Diabetic Retinopathy Image Lesion Segmentation with Feature Fusion Relation Transformer Network

[1,2]Shaymaa Hussein Nowfal, [3]Eswaramoorthy V, [4]Vishnu Priya Arivanantham, [5]Bhaskar Marapelli, [6]Swaroopa K and [7]Ezhil Dyana M V

[1]Medical Physics Department, College of Science, University of Warith Al-Anbiyaa, Karbala, Iraq.
[2]Medical Physics Department, College of Applied Medical Sciences, University of Kerbala, Karbala, Iraq.
[3]Department of Artificial Intelligence & Data Science, Bannari Amman Institute of Technology, Sathyamangalam, Tamil Nadu, India.
[4]Department of Computational Intelligence, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India.
[5]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India.
[6]Department of Computer Science and Engineering, Aditya University, Surampaleam, Andhra Pradesh, India.
[7]Department of Computer Science and Engineering, St. Joseph's Institute of Technology, Chennai, Tamil Nadu, India.
[1,2] shaymaa@uowa.edu.iq, [3]eswarinfotech@gmail.com, [4]a.vishnupriya@vit.ac.in, [5]bhaskarmarapelli@gmail.com, [6]drksp.cse@gmail.com, [7]dyanaezhil@gmail.com

Correspondence should be addressed to Eswaramoorthy V **:** eswarinfotech@gmail.com

Article Info
Journal of Machine and Computing (http://anapub.co.ke/journals/jmc/jmc.html)
Doi : https://doi.org/10.53759/7669/jmc202404096
Received 28 March 2024; Revised from 10 June 2024; Accepted 30 July 2024.
Available online 05 October 2024.

Abstract – Diabetes is a common disease that affects different vital organs of the human body, including the eyes. In diabetic patients, a change in blood sugar level leads to eye problems. Around 80% of the patients who have diabetes for more than 10 years have severe eye-related pathological disorders such as retinopathy and maculopathy. Proper detection, diagnosis, and treatment of eye-related pathologies prevent damage to the eye during the earliest stages of diabetic disease—the developed stage findings in patients losing their vision. The retinal damage due to diabetes is termed Diabetic Retinopathy (DR). The treatment of DR involves detecting the presence of the disease in the form of microaneurysms (MA), hemorrhages (HE), and exudates (EX) in the retinal area. The process of segmenting a massive segment of Retinal Images (RI) performs a prominent role in DR classification. The existing research concentrates on Optic Disc (OD) segmentation. This article focuses on the segmentation of MA, HE, and EX using a Feature Fusion Relation Transformer Network (FFRTNet). In this research, the benchmark dataset, the Indian Diabetic Retinopathy Image Dataset (IDRID), is used for the ablation study to evaluate the use of every module. The proposed method, FFRTNet, is compared with state-of-the-art methods. The evaluation of FFRTNet enhances the segmentation by 3.56%, 4.34%, and 3.75% on metrics, namely sensitivity, Intersection-over-Union (IoU), and Dice coefficient (DICE). The qualitative and quantitative results proved the superiority of FFRTNet in segmenting lesions in DR.

Keywords – Diabetes, Segmentation, Diabetic Retinopathy, Neural Network, Convolution, Feature Fusion, Lesion, Vessel.

I. INTRODUCTION

Authors must adhere to this Microsoft Word template when preparing their manuscripts for submission. It will speed up the review and typesetting process. Diabetic Retinopathy (DR) is a vision loss and blindness illness that affects people with diabetes. Diabetes is caused by high blood glucose, often known as blood sugar [1]. The retina's blood vessels are involved in the rear view of the eye's light-sensitive tissue. If a person has diabetes, that person should have a dilated eye exam at least once a year. There may be no signs of DR at first. Early detection can help in protecting the vision. Approximately 250 million individuals in the world currently have diabetes.

When individuals have diabetes, their blood glucose levels rise, which affects their retina and causes blindness or visual loss. The leading causes of vision loss and blindness in people include cataracts, glaucoma, macular degeneration due to age, and DR. The common name for the diabetic eye illness that causes blindness is DR. The blood vessels in the retina

are damaged, which might lead to blindness. One in four, almost 4.1 million persons with the same type of DR, have a visual loss.

According to the National Eye Institute, DR is the most important cause of blindness and blurred vision in persons between 20 and 74 and is the top cause of visual impairment. Retinopathy is a condition when there is damage to the blood vessels that supply blood to the retina. This disorder may result in impaired vision, eye haemorrhage, and, at last, total vision loss. This disease impairs the optic nerve by obstructing blood flow.

The application of image processing techniques and Computer Vision (CV) in different fields of science and engineering is growing speedily. There have been successful developments in computerized processing systems, such as image transformation. In the medicinal field, the present evolution of such methods to prevent the epidemic's progression is to recognize therapeutic disorders. It is achieved early by improving the complexity of the time needed to reveal the illness. These strategies rely on the accuracy of algorithms for diagnosing diseases in an actual period. It has predictable, distinct automated systems to detect DR from Retinal Images (RI) and Fundus Images (FI). Computer-aided screening and diagnosis save time and money for doctors by reducing the possibility of misdiagnosis. DL algorithms have advanced to the point where they can now analyze complicated aspects of medical data, resulting in rapid breakthroughs in automation. In ophthalmology, for retinopathy diagnosis and severity evaluation, such attempts have been assumed to analyze RI and develop models based on analysis.

A Clinical Decision Support System (CDSS) makes decisions about a patient's healthcare. CDSS can aid patients in achieving better results and receiving better treatment. The primary of CDSD is to give doctors, patients, and others prompt information to make better healthcare decisions. Sets are ordered for specific conditions or groups of patients, as well as ideas and databases with information about patients; CDSS tools include reminders for preventive care and alarms regarding potentially unsafe conditions. CDSS helps you save money, improve efficiency, and make patients feel better [16]. CDSS can simultaneously address all three challenges, such as alerting doctors to the likelihood of repeat testing on a patient.

Image Segmentation (IS) is the first step in image processing, and the required measures from the objects are retrieved. The level of segregation depends on the complexity of the problem. The IS paused once it spots the edge region. Thus, retrieving the lesions from the background image is the primary objective of segmentation. In edge detection, darker areas make convolution, and it is rectified by the approaches of Histogram Equalization (HE) and thresholding. Image segmentation is a primary procedure in image processing, and it divides the image into several regions. Through IS, information from the digital image needed is retrieved. Several techniques conduct IS, and the complexity of the problem determines the choice of the required method.

This research aims to implement the automatic identification of DR using digital FI. Designed systems and algorithms are a step toward achieving computer-aided screening as a tool for physicians and medical experts. Further, insights from art and emerging systems are displayed in the images in hopes of further updating and accelerating more work in automated object recognition. This work proves the automatic detection of DR by applying the FI of the retina to specific lesions of DR. In this work, the identification of the haemorrhages, microaneurysms, and exudates has been developed through a learning-based method. Medical decisions are an information source for CDSS.

The Primary Aim of This Research Paper is

(a) Pre-process the image to remove unwanted noise and image objects from the image. Medical images are attained with a specific volume of noise data, which can eventually show the impact on the further diagnosis of the disease. The Modified Mean Filter (MMF) 's pre-processing procedure enhances the image's quality.
(b) The Feature Fusion Relation Transformer Network (FFRTNet) segments the significant segments of vessels and lesions. The main sections are highlighted with FFRTNet, which is used for further prominent processing.

The research work is systematized as follows: the overview of DR, the significance of segmentation, and contributions are detailed in Section 1; the comprehensive DR segmentation is given with gap analysis done in Section 2; the dataset preparation, pre-processing methodology, and the system of proposed FFRTNet are described in Section 3, the outcome of DR is illustrated with comparative analysis and discussion in Section 4, and the article is concluded with future scope in Section 5.

II. RELATED WORKS

RI investigation is still a challenging and adaptable field of study [2]. Different retinal, cardiovascular, and major disorders are distinguished by observing the retinal vasculature and the morphological changes. Albeit retinal vascular adjustments are unpretentious, early acknowledgment of the signs is fundamental for avoiding ophthalmologic entanglements and vision misfortune. A detailed analysis of DR images is given in this section.

In [3], they proposed an end-to-end encoder-decoder model called Decoder Network (DRNet), which is used to segment Fovea centers and OD. The research recommended a skip link in DRNet called a residual skip connection to compensate for the spatial information lost because of pooling in the encoder. Computer-Aided Screening Tools (CAST) are essential for non-intrusive diagnostic procedures in contemporary ophthalmology. They play a significant role in precisely segmenting fovea centres and the optic disc (OD). Small dataset sizes, inconsistent spatial, texture, and shape information between the OD and Fovea, and the existence of different things make it challenging to create such an automated method.

In [4-5], segmentation that employs an encoder-decoder model, skipping connection, and elaborated convolutions are performed using the DeepLabv3 network. The DeepLabv3 is input to the Extreme Inception (Xception) framework for segmenting DR lesions in this research work. A clinical experiment study is run to find the best hyperparameters for the segmentation model's training process that produced successful segmentation results during testing.

A Cascade Attentive RefineNet (CARNet) was introduced in [6-7] to automate and precisely segment DR with many lesions. The FI's fine local features and coarse global data may be fully used. Global and local image encoder and attention refinement decoder structure of CARNet. It used Residual Network 50 (ResNet50) and ResNet101 to downscale the entire and patch images to extract lesion characteristics. The high-level refinement decoder generates accurate predictions using features from two encoders and the low-level attention refined module.

[8-9] have proposed a method for diagnosing DR using desktop diagnostics. Untreated DR causes blindness in the retina. Microaneurysms and RI/FI localization are first found. Dynamic thresholding and multi-scale correlation filtering are used to find the microaneurysms in the RI/FI. In this method, microaneurysm candidate identification and actual microaneurysm classification are two levels. The online retinopathy challenge and standard DR are efficient and effective open-source datasets.

[10-11] has presented an Optic Disc (OD) position-based technique with a vessel's location-matched filter. As a primary step, a binary mask is made after the image brightness and contrast equalization. Next, the retinal vasculature is sectioned, and the positions of the blood vessels are processed to the filter, thereby presenting the vessel's places in the vicinity of the OD.

[12-13] have analyzed the procedure for identifying the disease by using an OD boundary. Detecting the location of the macula is considered an advantage in DR analysis. However, the limitation is that it does not cover the complete anatomical structure of retinopathy images.

The Segmentation of retinal blood vessels using Artificial Neural Networks (ANN) for early detection of DR by [14] has worked hard to distinguish blood vessels in the retina using ANN concepts to detect disease at an early stage. The authors used monitoring techniques to find targets. They eventually concluded that they could achieve exceptionally reliable results by using ANNs and NNs and an algorithm to detect DR in the preliminary stages in humans. Still, they only worked for limited images from standard databases.

This section explains recent methods of DR segmentation, in which few researchers concentrated on ineffective OD segmentation over other DR-causing features. The segmentation of other significant components, MA, HE, and EX, can give an accurate result. This research work considers the gap in the existing research studies and formulates a Feature Fusion Relation Transformer Network (FFRTNet) for the segmentation. The lesions and vessels in the FI are highlighted with the help of the proposed FFRTNet [15].

III. MATERIALS AND METHODS

This section explains the data preparation and selection process, the pre-processing stage involved in this research, and the segmentation process. The dataset involved in this research is free to access, and the incidence of Gaussian noise is removed using a Modified Median Filter (MMF) [16-18]. The segmentation uses a Feature Fusion Relation Transformer Network (FFRTNet), where the significant regions are highlighted. **Fig 1** shows the complete block diagram of the proposed method.

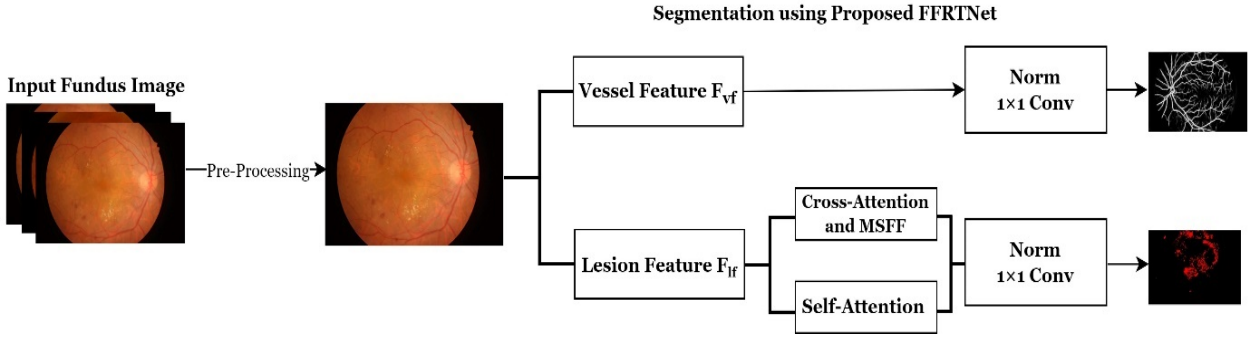
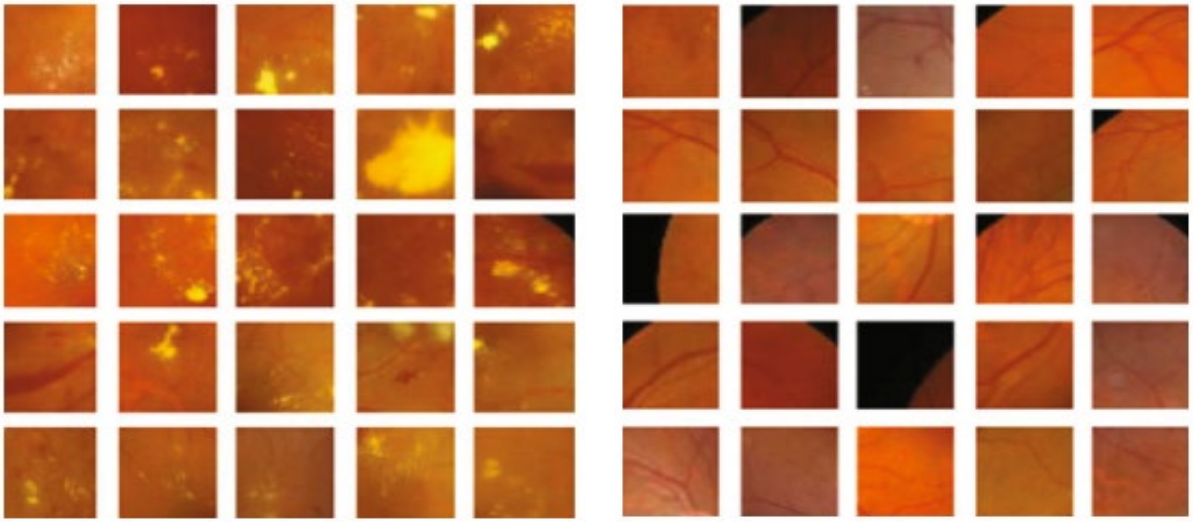


Fig 1. Block Diagram of FFRTNet.

Data Preparation

The study implemented a dataset that was collected from the Indian Diabetic Retinopathy Image Dataset (IDRID). Data from a patient's FI, an actual clinical study, and an eye clinic in India has been used in this research work; each image in the dataset was captured with a Kowa VX-10 colour fundus camera that had a field of view of 50 degrees and was positioned close to the macula. Each image has a 4288×2848 resolution and was created in JPG format. The research chose 81 color FI with pixel-level annotations out of 516 for this clinical study. Three typical DR anomalies are considered in this dataset, as shown in **Fig 2** Differentiation labels are used to divide the IDRID into training and testing sets. Empirically, **Table 1** shows the distribution results.



(a) Positive (b) Negative

Fig 2. Colour Fundus Image with (a) Positive and (b) Negative Samples.

Table 1. The Scope of IDRID's Distribution

Lesion Type	Training Set	Testing Set
Soft Exudates	26	14
Microaneurysms	54	27
Haemorrhage	53	27
Hard exudates	54	27

Pre-processing of Fundus Image

Pre-processing is a technique for improving image data that supports subsequent image processing by clipping redundant distortions or noise and improving the visual characteristics. Pre-processing techniques like intensity normalization, contrast augmentation, de-noising, and others significantly influence the results of fundus lesion segmentation. Poor image quality and low contrast make processing time-consuming and inaccurate, affecting how well the segmentation works. Typically, noise is a necessary segment of any image. The image denoising process is used to eliminate noise. Noise in the picture must be reduced to find the accurate Region of Interest (ROI) between normal and pathological tissues, which generates reliable results. Different methods exist to lessen noise and enhance the contrast between the FI. One such technique is the use of pre-processing filters.

Noises in the digital images may be redundant signals that randomly interfere with the desired signal or odd pixels that do not stand for the actual scene. During image acquisition, the bizarre arrangement of the image sensors may produce minimum noise. Noise removal is an essential pre-processing technique for different image-processing functions such as image registration, IS, and object recognition. Gaussian noise is frequently encountered in acquired images. Gaussian noise affects the entire image consistently and alters each pixel in the image from its original value based on the standard deviation of the noise.

Gaussian Filter

Gaussian function as an input filter creates a high pass in the step function of input where the initiation and finishing time are minimized. The input data is converted into an Eigenstress transform using a Gaussian filter. The FI is smoothened using a linear spatial filter, reducing noise and frequency at the edges. Attain smoothing, the convolve in the Gaussian function is given in Eq. (1) and Eq. (2).

$$G_\sigma(a,b) * I(a,b) \tag{1}$$

$$G_\sigma(a,b) = \frac{1}{2\pi\sigma^2} e^{-\frac{a^2+b^2}{2\sigma^2}} \tag{2}$$

where the input image is shown using I(a, b) and the Gaussian function is using $G\sigma$ (a, b) with the spatial coordinate values (σ, a b). The convolution operator is indicated using *. The edges of the image are located, and the find-out gradient is transmitted to the FI, which is in Eq. (3).

$$\nabla(G_\sigma(a,b) * I(a,b)) \tag{3}$$

where the gradient operator '∇' is stated to value the directional alteration in the intensity values. The boundary map is in Eq. (4).

$$M_Y(a,b) = \nabla(-G_\sigma(a,b) * I(a,b)) \tag{4}$$

The normalized factor of the boundary map is equated in Eq. (5).

$$M_{NY}(a,b) = \frac{M_Y(a,b) - \min(M_Y(a,b))}{\max(M_Y(a,b) - minM_Y(a,b))} \tag{5}$$

The threshold value is $T \in [0,1]$, which is used for the bound map with binary values, and it is given in Eq. (6.)

$$M_{YY}(a,b) = \begin{cases} if\ M_{NY}(a,b) & 1 \\ else & 0 \end{cases} \tag{6}$$

The contrast and intensity of the FI determine the threshold value, which may vary depending on the distribution. The threshold value used to remove the input fundus image's low-intensity region and object continuity is 0.1. The extracted edge specifies an encompass, ensuring the final convergence is within bounds.

Modified Mean Filter (MMF)

The eigenvalue is fundamental in several image processing functions. This pre-processing is enhanced using eigenvalues as a threshold to remove Gaussian noise from the FI. To reduce Gaussian noise, thresholding is often considered in the transform area rather than the spatial area. Here, thresholding is used to compute better pixels using spatial coefficients. The noisy FI of size M×N is first computed for its standard deviation value of noise, and then mean filtering is applied. The noisy FI is subtracted from the mean filtered image to generate a variance mask, which is then used to determine a threshold factor. The value is generated based on the eigenvalues based. The mean filtered FI is tuned to produce the denoised FI by the threshold value. The processed current pixel is $X_{ij,}$ and the restored pixel is Y_{ij}. The sliding window size is W×W, which is cantered at $X_{ij,}$ where the 'W' is assigned with the positive integer 2L+1. The pixel set created in X_{ij} is figured out as $\{X_{i-u,j-v,} - L \le u, v \le L\}$. The difference of the noise approximation mask is 4.6, and the mean is '0'. The Eq. (7) is used to assess the SD, σ_{GNM} of Gaussian noise.

$$\sigma_{GNM} = \frac{1}{4.6} \cdot \frac{1}{MN} \sum_{i,j=1}^{M,N} \left| (X \times MASK)_{ij} \right| \tag{7}$$

The abovementioned method is vital for approximating noise densities in real-time FI. The projected noise Standard Deviation (SD) is often used to determine the window size. A 3×3 sliding window is used in the MMF method for the noise SD of less than 20 to drop Gaussian noise and improve edge retention. A 5×5 sliding mask drops Gaussian noise when the noise SD exceeds 20. Therefore, selecting a fixed window size reduces processing time. The Modified Mean Filter (MMF) procedure is given in Algorithm 1.

Algorithm 1 for Pre-Processing of Images Using MMF

Step 1. The image with noise X of size M×N as input

Step 2. The Gaussian noise model is calculated using EQU 7

Step 3. Choose the mask with the size $W = \begin{cases} 3 \times 3, & \sigma_{GNM} \le 20 \\ 5 \times 5, & \sigma_{GNM} \le 20 \end{cases}$

Step 4. The MMF in the noisy FI is given in $\widehat{X_{ij}} = \frac{1}{r} \sum_{i=1}^{r} \sum_{j=1}^{r} X_{ij}$

Step 5. The local mask *'r*-value is assigned as 25 or 7.

Step 6. Apply sliding window size of 3×3 and accept the deviation among MMF applied FI and noise image using $D_{ij} = \left| X_{ij}.. - \widehat{X_{ij}}.. \right|$

Step 7. Estimate of eigen variation mask value D_{ij} is predictable using $\det\left(D_{ij} - \lambda \mathrm{I}\right) = 0$

Step 8. Compute the absolute mean ($|\mu|$) of eigenvalues of difference mask D_{ij} and set T =$|\mu|$

Step 9. The pixel values greater than 'T' are trimmed, and the vector 'V' is $V = \{\widehat{X_1}, \widehat{X_2}, \widehat{X_3}, \ldots . \widehat{X_n}, \}$

Step 10. The pixel value restored is $Y_{ij} = mean(V)$

Step 11. The final de-noised image is Y.

Segmentation of Fundus Images

This section describes the summary of significant modules in the network, namely the Relation Transformer Block (RTB), Multiscale Feature Fusion Block (MSFF), Global Transformer Block (GTB), and loss function. The proposed FFRTNet is made up of four important parts: the segmentation head, Relation Transformer Block (RTB), Multiscale Feature Fusion

Block (MSFF), and Global Transformer Block (GTB). FFRTNet (**Fig 1**) is a dual-branch method that independently investigates vascular and pathogenic aspects, using transformers based on GTB and RTB to consider interactions.

The FI is passed via the network's backbone to get the feature map F with the channel count C and resolution $W \times H$. GTB is used in identifying the long-range of feature map dependencies that result in lesion feature F_{lf} and distinct vessel feature F_{vf}. To lessen the spatial loss of the image, a convolutional layer is added to the network in place of the pooling layer. Integrating the Contextual Channel Attention (CCA) framework into encoders, the MSFF block helps the web learn multiscale features and enriches data sent with skip connections and lower decoder resolution.

Further, RTB is integrated into the network to model the spatial relation between lesion and vessel due to the properties of an essential pathological connection that uses a Cross-attention Head (*CH*) and Self-attention Head (*SH*). The self-attentive feature F_{sf} is generated and shown using SH, where F_{lf} accepted as input, and CH received (F_{lf}, F_{vf}). The cross-attentive feature F_{cf} is generated by the combination of fine-grained vessel structures. Create the output of the RTB, F_{sf} and F_{cf} are combined. Using vessel characteristics and lesion information, two sibling heads predict vascular and pathological masks using a Norm layer and 1×1 convolution.

The GTB comprises one head, and the RTB has two heads, which is based on the structure of the transformer. It created the key, value, and query relation for reasoning. The GTB query head is identical to channel-wise weight, and RTB encompasses a matching size. During training, GTB generates distinct vessel features and multi-lesion that maintain interested regions. The pathogenic relation among vessels and multi-lesion is found using RTB, which removes noise and implies the location data.

Two consecutive subdivisions of the same design GTB independently extract the characteristics of vessels and lesions. The fact that vessels and lesions impose radically distinct visual patterns accountable for such a dual-branch model. Randomly distributed lesions are discrete patterns. The central retinal artery and ciliary artery follow standard guidelines, but specialized branches are needed in order to understand branching.

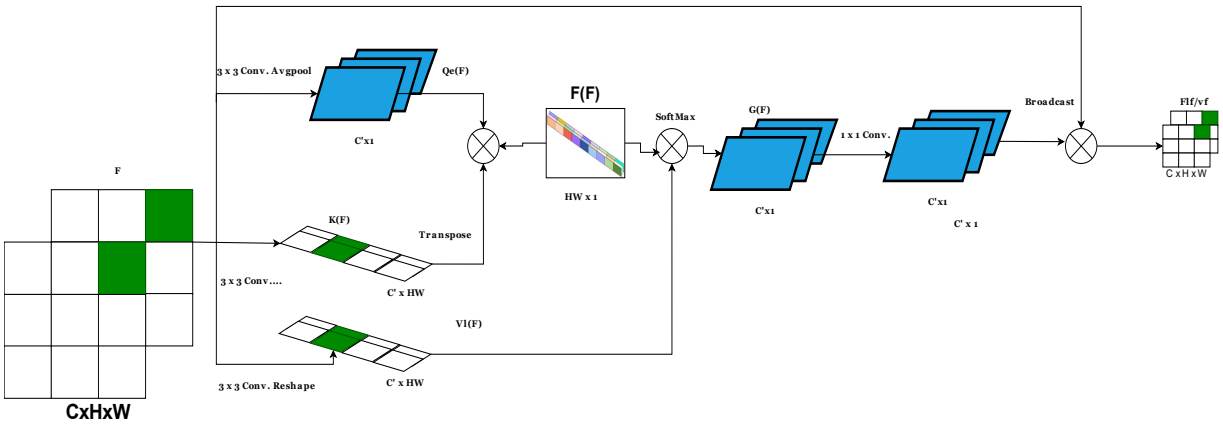


Fig 3. Structure of GTB.

The detailed model of GTB is given in **Fig 3**, and it assumes $F \in \mathbb{R}^{C\times W\times H}$ as input that is generated from the attentively refined output vessels $i \in \{lf, vf\}$ and lesions $F_i \in \mathbb{R}^{C\times W\times H}$. The F is transmitted into the generators, namely value (*Vl*), key (*Ke*), and query (*Qe*). The query is set out using a 3 × 3 convolutional layer that is followed by MSFF that gives an output vector $Qe(F) \in \mathbb{R}^{C'\times 1}$ with $C' = C/8$ as the channel number.

The generators Vl and Ke have a similar architecture as Qe with the reshape operation leading to $Ke(F), Vl(F) \in \mathbb{R}^{C'\times HW}$. The pairwise matrix multiplication is determined as Eq. (8).

$$\mathcal{F}(F) = Ke(F)^T Qe(F) \tag{8}$$

where the transport operator is indicated using 'T'. The Qe vector function as a feature selector for the key matrix channel, and the product value $\mathcal{F}(F) \in \mathbb{R}^{HW\times 1}$ function as a feature selector for the value matrix in the spatial position. The GTB is an attention mechanism that merges spatial and channel-wise weighted features with input data. The transform operation in the global layer is determined as in Eq. (9).

$$G(F) = Ve(F)softmax(\mathcal{F}(F)) \in \mathbb{R}^{C'\times 1} \tag{9}$$

where the value of $\mathcal{F}(F)$ is normalized using *SoftMax*. The bought attentive feature $G(F)$ and linearly embedded remaining terms are considered as input to the feature map. The final value is attained via the residual connection in Eq. (10).

$$F_i = WG(F) + F, i \in \{lf, vf\} \tag{10}$$

where the sum function across the element-wise broadcast channel is written down as + and the linear embedding factor is indicated as 'W', where these values are deployed on a 1×1 convolutional layer for converting the intermediate channel number from C' to C. Consequently, the output features, which have been enhanced with specific vessel and lesion features, are received in the same format as the input features.

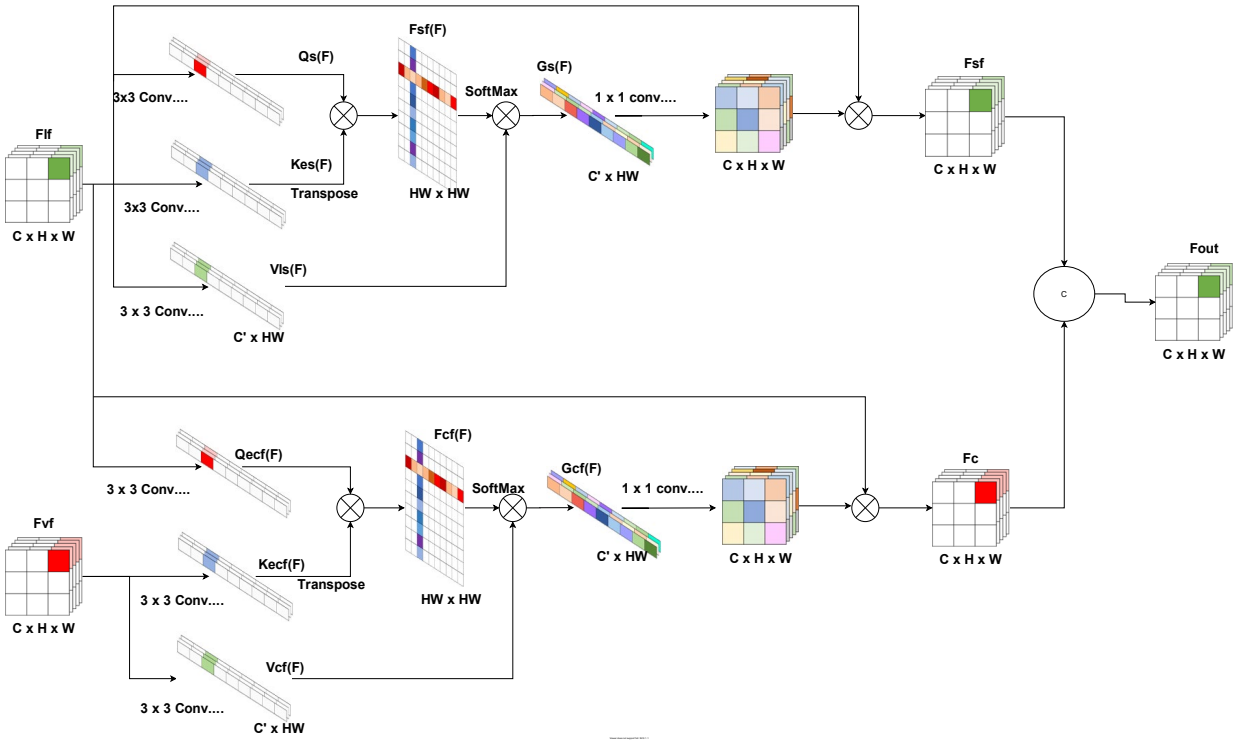


Fig 4. Structure of RTB.

The GCNet is used as a model for the GTB structure where GTB differs from GCNet. While producing the per-channel weights, they both adhere to the transformer mechanism concept. The GCNet and GTB get their weight vectors from the matrix multiplication. The generators highlight the significant channels and get spatial and channel attention. The fundus lesion in the tiny discrete one is confused with idiosyncratic tissue or objects. The prominent data is available in fewer pixels of a specific network channel. The feasibility of FI segmentation is aimed at improving the method.

As illustrated in **Fig 4**, RTB comprises two heads: self-attention and cross-attention—these heads record intraclass interdependence between (lesions and interclass) and (lesions and vessels). RTB is deployed on a 3×3 convolutional layer, and it is followed by the reshape techniques, namely value (Vl_i), key (Ke_i), and query (Ge_i) that belongs to $\{s_f, c_f\}$. The cross-attention and self-attention in pairwise estimation are equated in Eq. (11).

$$\begin{aligned} \mathcal{F}_{sf}(F_{lf}) &= Ke_{sf}(F_{lf})^T Qe_{sf}(F_{lf}) \\ \mathcal{F}_{cf}(F_{lf}, F_{vf}) &= Ke_{cf}(F_{vf})^T Qe_{cf}(F_{lf}) \end{aligned} \tag{11}$$

The evaluation is significant where the query is derived from the input lesion feature by the head of self-attention, and the key is derived from the vessel feature by the head of cross-attention. Subsequently, attentive features are assessed from the heads using Eq. (12) and Eq. (13).

$$G_{sf}(F_{lf}) = Ve_{sf}(F_{lf}) SoftMax(\mathcal{F}_{sf}(F_{lf})) \tag{12}$$

$$G_{cf}(F_{lf}, F_{vf}) = Ve_{cf}(F_{vf})\, SoftMax(\mathcal{F}_{cf}(F_{lf}, F_{cf})) \tag{13}$$

Every head adopts the residual learning process and generates the output using Eq. (14).

$$F_i = W_i G_i(F_{lf}, F_{vf}) \oplus F_{lf}\ i \in \{sf, cf\} \tag{14}$$

where the linear embedding element is denoted as W_i, and the element-wise addition is conducted using $\oplus$.

By weighing all locations, the self-attention head accurately captures long-range relationships. Self-attention can model intra-class pairwise interactions with multiple DR lesions regardless of location. In lesion segmentation, the head distinguishes between the edges of multiple lesions and sharpens broad patterns. The cross-attention head incorporates interactions among lesions and vessels by querying the global vascular model from vessel attributes. Given the strong pathogenic links that lesions and vasculature already have, cross-attention can aid in more accurately finding Micro-Aneurysms (MAs) and Soft Exudates (SEs) while also removing FP for hard EXudates (EXs) and MAs due to vessel reflection and capillary confusion, respectively.

The outcome of features F_{sf} from the self-attention and F_{cf} from cross-attention generates the final RTB using Eq. (15).

$$F_{out} = [F_{sf}; F_{cf}] \tag{15}$$

where the channel dimension addition is shown using [;].

The vessel and lesion segmentation are done using the loss function $\mathcal{L}_{vessel}$ and $\mathcal{L}_{lesion}$. The total loss in this network is stated by Eq. (16).

$$L = \lambda \mathcal{L}_{vessel} + \mathcal{L}_{lesion} \tag{16}$$

where the multi-lesion segmentation's weighted cross-entropy with 5-Class is indicated using $\mathcal{L}_{lesion}$ and vascular feature learning of binary weighted cross entropy across the loss is indicated using $\mathcal{L}_{vessel}$. The weight value in the loss is determined by and assigned with the value 0.0. Only the multi-lesion features are used to perfect the network, and vascular information becomes significant if it increases.

IV. RESULT AND DISCUSSION

The proposed FFRTNet is performed on an NVIDIA GPU with 24 GB of RAM, deploying PyTorch as the backend. The batch size for the training process is 16. The starting learning rate is assigned as 0.001 and gradually decreases after 120 epochs to 0.1 times the initial value. Every model is trained using the Stochastic Gradient Descent (SGD) optimizer for 250 iterations, with momentum set to 0.9 and weight decay set to 0.0005. The return loss is assigned as 0.1. The weight of $\mathcal{L}_{lesion}$ of MA is 1.0, HA is 0.001, HE is 0.1, and SE is 0.1. The coefficient of $\mathcal{L}_{vessel}$ is assigned as 1.0 and 0.01. This section describes the pre-processing and segmentation of FI using comparative analysis.

Pre-Processing of Fundus Image

The FI is initially smoothed using image smoothening. Smoothing is used to generate images with fewer pixels and less noise. The different smoothing methods rely on low-pass filters, but they also use a kernel that is a moving collection of pixels to smooth an image by capturing the average (or) median measurement for the group of pixels. The smoothened Gaussian filter and Modified Mean Filter (MMF) applied images are given in **Fig 5**.

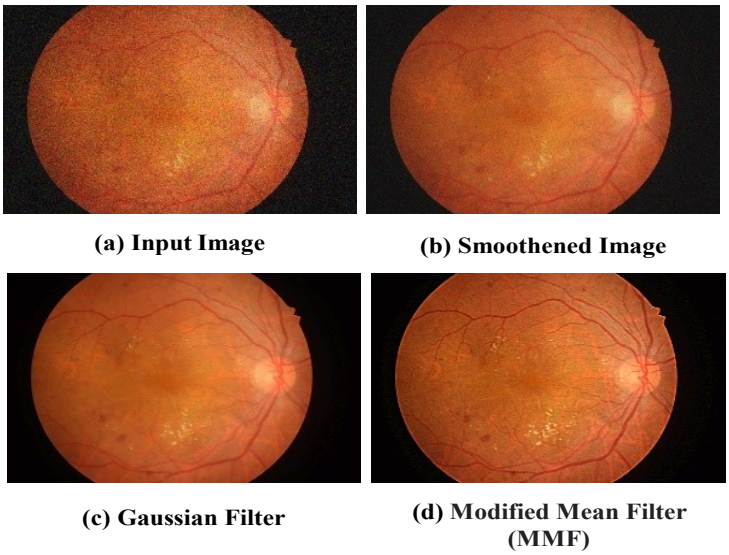


Fig 5. Comparison of Pre-Processing Approaches.

Fig 5 illustrates the output of pre-processing techniques. **Fig 5 (a)** depicts the input image, which is smoothened in **Fig 5 (b).** Further, the image is applied with the filters, namely existing Gaussian and proposed MMF. The Proposed MMF effectively removes the redundant regions from the FI. The prepared image is passed to the following IS technique.

Segmentation of Fundus Images

Implementing generalization over different domains and imaging scenarios is challenging yet fundamental for medical images. Models are evaluated on the IDRiD dataset [17], which was attained from an added source after they have been trained using images from the training set of the DDR dataset. **Fig 6** compares the findings of the IS analysis. It shows that the proposed method performs at its best by reducing the distance between images captured under different settings.

Fig 6 depicts the IS of MA, HA, HE, and SE from the FI using the proposed FFRTNet. The convolution layer is deployed to minimize the spatial loss of the data, and MSFF effectively learns the multiscale feature. It obtains the IS process effective over the existing *state-of-the-art* technique, which is discussed in a subsequent section.

When evaluating the process of IS, performance evaluation metrics are necessary. The research employs different metrics, such as Sensitivity, Dice coefficient (DICE), and IoU, to assess the performance of segmenting DR images and quantitatively analyze the experimental clinical test data. Initially, the True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) values are determined. The predicted and true lesion identification is TP, the predicted and true normal lesion is FP, the said normal and true normal is TN, and the true and predicted normal is FN. The performance measure is stated in Eq. (17) to Eq. (19).

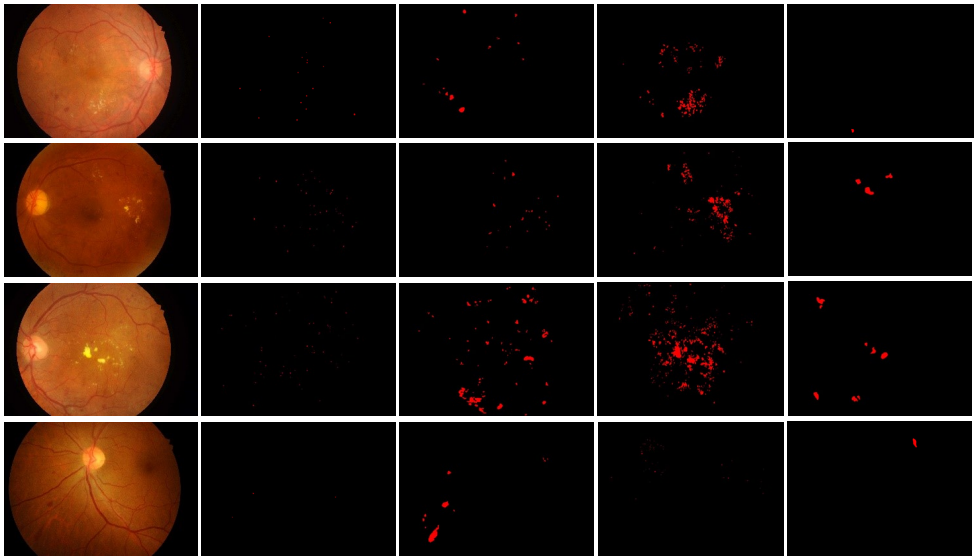


Fig 6. Comparative Analysis of IS.

$$Sensitivity = \frac{TP}{TP+FN} \tag{17}$$

$$DICE = \frac{2TP}{2TP+FP+FN} \tag{18}$$

$$IoU = \frac{TP}{TP+FP+FN} \tag{19}$$

Sensitivity is the frequency of diseases being misdiagnosed. This work refers to the ratio of accurate to total lesion area, a crucial consideration for patients and medical experts. This research has concentrated on reducing the rate of misdiagnosis in real-world applications. IoU is an evaluation metric that determines the degree to which predicted results and actual outcomes overlap. Using this value, one can determine how accurate a particular semantic segmentation method is. DICE measures the similarity of ground truth and prediction, which predicts FP and FN techniques. The weights close to 1 of sensitivity, DICE, and IoU determine the effectiveness of the segmentation.

Table 2. Performance Comparison of Segmentation of MA and HA

Feature	**MA**			**HA**		
Segmentation Method	Sensitivity	DICE	IoU	Sensitivity	DICE	IoU
DRNet	0.5171	0.6354	0.4791	0.6698	0.7898	0.6572
DeepLabV3	0.5518	0.6789	0.5367	0.6946	0.8577	0.6937
CARNet	0.5946	0.7398	0.5631	0.7363	0.8654	0.7468
FFRTNet	0.6145	0.7561	0.6134	0.7451	0.8781	0.7615

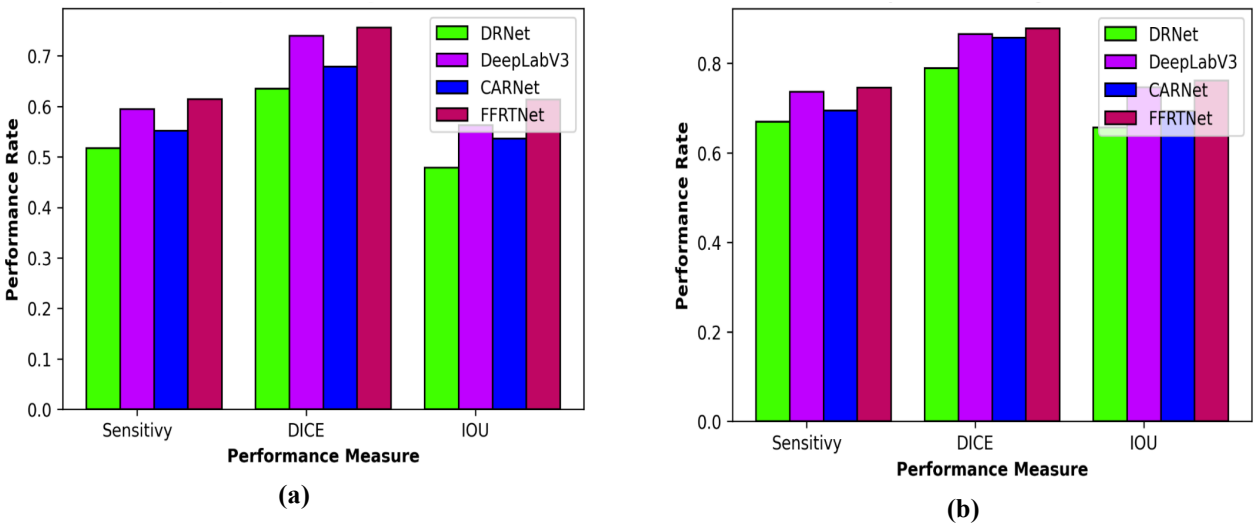


Fig 7. Performance Comparison of (a) MA Segmentation and (b) HA Segmentation.

In **Fig 7**, the segmentation of MA and HA is illustrated where the performance metrics, namely sensitivity, DICE, and IoU, are compared with existing methods, namely DRNet, DeepLabV3, and CARNet. The sensitivity of MA segmentation for FFRTNet is higher than 9.74%, 6.27%, and 1.99% for the techniques DRNet, DeepLabV3, and CARNet. The DICE of MA segmentation for FFRTNet is more elevated than 12.07%, 7.72%, and 1.63% for DRNet, DeepLabV3, and CARNet methods. The IoU of MA segmentation for FFRTNet is higher than 13.43%, 7.67%, and 5.03% for DRNet, DeepLabV3, and CARNet methods. The sensitivity of HA segmentation for FFRTNet is higher than 7.53%, 5.05%, and 0.88% for the techniques DRNet, DeepLabV3, and CARNet. The DICE of HA segmentation for FFRTNet is higher than 8.83%, 2.04%, and 1.27% for DRNet, DeepLabV3, and CARNet methods. The IoU of HA segmentation for FFRTNet is more elevated than 10.43%, 6.78%, and 1.47% for DRNet, DeepLabV3, and CARNet methods.s The higher rate of performance measures shows that the proposed system is highly effective. **Table 2** shows the performance comparison of segmentation of MA and HA

Table 3. Comparison of Performance of Segmentation of HE and SE

Feature	**HE**			**SE**		
Type of Segmentation	Sensitivity	DICE	IoU	Sensitivity	DICE	IoU
DRNet	0.5598	0.6914	0.5241	0.6995	0.8228	0.6990
DeepLabV3	0.4997	0.6576	0.4732	0.6518	0.7810	0.6407
CARNet	0.4932	0.6464	0.4666	0.6944	0.8068	0.6761
FFRTNet	0.5954	0.7348	0.5701	0.7542	0.8450	0.7365

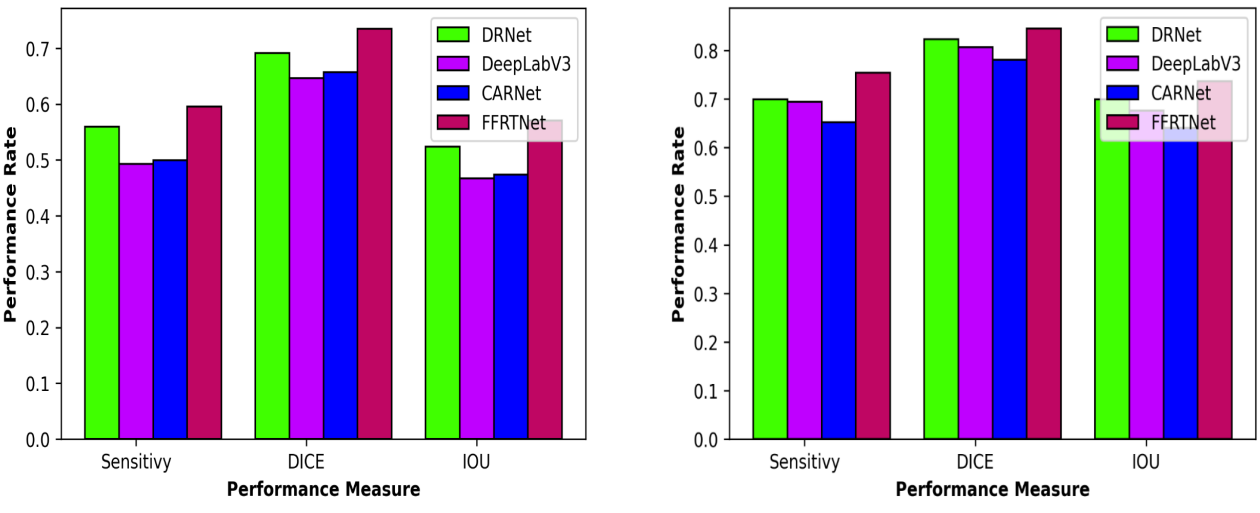


Fig 8. Performance Comparison of HE and SE Segmentation.

In **Fig 8**, the segmentation of HE and SE is demonstrated where the performance metrics, namely sensitivity, DICE, and IoU, are compared with existing methods, namely DRNet, DeepLabV3, and CARNet. The sensitivity of HE segmentation for FFRTNet is higher than 3.56%, 9.57%, and 10.22% for DRNet, DeepLabV3, and CARNet methods. The DICE of HE segmentation for FFRTNet is higher than 4.34%, 7.72%, and 8.84% for the methods DRNet, DeepLabV3, and CARNet. The IoU of HE segmentation for FFRTNet is higher than 4.6%, 9.69%, and 10.35% for DRNet, DeepLabV3, and CARNet methods. The sensitivity of SE segmentation for FFRTNet is higher than 5.47%, 10.24%, and 5.98% for the methods DRNet, DeepLabV3, and CARNet. The DICE of SE segmentation for FFRTNet is higher than 2.22%, 6.4%, and 3.82% for DRNet, DeepLabV3, and CARNet methods. The IoU of SE segmentation for FFRTNet is higher than 3.75%, 9.58%, and 6.04% for DRNet, DeepLabV3, and CARNet methods. The higher rate of performance measures shows that the proposed method is highly effective. **Table 3** shows the comparison of performance of segmentation of HE and SE

Table 4. Performance Comparison of Running Time

Type of Segmentation	Epoch 50	Epoch 100	Epoch 150	Epoch 200
DRNet	801	845	902	1011
DeepLabV3	1067	1100	1131	1176
CARNet	810	860	951	1091
FFRTNet	490	516	598	665

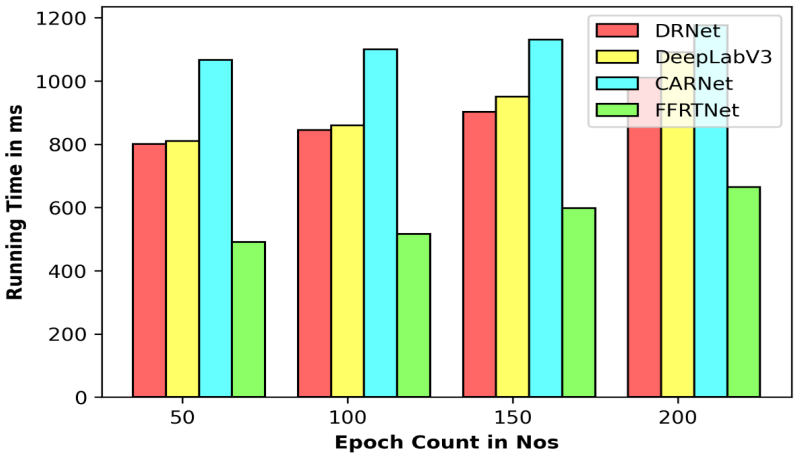


Fig 9. Performance Comparison of Running Time with DRNet, DeepLabV3, and CARNet.

From the observation of **Fig 9**, it is shown that the proposed FFRTNet achieves minimal running time. For epoch 250, FFRTNet attains 665 milliseconds (*ms*), less than existing *state-of-the-art* techniques. **Table 4** shows the performance comparison of running time.

V. CONCLUSION AND FUTURE WORK

Chronic Diabetic Retinopathy (DR) is a significant threat to diabetic patients, with its progression damaging the retina's blood vessels and causing abnormalities in the macular region. DR, if left untreated, can lead to blurred eye vision and blindness. This research uses a new network that simultaneously segments the four DR lesions using a dual-branch design with GTB and RTB integrated with MSFF. GTB and RTB examine the intra-class dependencies between inter-class and multi-lesion relationships of vessels and lesions. The investigation of vessel and lesion extraction is responsible for the network's experimental findingsLearning's coarse-grained pseudo masks generate the proposed network insufficiently. DR multi-lesion segmentation requires expert pixel-level annotations. FFRTNET outperforms DRNet, DeepLabV3, and CARNet in sensitivity, IoU, and DICE. FFRTNet achieves 665 *ms,* which is minimal compared to existing methods that indicate the effectiveness of the proposed method.

The proposed techniques deal with only FI. The RI from other medical modalities will be considered. The computation of measures like tortuosity and diameter from the segmented vessels will be included. Increasing the number of training and testing images may also improve the segmentation process.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding
No funding agency is associated with this research.

Competing Interests
There are no competing interests

References

[1]. S. Vujosevic et al., “Screening for diabetic retinopathy: new perspectives and challenges,” The Lancet Diabetes & Endocrinology, vol. 8, no. 4, pp. 337–347, Apr. 2020, doi: 10.1016/s2213-8587(19)30411-5.
[2]. Z. L. Teo et al., “Global Prevalence of Diabetic Retinopathy and Projection of Burden through 2045,” Ophthalmology, vol. 128, no. 11, pp. 1580–1591, Nov. 2021, doi: 10.1016/j.ophtha.2021.04.027.
[3]. D. A. Antonetti, P. S. Silva, and A. W. Stitt, “Current understanding of the molecular and cellular pathology of diabetic retinopathy,” Nature Reviews Endocrinology, vol. 17, no. 4, pp. 195–206, Jan. 2021, doi: 10.1038/s41574-020-00451-4.
[4]. S. Qummar et al., “A Deep Learning Ensemble Approach for Diabetic Retinopathy Detection,” IEEE Access, vol. 7, pp. 150530–150539, 2019, doi: 10.1109/access.2019.2947484.
[5]. L. Dai et al., “A deep learning system for detecting diabetic retinopathy across the disease spectrum,” Nature Communications, vol. 12, no. 1, May 2021, doi: 10.1038/s41467-021-23458-5.
[6]. L. A. Levin, M. Sengupta, L. J. Balcer, M. J. Kupersmith, and N. R. Miller, “Report From the National Eye Institute Workshop on Neuro-Ophthalmic Disease Clinical Trial Endpoints: Optic Neuropathies,” Investigative Opthalmology & Visual Science, vol. 62, no. 14, p. 30, Nov. 2021, doi: 10.1167/iovs.62.14.30.
[7]. T. R. Gadekallu, N. Khare, S. Bhattacharya, S. Singh, P. K. R. Maddikunta, and G. Srivastava, “Deep neural networks to predict diabetic retinopathy,” Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 5, pp. 5407–5420, Apr. 2020, doi: 10.1007/s12652-020-01963-7.
[8]. A. Grzybowski et al., “Artificial intelligence for diabetic retinopathy screening: a review,” Eye, vol. 34, no. 3, pp. 451–460, Sep. 2019, doi: 10.1038/s41433-019-0566-0.
[9]. J. V. Forrester, L. Kuffova, and M. Delibegovic, “The Role of Inflammation in Diabetic Retinopathy,” Frontiers in Immunology, vol. 11, Nov. 2020, doi: 10.3389/fimmu.2020.583687.
[10]. R. Cheloni, S. A. Gandolfi, C. Signorelli, and A. Odone, “Global prevalence of diabetic retinopathy: protocol for a systematic review and meta-analysis,” BMJ Open, vol. 9, no. 3, p. e022188, Mar. 2019, doi: 10.1136/bmjopen-2018-022188.
[11]. O. Simó-Servat, C. Hernández, and R. Simó, “Diabetic Retinopathy in the Context of Patients with Diabetes,” Ophthalmic Research, vol. 62, no. 4, pp. 211–217, 2019, doi: 10.1159/000499541.
[12]. Y. Zhou, B. Wang, L. Huang, S. Cui, and L. Shao, “A Benchmark for Studying Diabetic Retinopathy: Segmentation, Grading, and Transferability,” IEEE Transactions on Medical Imaging, vol. 40, no. 3, pp. 818–828, Mar. 2021, doi: 10.1109/tmi.2020.3037771.
[13]. N. Sambyal, P. Saini, R. Syal, and V. Gupta, “Modified U-Net architecture for semantic segmentation of diabetic retinopathy images,” Biocybernetics and Biomedical Engineering, vol. 40, no. 3, pp. 1094–1109, Jul. 2020, doi: 10.1016/j.bbe.2020.05.006.
[14]. S. Kumar, A. Adarsh, B. Kumar, and A. K. Singh, “An automated early diabetic retinopathy detection through improved blood vessel and optic disc segmentation,” Optics & Laser Technology, vol. 121, p. 105815, Jan. 2020, doi: 10.1016/j.optlastec.2019.105815.
[15]. M. U. Akram, S. Akbar, T. Hassan, S. G. Khawaja, U. Yasin, and I. Basit, “Data on fundus images for vessels segmentation, detection of hypertensive retinopathy, diabetic retinopathy and papilledema,” Data in Brief, vol. 29, p. 105282, Apr. 2020, doi: 10.1016/j.dib.2020.105282.
[16]. A. Garifullin, L. Lensu, and H. Uusitalo, “Deep Bayesian baseline for segmenting diabetic retinopathy lesions: Advances and challenges,” Computers in Biology and Medicine, vol. 136, p. 104725, Sep. 2021, doi: 10.1016/j.compbiomed.2021.104725.
[17]. P. Porwal, S. Pachade, M. Kokare, G. Deshmukh, J. Son, W. Bae, *et al.,* “IDRiD: Diabetic Retinopathy – Segmentation and Grading Challenge,” Medical Image Analysis, vol. 59, 2020, doi: 10.1016/j.media.2019.101561.
[18]. G. T. Zago, R. V. Andreão, B. Dorizzi, and E. O. Teatini Salles, “Diabetic retinopathy detection using red lesion localization and convolutional neural networks,” Computers in Biology and Medicine, vol. 116, p. 103537, Jan. 2020, doi: 10.1016/j.compbiomed.2019.103537.