-disease image dataset (RFMiD): a dataset for multi-disease detection research," *Data*, vol. 6, no. 2, p. 14, 2021, doi: 10.3390/data6020014.

[24] D. S. W. Ting, L. R. Pasquale, L. Peng, J. P. Campbell, A. Y. Lee, R. Raman, G. S. W. Tan, L. Schmetterer, P. A. Keane, and T. Y. Wong, "Artificial intelligence and deep learning in ophthalmology," *British Journal of Ophthalmology*, vol. 103, no. 2, pp. 167–175, 2019, doi: 10.1136/bjophthalmol-2018-313173.

[25] H. Xiong, F. Long, M. S. Alam, and J. Sang, "Multi-GlaucNet: A multi-task model for optic disc segmentation, blood vessel segmentation and glaucoma detection," *Biomedical Signal Processing and Control*, vol. 99, p. 106850, 2025, doi: 10.1016/j.bspc.2024.106850.

[26] S. Al-Fahdawi, A. S. Al-Waisy, D. Q. Zeebaree, R. Qahwaji, H. Natiq, M. A. Mohammed, J. Nedoma, R. Martinek, and M. Deveci, "Fundus-DeepNet: Multi-label deep learning classification system for enhanced detection of multiple ocular diseases through data fusion of fundus images," *Information Fusion*, vol. 102, p. 102059, 2024, doi: 10.1016/j.inffus.2023.102059.

[27] J. Lin, X. Huang, H. Zhou, Y. Wang, and Q. Zhang, "Stimulus-guided adaptive transformer network for retinal blood vessel segmentation in fundus images," *Medical Image Analysis*, vol. 89, p. 102929, 2023, doi: 10.1016/j.media.2023.102929.

[28] W. Zhou, W. Bai, J. Ji, Y. Yi, N. Zhang, and W. Cui, "Dual-path multi-scale context dense aggregation network for retinal vessel segmentation," *Computers in Biology and Medicine*, vol. 164, p. 107269, 2023, doi: 10.1016/j.compbiomed.2023.107269.

[29] Y. Li, Y. Zhang, J.-Y. Liu, K. Wang, K. Zhang, G.-S. Zhang, X.-F. Liao, and G. Yang, "Global transformer and dual local attention network via deep-shallow hierarchical feature fusion for retinal vessel segmentation," *IEEE Transactions on Cybernetics*, vol. 53, no. 9, pp. 5826–5839, 2022, doi: 10.1109/TCYB.2022.3194099.

[30] Y. Yeganeh, A. Farshad, G. Guevercin, A. Abu-zer, R. Xiao, Y. Tang, E. Adeli, and N. Navab, "Scope: Structural continuity preservation for medical image segmentation," *arXiv preprint arXiv:2304.14572*, 2023, doi: 10.48550/arXiv.2304.14572.

[31] H. Zhang, W. Ni, Y. Luo, Y. Feng, R. Song, and X. Wang, "TUnet-LBF: Retinal fundus image fine segmentation model based on transformer Unet network and LBF," *Computers in Biology and Medicine*, vol. 159, p. 106937, 2023, doi: 10.1016/j.compbiomed.2023.106937.

[32] J. Ryu, M. U. Rehman, I. F. Nizami, and K. T. Chong, "SegR-Net: A deep learning framework with multi-scale feature fusion for robust retinal vessel segmentation," *Computers in Biology and Medicine*, vol. 163, p. 107132, 2023, doi: 10.1016/j.compbiomed.2023.107132.

[33] Y. Xie, J. Shang, Q. Yang, X. Qian, H. Zhang, and X. Tang, "ARSA-UNet: Atrous residual network based on structure-adaptive model for retinal vessel segmentation," *Biomedical Signal Processing and Control*, vol. 96, p. 106595, 2024, doi: 10.1016/j.bspc.2024.106595.

[34] Y. Zhou, H. Yu, and H. Shi, "Study group learning: Improving retinal vessel segmentation trained with noisy labels," in *Medical Image Computing and Computer Assisted Intervention–MICCAI*, 2021, doi: 10.1007/978-3-030-87193-2_6, pp. 57–67.

[35] M. T. Sadrabadi and H. Agahi, "Comparative analysis of retinal vessel segmentation utilising convolutional and transformer-based architectures," pp. 1–12, 2024, doi: dx.doi.org/10.2139/ssrn.5003364.

[36] R. Li, M. Li, J. Li, and Y. Zhou, "Connection sensitive attention U-NET for accurate retinal vessel segmentation," *arXiv preprint arXiv:1903.05558*, 2019, doi: 10.48550/arXiv.1903.05558.

[37] Z. Fan, J. Mo, B. Qiu, W. Li, G. Zhu, C. Li, J. Hu, Y. Rong, and X. Chen, "Accurate retinal vessel segmentation via octave convolution neural network," *arXiv preprint arXiv:1906.12193*, 2019, doi: 10.48550/arXiv.1906.12193.

[38] S. A. Kamran, K. F. Hossain, A. Tavakkoli, S. L. Zuckerbrod, K. M. Sanders, and S. A. Baker, "RV-GAN: Segmenting retinal vascular structure in fundus photographs using a novel multi-scale generative adversarial network," in *Medical image computing and computer assisted intervention–MICCAI*, 2021, doi: 10.1007/978-3-030-87237-3_4.

[39] X. Sun, H. Fang, Y. Yang, D. Zhu, L. Wang, J. Liu, and Y. Xu, "Robust retinal vessel segmentation from a data augmentation perspective," in *Ophthalmic Medical Image Analysis: 8th International Workshop, OMIA 2021, Held in Conjunction with MICCAI*, 2021, doi: 10.1007/978-3-030-87000-3_20.

[40] F. T. J. Faria, M. B. Moin, P. Debnath, A. I. Fahim, and F. M. Shah, "Explainable convolutional neural networks for retinal fundus classification and cutting-edge segmentation models for retinal blood vessels from fundus images," *arXiv preprint arXiv:2405.07338*, 2024.

[41] M. Tan and Q. V. Le, "EfficientNetV2: Smaller models and faster training," *arXiv preprint arXiv:2104.00298*, 2021, doi:10.48550/arXiv.2104.00298.

[42] N. Park and S. Kim, "How do vision transformers work?" *arXiv preprint arXiv:2202.06709*, 2022.
[43] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *arXiv preprint arXiv:1905.11946*, 2020, doi: 10.48550/arXiv.1905.11946.
[44] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, doi: 10.1109/CVPR.2016.308, pp. 2818–2826.
[45] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, doi:10.1109/CVPR.2017.243, pp. 4700–4708.
[46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2016, doi:10.1109/CVPR.2016.90, pp. 770–778.
[47] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2017, doi:10.1109/CVPR.2017.634, pp. 1492–1500.
[48] X. Xia, Y. Li, G. Xiao, K. Zhan, J. Yan, C. Cai, Y. Fang, and G. Huang, "Benchmarking deep models on retinal fundus disease diagnosis and a large-scale dataset," *Signal Processing: Image Communication*, vol. 127, pp. 117–131, 2024, doi: 10.1016/j.image.2024.117151.
[49] C. Priyadharsini *et al.*, "Deep hybrid architecture with stacked ensemble learning for binary classification of retinal disease," *Results in Engineering*, vol. 24, pp. 103–122, 2024, doi: 10.1016/j.rineng.2024.103219.
[50] N. Gour and P. Khanna, "Multi-class multi-label ophthalmological disease detection using transfer learning based convolutional neural network," *Biomedical Signal Processing and Control*, vol. 66, p. 102329, 2021, doi: 10.1016/j.bspc.2020.102329.
[51] N. Li, T. Li, C. Hu, K. Wang, and H. Kang, "A benchmark of ocular disease intelligent recognition: One shot for multi-disease detection," in *Benchmarking, Measuring, and Optimizing*, 2021, doi:10.1007/978-3-030-71058-3_11, pp. 177–193.
[52] X. Ou, L. Gao, X. Quan, H. Zhang, J. Yang, and W. Li, "BFENet: A two-stream interaction CNN method for multi-label ophthalmic diseases classification with bilateral fundus images," *Computer Methods and Programs in Biomedicine*, vol. 219, p. 106739, 2022, doi: 10.1016/j.cmpb.2022.106739.
[53] J. He, C. Li, J. Ye, Y. Qiao, and L. Gu, "Multi-label ocular disease classification with a dense correlation deep neural network," *Biomedical Signal Processing and Control*, vol. 63, p. 102167, 2021, doi: 10.1016/j.bspc.2020.102167.

Akshya Kumar Sahoo is an Assistant Professor in the Department of Electrical and Electronics Engineering at GIET University. He has received his B. Tech and M. Tech degrees from Biju Patnaik University of Technology, Odisha. He obtained his doctoral degree from GIET University, Gunupur in 2024. His research interests mainly focus on different computer vision applications for biomedical images.

Priyadarsan Parida is an Associate Professor in the Department of Electronics and Communication Engineering at GIET University. He has received his B. Tech and M. Tech degrees from Biju Patnaik University of Technology, Odisha. He obtained his doctoral degree from V. S. S. U. T., Burla, India, in 2019. His research interests mainly focus on different computer vision applications for biomedical images and secured communication.

Manoj Kumar Panda is an Assistant Professor in the Department of Electronics and Communication Engineering, GIET University, Gunupur, Rayagada, Odisha, India. He received M.Tech degree in Electronics and Communication from the NIST, Odisha, India, in 2011 and PhD degree from the IIT Jammu, India, in 2022. His current research interests include Image and Video Processing, Deep Learning.

Chittaranjan Nayak received the Ph.D. degree in engineering from the NIT, Agartala, India, in 2017. He is currently working as an Associate Professor in the Department of Communication Engineering, School of Electronics, VIT, Vellore. His current research interests include soft computing, 1-D photonic multilayers, and the formation of photonic nanojets for different optoelectronic applications.

N. Mohankumar received his B.E. Degree from Bharathiyar University, Tamilnadu, India in 2000 and M.E. & Ph.D Degree from Jadavpur University, Kolkata in 2004 & 2010. He is currently working as a Research Professor at SIT, Nagpur Campus, Symbiosis (International) Deemed University, Pune, India. His research interest includes modeling and simulation study of HEMTs and optimization of devices for RF applications.