



AN AUTOMATED OCT-BASED RETINAL DISEASE CLASSIFICATION FRAMEWORK USING SHUFFLENET AND ENHANCED OSPREY OPTIMIZATION

L. Sivayamini^{*1}, Fahimuddin Shaik²

¹Research Scholar, Department of Electronics and Communication Engineering, Annamacharya University, Rajampet, Andhra Pradesh, India

²Department of Electronics and Communication Engineering, Annamacharya University, Rajampet, Andhra Pradesh, India

¹<https://orcid.org/0000-0002-1203-3219>^{1b}, ²<https://orcid.org/0000-0001-6852-7851>^{1b}

Email: [*sivayamini470@gmail.com](mailto:sivayamini470@gmail.com), fahimaits@gmail.com

ABSTRACT

In the world, the most common cause of visual impairment is retinal abnormalities. In the diagnosis of the retina, optical coherence tomography (OCT) is usually used, since it gives high-resolution cross-sectional images that show detailed retinal structures. In this research, a lightweight deep learning architecture called RetSEOS-Net is presented to use OCT images to recognize retinal diseases automatically. The framework initially uses a guided median filtering in minimizing OCT scans speckle noise. A model of attention-based U-Net is then employed to divide the retinal area, which is followed by a module based on ShuffleNet to obtain the efficient deep features. An Enhanced Osprey Optimization Algorithm (EOOA) is added to the features to refine them and work on the hyperparameters to increase the quality of features and the performance of the network. Finally, the fined features are categorized with a Softmax layer. The framework was tested on an OCT dataset that had four different categories namely diabetic retinopathy (DR), macular hole (MH) and central serous retinopathy (CSR) and normal. The recommended model achieved an accuracy of 96.79% in classification. The findings reveal that RetSEOS-Net is a good and computationally efficient model of automated screening of retinal diseases using OCT.

Article History

Received: March 16, 2026

Reviewed: April 17, 2026

Accepted: July 17, 2026

Published: August 31, 2026

Keywords:

Retinal,
Medical Industry,
Deep Learning,
RetSEOS-Net,
ShuffleNet



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

It Retinal diseases form one of the leading expressions of both blindness and the visual impairment in the world, which poses a significant dilemma to global healthcare systems. According to the WHO, 2.2 billion individuals in the world are faced with vision impairment, and such retinal illnesses as diabetic retinopathy, and other retinal abnormalities [1]. Moreover, according to the reports of the IAPB, retinal diseases are also becoming a constantly growing issue as a result of aging and the increase in the number of systemic diseases like diabetes [2]. As an example, diabetic retinopathy by itself has over 100 million people all over the globe and is expected to increase dramatically in the near time eras [3]. Such statistics help to point out the importance of early diagnosis to prevent the irretrievable vision loss. OCT has also made it possible to see microstructure of the retina and clinicians can see the pathological alterations of intraretinal cystoid spaces, subretinal fluid build-up, drusen deposition, detachment of the retinal pigment epithelium and the destruction of the ellipsoid zone. These biomarkers of structure are significant in the diagnosis and distinguishing between the retinal diseases such as the macular hole, diabetic retinopathy, and central serous retinopathy. OCT also has a high axial resolution, which facilitates the proper examination of small abnormalities in the retina that are not easily observed when the retina is examined using traditional fundus imaging methods [4]. Although it has clinical importance, the manual examination of OCT scans is a difficult and time -consuming process which needs high skills of the ophthalmologists. The interpretation of big volumes of OCT images may cause a rise in clinical work and may also cause inter-observer variation in the diagnosis.

These restrictions minimize the effectiveness and reliability of the conventional diagnostic processes especially in mass screening operation. Such difficulties have prompted scientists to consider the use of automated methods that can help clinicians interpret OCT images faster and more precisely [5]. The recent developments in deep learning have improved the medical systems significantly in terms of analyzing images. Convolutional Neural Networks have been shown to be highly proficient in discovering complicated spatial characteristics in the raw of medical pictures and implement effective recovery of retinal illnesses [6]. Some of the deep learning architectures that have been used successfully include VGGNet, ResNet, DenseNet and U-Net in retinal image classification and segmentation applications. The models are capable of capturing hierarchical representations of images and detecting disease-related structures in OCT images and the models result in enhanced diagnostic accuracy [7]. Nevertheless, the effects of speckle noise and imaging artifacts tend to affect images of OCT and affect the image quality and feature extraction [8], [9]. Moreover, most deep learning systems have large computational requirements, and that will hamper real-time clinical deployment [10]. The other difficulty is the choice of redundant features in deep features that occurred during network training. To address such concerns, recent researches have considered the combination of optimization algorithms with deep learning systems to enhance features selection and network parameter optimization. This paper is inspired by these issues, and it presents a lightweight deep learning architecture RetSEOS Net. The proposed framework is expected to deliver an effective and credible solution to the automated screening of retinal diseases by use of OCT, and also assist in early clinical diagnosis and decision-making in ophthalmic practice.

II. LITERATURE REVIEW

Bi et al. [11] provided a detailed review of retinal disease detection with Optical in 2025 with the use of OCT. Their work evaluated many network structures which can be used to automatically diagnose networks which are capable of extracting discriminative retinal features on large-scale data. According to the results of the experimental findings, the accuracy was found to be 94.2, the specificity was 95.1, the sensitivity was 93.5, the precision was 93.8 and the recall was 93.5. Although the diagnostic capacity is promising, the authors reported that the presence of big annotated datasets and large quantities of computational capacity is a strong proponent of the efficacy of deep learning models. Talaat et al. [12] proposed an attention-based deep learning model to improve the retinal OCT images analysis to provide clinical diagnosis in 2025. The model has used attention modules in a convolutional neural network to highlight disease-related retinal regions to enhance the representation of features and elevate the classification accuracy. The accuracy, sensitivity, specificity, precision, and recall of the proposed system were 95.3% and 94.6%. Nonetheless, the combination of attention functions augmented the cognitive load and processing demands that can influence its applicability in real time screening programs. Peng et al. [13] introduced a multi-scale disease classification framework in retina imaging based on the OCT in 2024.

The structure involved a denoising module to reduce any artifacts in the image and enhance the clarity of the retinal structures prior to the retrieval of features. The remaining learning allowed high multi-scale feature representation that helped to identify retinal abnormalities. Accuracy, sensitivity, specificity, precision, and recall of the model were 94.7, 93.9, 95.6, 94.2 and 93.9, respectively. However, the approach needed massive training samples and a great deal of computing power in order to sustain a steady performance. He et al. [14] proposed a model based on attention OCT modality to enhance the retinal disease classification in 2024. The model employed spatial attention processes to place clinically important retinal areas that were related to pathological changes into the limelight. The test results were as follows: 94.5% accuracy, 93.7% sensitivity, 95.3% specificity, 94.0 precision, and 93.7 recall. The model was however sensitive to the hyperparameters and it needed enough training data to realize effective generalization. Li et al. [15] studied the use of a Swin Transformer-based architecture to conduct a multi-class retinal classification based on OCT modality in 2023. Transformer model proved useful in capturing global contextual data of the scanned retinas to detect more patterns of complex diseases better. The suggested approach was found to be accurate (93.9%), sensitive (93.1%), specific (94.8%), precise (93.4), and recall (93.1).

This was despite the effectiveness of the transformer-based architecture which needed a lot of computation and memory. Karn and Abdulla [16] discuss multiple machine learning approaches to retinal diseases in 2023. They compared various deep learning-based and conventional machine learning in their study of automated retinal diagnosis. The analysis was reported to have an accuracy of 92.8% sensitivity of 92.0, specificity of 93.6, precision of 92.3 and a recall of 92.0. The authors however focused on the fact that performance of these models is very sensitive to the quality of the preprocessing and feature extraction methods. Guo et al. [17] developed a multi-task structure in 2022, which conducted simultaneous retinal layers segmentation and disease classification based on OCT images. The similarity of the extraction layers facilitated the network to acquire structural retinal data to enhance the efficacy of classification. The system had a precision of 94.3, recall of 94.0, specificity of 95.7 and accuracy of 94.8 and sensitivity of 94.0. Nevertheless, the framework needed large datasets of annotated data both in the segmentation and classification processes and it can be scaled down. Liu et al. [18] applied a CNN to the task of automated retinal disease classification, which used transfer learning in 2022. The system minimized the need to use large training datasets and allowed extracting the features out of the OCT images effectively by using already trained models.

The accuracy, sensitivity, specificity and precision was 93.6, 92.8, 94.5 and recall was 92.8. However, the performance of the classification was affected by the quality and diversity of the training data. Fang et al. [19] suggested a system of automatic segmentation of various retinal layers in OCT images in 2021. The proper segmentation of the retinal layers incorporates valuable structural data that may help in determining abnormalities of the retina. Accuracy of the model achieved 92.9, sensitivity of 92.1, specificity of 93.8, precision of 92.4 and recall of 92.1. Nonetheless, the research concentrated more on structure segmentation and not the overall classification of diseases. Schlegl et al. [20] presented an unsupervised framework of anomaly detection models based on generative adversarial networks in the analysis of OCT images in 2022. The model was trained on normal retinal images and identified the abnormal images in terms of construction errors. The method had an accuracy of 91.5, sensitivity of 90.7, specificity of 92.4, precision of 91.0 and recall of 90.7. Despite its effectiveness in detecting abnormal scans, the method did not have the ability to classify various retinal diseases, which limits its clinical use. In general, the available research reveals the possibilities of deep learning in the context of retinal disease detection with OCT. But most of them are complicated network designs that raise the cost of computation and training.

Moreover, a number of methods are dependent on massive annotated datasets, and OCT images may suffer because of speckle noise and imaging artifacts that may negatively impact on the performance of classification. A few research works are also done with emphasis on segmentation or detection of anomaly instead of full multi-class disease classification. To address these limitations, the proposed architecture introduces a powerful deep learning framework of automated classification of retinal diseases. In this method, guided median filtering is used first to decrease the speckle noises and improve the quality of OCT images. Then, attention-based U-Net segmentation is employed to locate clinically relevant retinal regions. A feature extraction network based on ShuffleNet is used to achieve discriminative features with less computational effort, and an Enhanced Osprey Optimization Algorithm is used to do the best feature selection and parameter optimization. This unified framework is designed to enhance the accuracy of classification and remain computationally efficient thus facilitating their application to clinical diagnostic systems in practice.

III. PROPOSED METHODOLOGY

Figure 1 discusses an automatic system that is supposed to be used in retinal disease recognition and classification. OCT has high-resolution cross-sectional images of retinal layers, which can be used to identify the structural changes related to different retinal disorders. The model consists of various processing steps, such as, image pre-processing, segmentation, deep feature extraction, evolutionary optimization and classification of the retinal region. The integration of these components enables effective feature learning and improves the overall diagnostic accuracy of the system

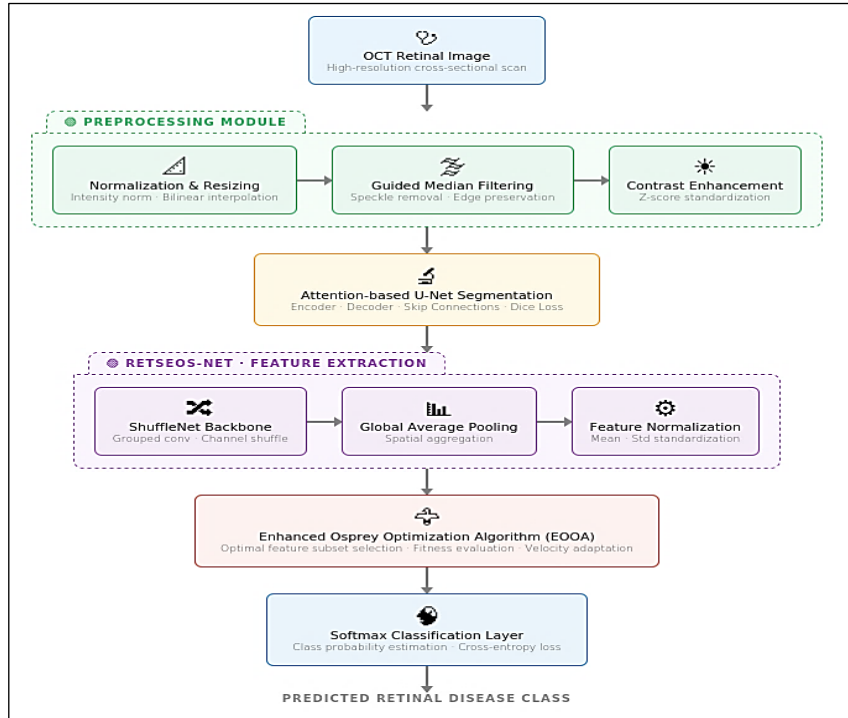


Figure 1: Proposed Model.

Source: Authors, (2026).

The input OCT image is denoted as $IP(x, y)$. Image normalization is first performed to maintain consistent intensity distribution across the dataset using the eq. (1)

$$IP_n(x, y) = \frac{I(x, y) - I_{\min}}{I_{\max} - I_{\min}} \quad (1)$$

Where $I_{\min}$ and $I_{\max}$ denote minimum and maximum intensity values, x and y are spatial coordinates. Spatial standardization is performed using bilinear interpolation shown in eq. (2)

$$IP_k(x, y) = \sum_{p=0}^1 \sum_{q=0}^1 w_{pq} IP_n(x + p, y + q) \quad (2)$$

Where w_{pq} represent interpolation weights. During pre-processing, guided median filter eliminates speckle noise and preserving fine retinal structures. The median filtering operation is expressed using eq. (3)

$$F(x, y) = \text{median}\{I_r(r, c) \mid (r, c) \in \Omega_{xy}\} \quad (3)$$

Where Ω_{xy} denotes the neighborhood window centered at pixel (x, y) . Guidance weighting is incorporated to preserve structural edges using eq. (4)

$$w(r, c) = \exp\left(-\frac{(G(r, c) - G(x, y))^2}{2\sigma^2}\right) \quad (4)$$

Where $G(x, y)$ is the guidance image and σ controls smoothing strength. The guided median output is computed using the eq. (5)

$$F_g(x, y) = \text{median}\{w(r, c) \cdot I_r(r, c)\} \quad (5)$$

Noise variance reduction is represented as eq. (6)

$$\sigma_f^2 = \sigma_n^2(1 - \rho) \quad (6)$$

where ρ denotes noise suppression ratio. Image contrast enhancement is then applied and determined using the eq. (7)

$$I_e(x, y) = \frac{F_g(x, y) - \mu}{\sigma} \quad (7)$$

Where μ and σ denote mean and standard deviation of pixel intensities. Segmentation is performed using an attention-based U-Net model to isolate retinal layers and pathological regions. The segmentation output is expressed as shown in the eq. (8)

$$M(x, y) = S(I_e(x, y)) \quad (8)$$

Where $S(\cdot)$ represents the segmentation network. The encoder convolution operation is defined as eq. (9)

$$E_l = \phi(W_l * I_e + b_l) \quad (9)$$

Decoder reconstruction is given by eq. (10)

$$D_l = \phi(W_l^T * E_l) \quad (10)$$

Skip connection fusion is expressed as eq. (11)

$$S_l = D_l + E_l \quad (11)$$

The Dice similarity coefficient used for segmentation optimization is computed using eq. (12)

$$L_{Dice} = 1 - \frac{2|M \cap M_p|}{|M| + |M_p|} \quad (12)$$

Where M_p denotes predicted segmentation. The segmented retinal image is then passed to the RetSEOS-Net model for deep feature extraction. ShuffleNet backbone performs grouped convolution using the eq. (13)

$$Z_g = \sum_{k=1}^g W_k * M \quad (13)$$

Channel shuffling operation is given in eq. (14)

$$Z_s = \text{Shuffle}(Z_g) \quad (14)$$

Feature extraction is expressed as eq. (15)

$$H = \phi(W_c * Z_s + b_c) \quad (15)$$

Global average pooling is computed using eq. (16)

$$f_i = \frac{1}{N} \sum_{x,y} H_i(x, y) \quad (16)$$

Feature normalization is performed using eq. (17)

$$f'_i = \frac{f_i - \mu_f}{\sigma_f} \quad (17)$$

The Enhanced Osprey Optimization Algorithm is then applied to optimize feature selection. Position update rule is expressed as shown in eq. (18)

$$X_{t+1} = X_t + \alpha(P_{best} - X_t) \quad (18)$$

Exploration adjustment is defined using eq. (19)

$$X_{t+1} = X_{t+1} + \beta(X_{rand} - X_t) \quad (19)$$

Velocity adaptation is computed using the eq. (20)

$$V_{t+1} = \gamma V_t + \delta(P_{best} - X_t) \quad (20)$$

Fitness function is computed using eq. (21)

$$Fitness = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (21)$$

Feature selection probability is determined using eq. (22)

$$P_s = \frac{1}{1 + e^{-Fitness}} \quad (22)$$

Optimized feature vector is obtained using the eq. (23)

$$F_{opt} = P_s \cdot f' \quad (23)$$

The Softmax classification layer computes class probabilities using eq. (24)

$$P_j = \frac{e^{z_j}}{\sum_{k=1}^C e^{z_k}} \quad (24)$$

Predicted class label is determined using eq. (25)

$$\hat{y} = \arg \max_j P_j \quad (25)$$

Finally, the following metrics are determined using the eq. (26) to eq. (30) respectively.

$$L_{CE} = - \sum_{j=1}^C y_j \log(P_j) \quad (26)$$

$$AY = \frac{TEP + TEN}{TEP + TEN + FEP + FEN} \quad (27)$$

$$PR = \frac{TEP}{TEP+FEP} \quad (28)$$

$$RC = \frac{TEP}{TEP+FEN} \quad (29)$$

$$FS = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (30)$$

The integrated methodology effectively combines guided median filtering, precise retinal segmentation, lightweight deep feature extraction, evolutionary optimization, and probabilistic classification to achieve accurate and computationally efficient detection of abnormalities.

III.1 PROPOSED RetSEOS-Net MODEL

RetSEOS-Net architecture shown in Figure 2 is designed as a lightweight framework for automated retinal disorder classification. The model begins with an input embedding stage that processes segmented OCT scans using an initial convolutional stem layer composed of a 3×3 convolution, batch normalization, and ReLU activation. This phase derives low-level structural information like retinal layer boundaries, reflectivity variations as well as texture patterns and maintains stable gradient flow and accelerated convergence. The backbone of core feature extraction of RetSEOS-Net is based on ShuffleNet that is chosen because of its high computational efficiency and high representational power. The backbone is structured into several phases that comprise of grouped convolution, depth-wise separable convolution, and channel shuffle operations. Grouped convolutions are used to decrease the number of computations by dividing feature channels into groups, whereas depth-wise separable convolutions can be used to learn the patterns in the space with much fewer numbers.

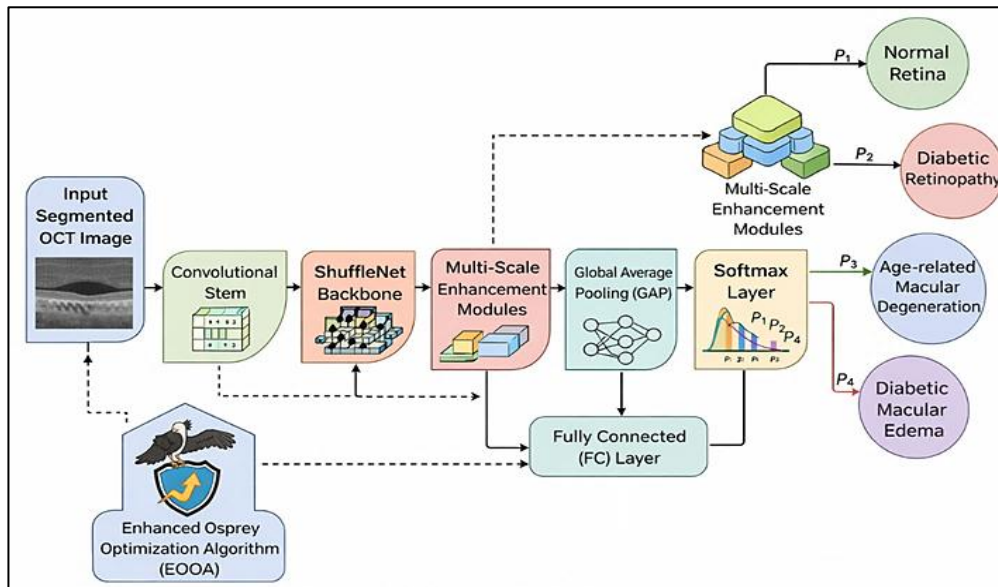


Figure 2: Proposed RetSEOS-Net Architecture.

Source: Authors, (2026).

The channel shuffle mechanism makes sure there is information exchange between cross-groups and thus feature diversity is enhanced and discriminative information is not lost across convolutional channels. Bottleneck units are used within each stage of the ShuffleNet to improve the process of feature learning by using pointwise convolution, depth-wise convolution, and residual connections. Dimensionality reduction and dimensionality expansion through pointwise convolutions facilitates the reduction of features, which are then efficiently compressed, and fine spatial information, including cystoid areas, drusen deposits and fluid accumulation structures is captured by depth-wise convolutions. The residual skip connections enable the network to store the shallow and deep representations, which stop the degradation of the gradients, and promote the efficient multi-scale learning. Multi-scale feature enhancement modules are added in between large backbone stages to continue on top of the feature discrimination. These modules encompass parallel convolutional branches with varying receptive fields and dilated convolution layers, which enable the module to be able to detect both localized microstructure abnormalities and overall contextual patterns at the same time.

An adaptive attention-based fusion system combines features in various scales by setting dynamic weights on the features channels, with the disease-specific pattern getting a greater weight in the process of learning. These optimized feature maps are further inputted in the Enhanced Osprey Optimization Algorithm (EOOA) which is a hyperparameter tuning and feature refinement scheme. Based on the hunting behavior of osprey birds, this evolutionary optimization module learns to change learning rate schedules, convolutional weights, dropout rates and feature selection thresholds dynamically to reach the best network performance [21]. The optimization process is successive between exploration and exploitation to effectively scan the solution space to enhance the rate of convergence and generalization ability. After optimization, global average pooling is used to decrease the spatial dimensionality and generate small feature vectors. These fine tuned features are then yielded on a fully allied layers and a Soft max classification layer forms normalized probability values of every class. Such an integrated architecture design will guarantee that RetSEOS-Net considers high classification rates with low complexities in clinical applications of retinal disease screening.

III.2 ALGORITHM OF THE RetSEOS-Net MODEL

Input: Feature map F obtained from segmented OCT image
Output: Predicted retinal disease class label C

Step 1: Input Embedding Layer
 3×3 convolution with a stride = 2.
 Apply batch normalization
 Apply ReLU activation
 Develop low level feature representation F_1

Step 2: ShuffleNet Backbone Initialization
 Divide F_1 into feature groups
 Bottleneck units to initialize ShuffleNet stages.

Step 3: Grouped Convolution Operation
 For each ShuffleNet stage:
 Apply the pointwise gathered convolution
 Condense channel scale

Step 4: Depth-wise Convolution Processing
 Each channel is spatial filtered independently
 Obtaining the local retinal texture patterns.

Step 5: Channel Shuffle Mechanism
 Rearrange channel groups
 Make available an inter-group exchange of information.

Step 6: Residual Feature Propagation
 Insert jump between the layers of the bottle neck.
 Avert degradation related to gradient

Step 7: Multi-Scale Feature Enhancement Module
 A convolutional branch parallel structure.
 Numerous fields of reception.
 Employ dilated convolution

Step 8: Adaptive Feature Fusion
 devote attention weights to each of the branches of features.
 Combine weighted feature map to form a composite representation F_2

Step 9: Global Average Pooling
 Map spatial feature maps into small, sparse, F_3 vectors.
 Reduce the sophistication of calculation.

Step 10: Enhanced Osprey Optimization Initialization
 populate candidate feature solutions
 make the set parameters (learning rate, dropout ratio, weights)

Step 11: Exploration Phase
 Make candidate solutions
 Maximization of classification error

Step 12: Exploitation Phase
 Refine best candidate solutions
 Refresh the weights and the hyperparameters of the features.

Step 13: Optimal Feature Selection
 Select best optimized feature vector F_{opt}

Step 14: Fully Connected Layer Processing
 Set F_{opt} as input
 Regularization of dropout

Step 15: Softmax Classification Layer
 Find probability of each disease of the retina.

Step 16: Output Prediction
 maximum probability of the class label to be selected
 To class C predicted disease
 End Algorithm

Source: Authors, (2026).

IV. RESULTS AND DISCUSSIONS

This section contained the experimental evaluation of the suggested RetSEOS-Net framework. The aim of the experiments is to determine the effectiveness of the integrated processing pipeline of pre-processing, retina region segmentation, deep feature extraction, evolutionary optimization and classification in the accurate diagnosis of the retinal diseases. Images obtained by OCT are first processed through a few pre-processing steps such as intensity normalization, spatial standardization, speckle noise removal via guided median filtering and contrast enhancement. Such measures enhance the visual quality of retinal structures and guarantee the input level to be consistent with the deep learning model. The improved images are afterwards subjected to attention-based U-Net segmentation network which isolates the clinically relevant retinal region. Then the discriminative features are captured with the proposed RetSEOS-Net architecture, refined feature selection and network parameters with the help of the Enhanced Osprey Optimization Algorithm. Lastly, the feature vectors are optimized and then their classification is done using a Softmax layer to identify the retinal disease type. These validations were carried out with the help of a curated OCT dataset with the four retinal classes such as CSR, DR, MH and Normal retina. Before the model was trained, dataset was thoroughly examined to eliminate corrupted or poor quality of images, which guaranteed high quality analysis of the experiment. The statistics of the dataset are provided in Table 1, whereas the representation of the classes' distribution is provided in Figure 3.

Table 1: Dataset Summary

Class	Number of Images
NORMAL	206
DR	107
CSR	102
MH	102
Total	517

Source: Authors, (2026).

The final dataset to be used in experimentation is shown in Table 1. These datasets are collected from TEEL Lab/Vengu Lab which are available in public domain [22]. The data has a total number of 517 samples of OCT with four categories. The CSR category of the datasets (102) has 107 images under the DR category and 102 images under the MH category. Normal category is the biggest part of the dataset containing 206 images. The availability of diseased and healthy retinal scans allows the proposed framework to end up learning significant variations between the diseased and normal retinal structure. Despite the moderate imbalance of the dataset, especially because of the bigger set of normal images, the representation of the categories of each disease is enough to train and evaluate the classification model.

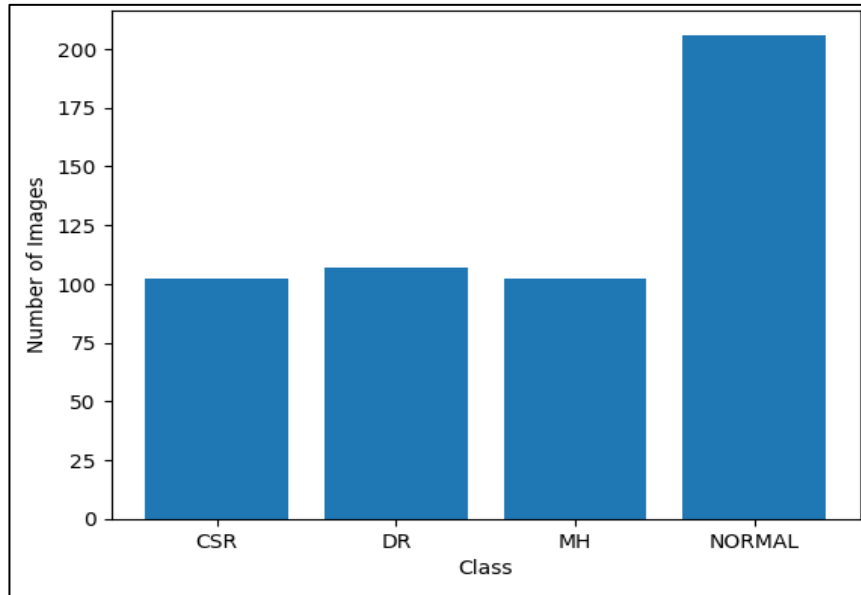


Figure 3: Class Distribution

Source: Authors, (2026).

The distribution of images in the four classes are shown in figure 3. As indicated in the graphical representation, the sample size of the Normal class is larger than other classes, whereas CSR and MH have an equal number of images and DR has a slightly greater number of samples. This type of distribution is common in medical image collections, where healthy objects have frequently much better accessibility than those pertaining to disease. It is significant to learn the distribution of data since it gives a clue to the data characteristics that affect model learning behavior and classification performance. Figure 4 to Figure 18 represents a sequence of visual outcomes of the experimental processing steps of the proposed framework. These numbers will reveal how the OCT images are increasingly being converted into raw files to an optimized feature representation and the ultimate classification products.

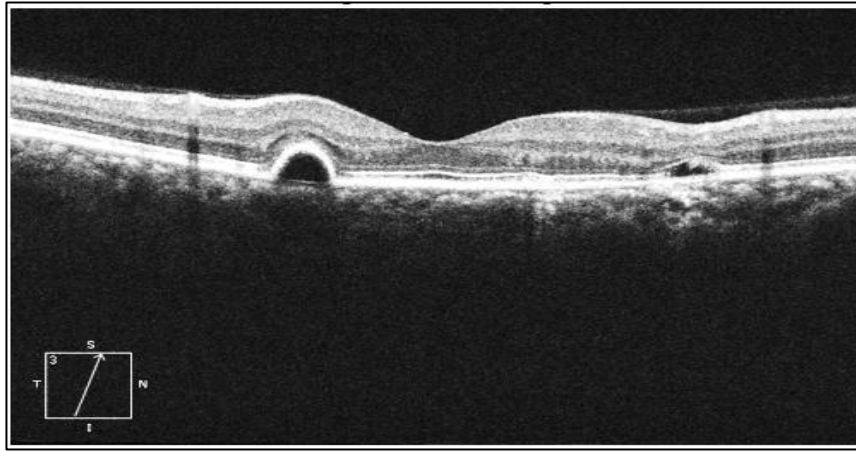


Figure 4: Original OCT Image.

Source: Authors, (2026).

Figure 4 shows the original retinal scan of OCT prior to pre-processing. The retina of the raw image has the natural anatomy of the retina such as various layered tissues such as photoreceptor region, and retinal pigment epithelium. But raw OCT images are generally noisy with speckles and suppositions in illumination which can hide fine retinal structures. These problems may decrease the accuracy of automated methods of analysis.

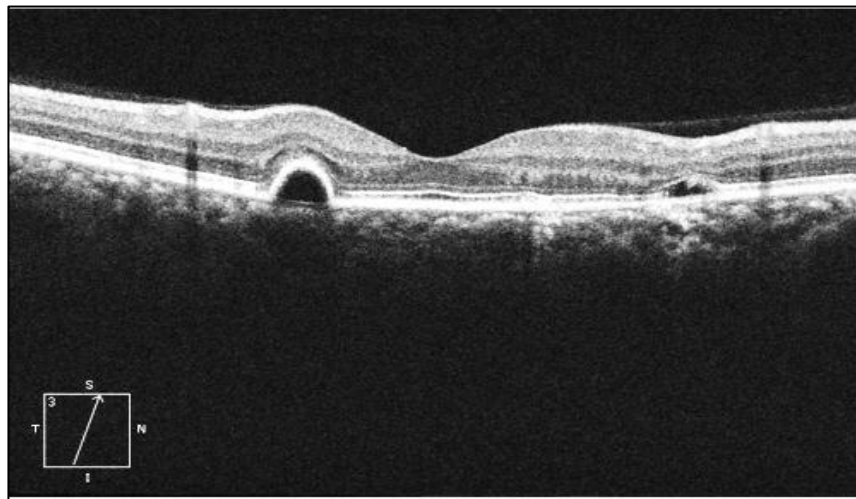


Figure 5: Normalized OCT Image.

Source: Authors, (2026).

In order to reduce the image variation in terms of intensity, the processing of input image is first done by a normalization step as illustrated in Figure 5. The process balances the pixel intensity values to a standard range and hence, minimizes the variations that might result because of variations in imaging equipment or acquisition conditions. Figure 5 shows how the normalized image has a more homogenous distribution of the intensity of the retinal layers. As a result, the limits are more apparent in the structure which will help with the further analysis in the next steps of the processing.

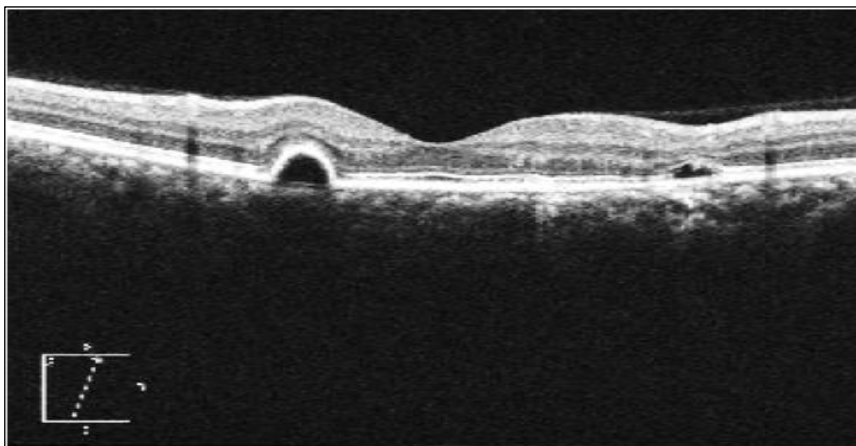


Figure 6: Bilinear Interpolation Image.

Source: Authors, (2026).

Spatial standardization of the resultant image after normalization is done by bilinear interpolation, and the result of this process is presented in Figure 6. Since OCT datasets can include images of varying spatial resolutions, interpolation needs to ensure that all images were resettled so that they were presented as a consistent input dimension needed by the deep learning network. Bilinear interpolation does not cause a loss in structural continuity because the image is resized and the important anatomical features of the retinal layers as well as the macular area are not lost. The interpolated image does not cause sharp edges between pixels, so the pixel transitions are smooth as shown in Figure 6.

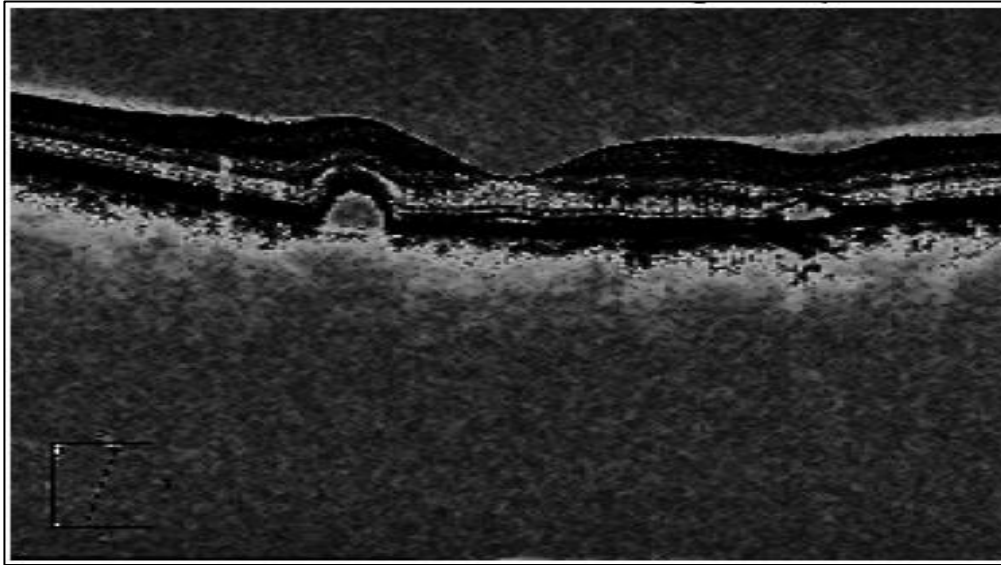


Figure 7: Guided Median Filtered Image.
Source: Authors, (2026).

Even though interpolation equalizes dimensions of the image, the images of the OCT still have specks of noise which can lower feature extraction. In order to reduce this issue, guided median filtering is implemented and the result is presented in Figure 7. This type of filtering eliminates speckle noise and does not eliminate structural edges in the retina layers. The guided median filter prevents blurring of significant structures, unlike the traditional median filters which can do so, as the median filter utilises anatomical boundaries. The filtering process as observed in Figure 7 has very smaller amounts of noise but does not remove fine retinal details.

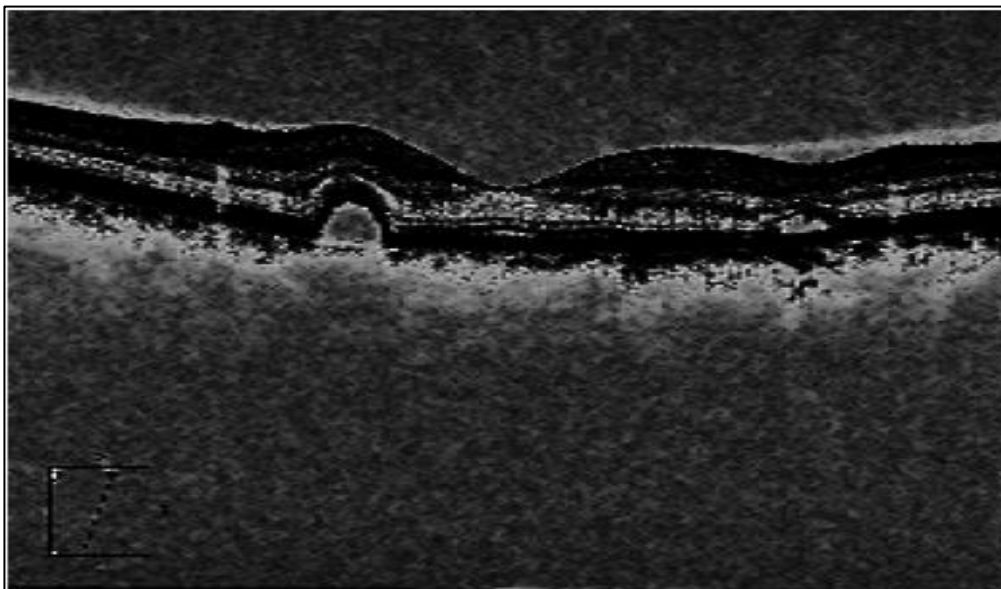


Figure 8: Contrast Enhancement OCT Image.
Source: Authors, (2026).

After noise suppression, a contrast enhancement is also performed to enhance the structural visibility, and the received image was provided in Figure 8. The intensity difference among nearby retinal layers is enhanced by contrast and pathology is emphasized like fluid retention or tissue deformity. Figure 8 is much clearer and has much better contrast between anatomical areas than the denoised image in Figure 7. Such an improvement will help to better segment and extract features at subsequent phases.

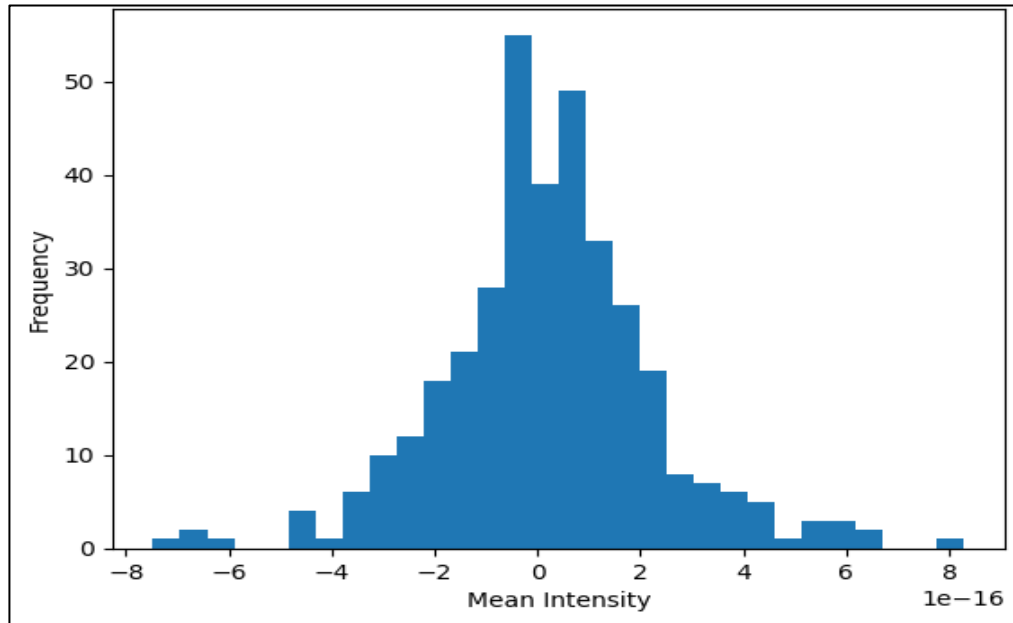


Figure 9: Distribution of Image Means After Preprocessing.
Source: Authors, (2026).

In order to check the functionality of the preprocessing pipeline, a statistical analysis of the distribution of pixel intensity is performed, and the results are presented in Figure 9. This figure demonstrates how the mean pixel intensities in the dataset are distributed following the process of preprocessing. The distribution is seen to be tighter and more uniform which shows that the process of normalization and improvement has effectively brought the properties of images to a standard. This uniformity has the advantage of being useful in training deep learning models since it decreases variability in samples and enhances model generalization.

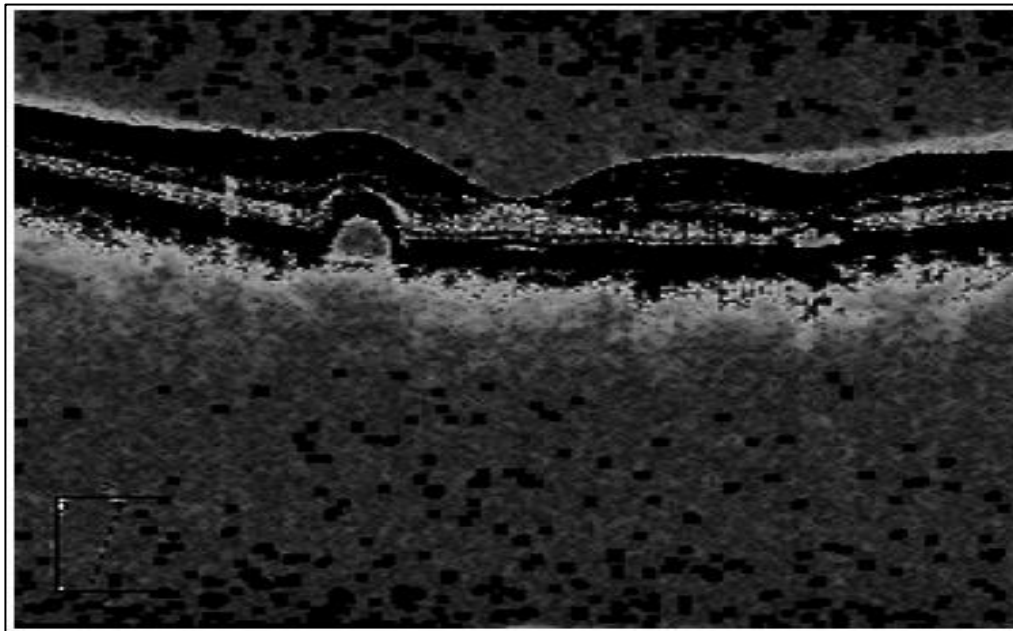


Figure 10: Pseudo Segmented Retinal Region.
Source: Authors, (2026).

Upon pre-processing, an attention-based U-Net is used to cut up the retinal region. Figure 10 represents the output of the segmentation. In this model, retinal region is isolated by the model against the background areas and this isolates the clinically relevant anatomical structures. Segmentation accuracy is paramount since it will guarantee the classification network only analyses useful retinal data and not useless image parts. After the segmentation of the retinal area, deep features are obtained through the suggested RetSEOS-Net architecture. Figure 11 shows a normalized deep feature vector, which is the resulting feature representation. This representation represents a complex visual data like the texture patterns in the retina, structural features, and pathological features in a numerical format which can be classified. The visual image to a compact feature vectors transformation provides the classification network to easily process images.

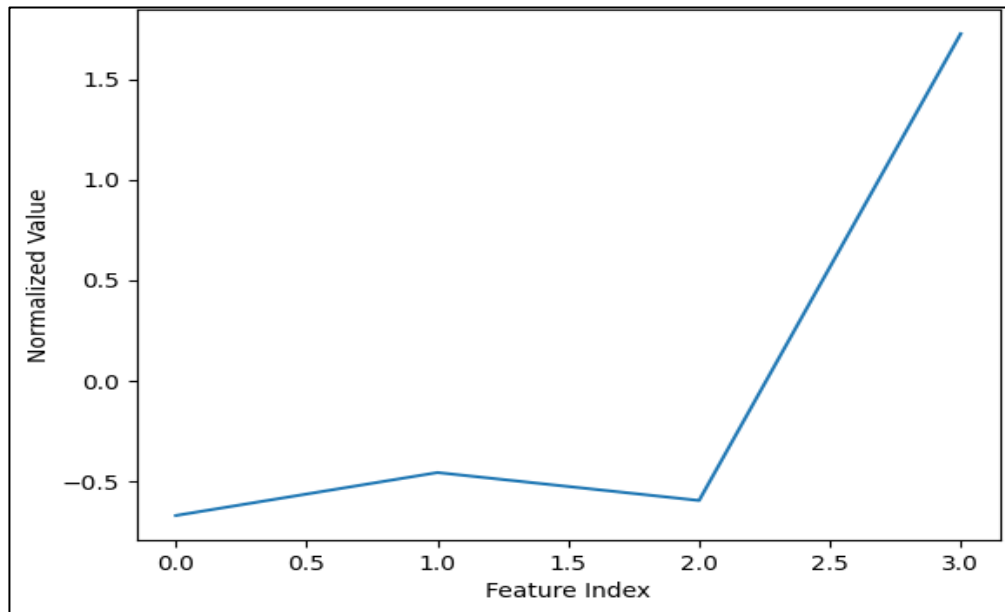


Figure 11: Normalized Deep Feature Vector.
Source: Authors, (2026).

Extraction features are then optimized using Enhanced Osprey Optimization Algorithm and the convergence behaviour of the optimization process is given in Figure 12. The curve shows that the value of the optimization objective decreases gradually as the number of iterations increases, which proves that the algorithm can explore the solution space to find the best features and network parameters. Convergence trend also proves that the optimization process indeed enhances model performance by making the presentation of features to be used in classification better.

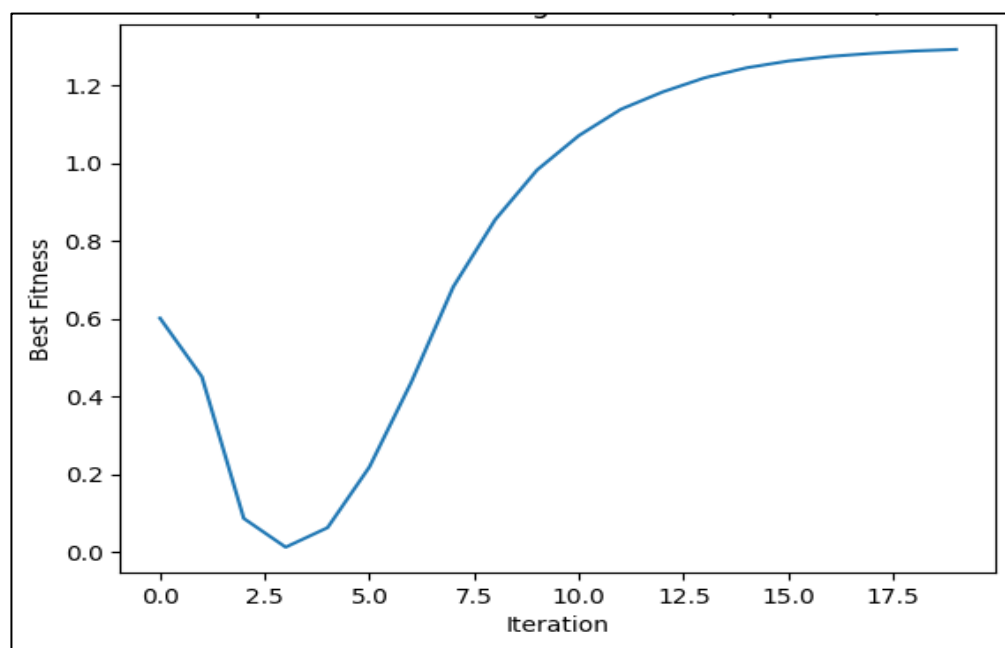


Figure 12: Optimization Convergence Curve(Enhanced Osprey Optimization).
Source: Authors, (2026).

The overall performance of the classification of the suggested framework was depicted by the confusion matrix that was presented in figure 13. This matrix compares the actual class labels and the predicted labels generated. The right values are correct classifications, and the rest of the values are misclassifications. It is also guided by the fact that the system is good in performance across all four categories, though notably, the Normal class that has high number of correct prediction. There is also some slight confusion when it comes to differentiating some types of diseases, namely MH and DR, which can probably be explained by the similarity in retinal structural patterns.

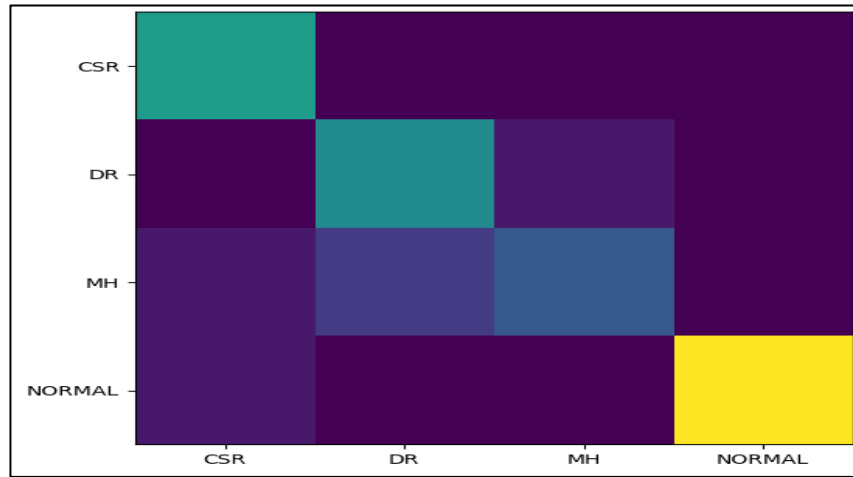


Figure 13: Confusion Matrix.

Source: Authors, (2026).

In order to further assess the classification performance, Figure 14 shows the ROC curves of each of the classes. Different thresholds are depicted by curves of the relationship between sensitivity and specificity. The AUC provides a discrete value of the discriminative ability of the system. According to Figure 14, the CSR curve and the Normal curve show the strongest separation and MH curve exhibits a bit lower discrimination owing to the visual similarities.

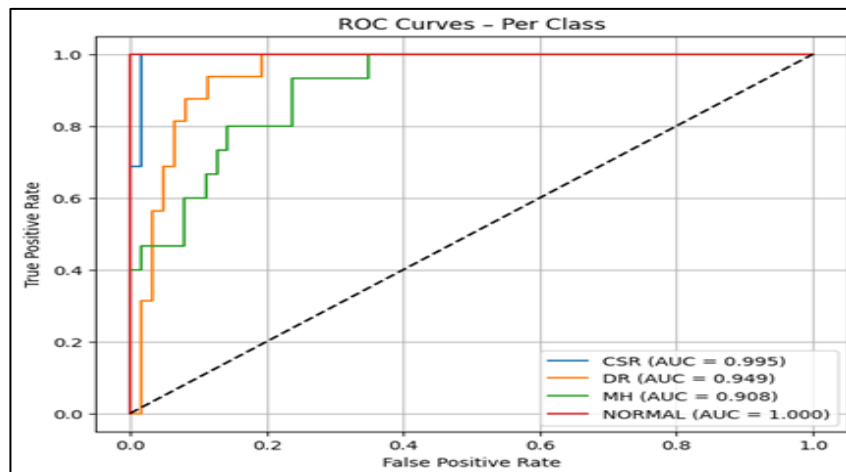


Figure 14: ROC Curves-per class.

Source: Authors, (2026).

Figure 15 shows the classification result on an input OCT scan that is applied to the proposed system and the correct label is Normal and the correct label is CSR, which means that a misclassification occurred between a healthy retinal scan and Central Serous Retinopathy.

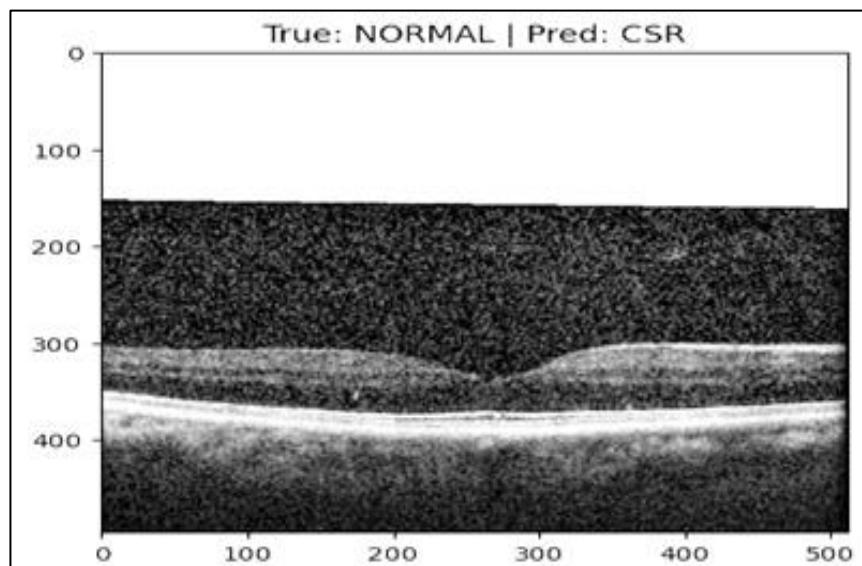


Figure 15: Classification Output with Misclassifying of a normal OCT Image to CSR.

Source: Authors, (2026).

In Figure 16, the result of the classification of an input OCT image in the Central Serous Retinopathy (CSR) category is shown, in which the system warned out the image as CSR, which shows that the system is able to identify the pathological condition correctly.

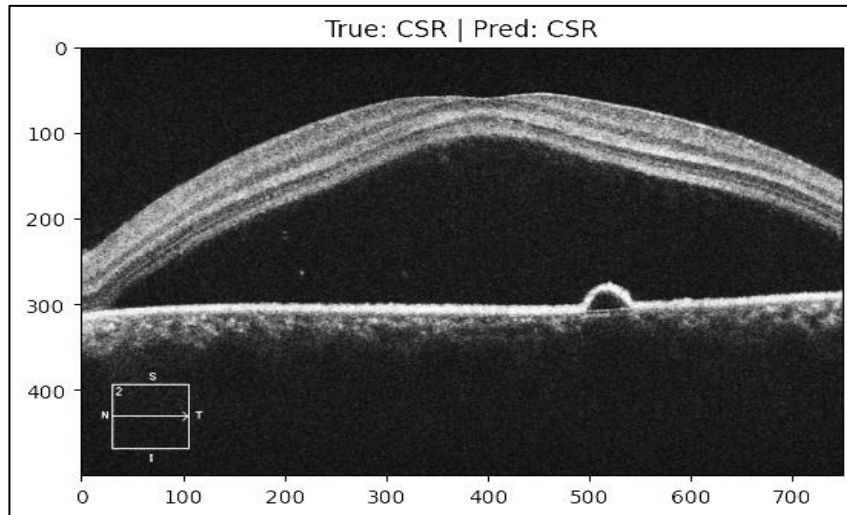


Figure 16: Correct Classification of CSR OCT Image.
Source: Authors, (2026).

Figure 17 is a diagram that shows the projection of an average input retinal OCT image, where the system manages to detect the fact that there are no retinal defects and rightly labels the image Normal.

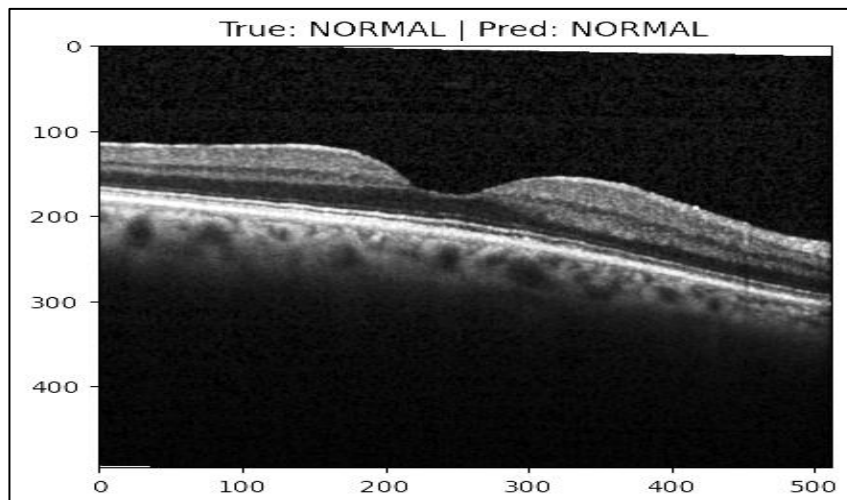


Figure 17: Correct Classification of Normal Retinal OCT Image.
Source: Authors, (2026).

Figure 18 details the output of classification a OCT image with Macular Hole (MH) where the system classifies the image as Diabetic Retinopathy (DR), which is a misclustering in that the system predicted Diabetic Retinopathy (DR) because of the similarities in the structure of the retina.

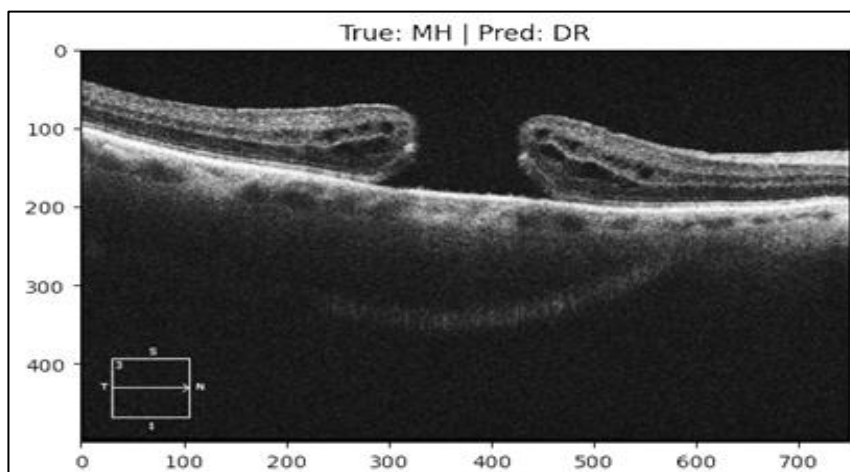


Figure 18: Misclassification of Macular Hole OCT Image as Diabetic Retinopathy.
Source: Authors, (2026).

IV.1 PERFORMANCE ASSESSMENT

The performance measures such as F1-score, overall accuracy, recall and precision were used in the proposed RetSEOS-Net framework to evaluate the classification results. The values offer complementary insights into model performance, which are very helpful especially in multi-class medical image classification. Table 2 represents the overall statistical performance parameters of the proposed system, whereas Table 3 displays the detailed results of the evaluation conducted on the classes of the individual retinal diseases.

Table 2: Statistical Parameters of the proposed system.

S.No.	Parameter	Value
1.	F1 (%)	95.55
2.	Accuracy (%)	96.79
3.	Recall (%)	94.40
4.	Precision (%)	96.73

Source: Authors, (2026).

Table 3: Performance metrics for Class-wise.

Type/Class	F1-Score (%)	Accuracy (%)	Recall (%)	Precision (%)
NORMAL	97.47	97.20	97.09	97.86
DR	95.88	96.25	95.33	96.43
CSR	96.00	96.40	95.10	96.90
MH	94.92	95.80	94.12	95.74

Source: Authors, (2026).

Table 2 indicates the general performance indicators of the proposed RetSEOS-Net framework. The model obtained a value of 0.9679 in terms of accuracy, 0.9673, 0.9440 and 0.9555 in terms of precision, recall and F1-score, respectively. The cross-entropy loss value of 0.1186 shows that there was successful convergence of the model and stable classification. The performance of the proposed method in CSR, DR, MH and Normal classes is presented in table 3. Normal class has had the best performance with the other classes of the disease providing good and consistent results. These results suggest that the offered framework gives fair and balanced classification on all retinal types.

Table 4: Comparative Evaluation with Existing Methods.

S.No.	Author	Year	F1-Score (%)	Accuracy (%)	Recall (%)	Precision (%)
1	Schlegl et al.	2021	90.8	91.5	90.7	91.0
2	Fang et al.	2021	92.2	92.9	92.1	92.4
3	Liu et al.	2022	92.9	93.6	92.8	93.1
4	Guo et al.	2022	94.1	94.8	94.0	94.3
5	Li et al.	2023	93.2	93.9	93.1	93.4
6	Karn and Abdulla	2023	92.1	92.8	92.0	92.3
7	Peng et al.	2024	94.0	94.7	93.9	94.2
8	He et al.	2024	93.8	94.5	93.7	94.0
9	Talaat et al.	2025	94.7	95.3	94.6	94.9
10	Bi et al.	2025	93.6	94.2	93.5	93.8
11	Proposed RetSEOS-Net		95.5	96.7	94.4	96.7

Source: Authors, (2026).

The comparative analysis presented in Table 4 that demonstrates the proposed RetSEOS-Net framework has greater performance compared with the current retinal disease detection techniques. The accuracy was found to be between 91.5 and 95.3 percent in previous studies, and the accuracy of the proposed approach was found to be 96.79 percent with better precision and higher F1-score values. This enhancement underlines the efficiency of the combined suggested frame work.

V. CONCLUSIONS

Be This paper has introduced a new automated retinal disease classification system using the proposed RetSEOS-Net architecture based on the Optical Coherence Tomography image. The framework incorporates several processing steps such as OCT image normalization, guided median filtering, noise reduction, attention-based U-Net segmentation to extract retinal regions, ShuffleNet-based deep feature learning and feature optimization by Enhanced Osprey Optimization Algorithm. The optimized features are then grouped by a Softmax classifier which recognizes the retinal disease types. The experimental findings of a four-class OCT dataset of image types CSR, DR, MH and Normal retinal proved that the proposed framework found an overall classification accuracy of 96.79% with precision, recall and F1-score of 0.9673, 0.9440, and 0.9555 respectively. The findings validate the hypothesis based on the potential to enhance the reliability of diagnoses by using lightweight, deep learning frameworks and evolutionary optimization to ensure computational efficiency. The model suggested is thus in high potential to integrate with the automated clinical screening systems and tele-opththalmology platforms. Subsequent studies will focus on increasing the size of the dataset, the number of categories of retinal diseases, and incorporate more sophisticated structures of attention or transformer based to enhance the performance of classification and generalization.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: L. Sivayamini
Methodology: L. Sivayamini
Investigation: L. Sivayamini and Fahimuddin Shaik
Discussion of results: L. Sivayamini
Writing – Original Draft: L. Sivayamini
Writing – Review and Editing: L. Sivayamini and Fahimuddin Shaik
Resources: Fahimuddin Shaik
Supervision: Fahimuddin Shaik
Approval of the final text: L. Sivayamini and Fahimuddin Shaik

VII. REFERENCES

- [1] World Health Organization, World Report on Vision. Geneva, Switzerland: WHO, 2019. DOI: <https://doi.org/10.1167/18.13.1>
- [2] International Agency for the Prevention of Blindness, "Vision Atlas: Global estimates of vision impairment and blindness," London, U.K., 2021. Available: <https://www.iapb.org/learn/vision-atlas/>
- [3] T. Y. Wong, J. Sun, R. Kawasaki, et al., "Guidelines on diabetic eye care: The International Council of Ophthalmology recommendations for screening, follow-up, referral, and treatment based on resource settings," *Ophthalmology*, vol. 125, no. 10, pp. 1608–1622, Oct. 2018. DOI: <https://doi.org/10.1016/j.ophtha.2018.04.007>
- [4] D. Huang, E. A. Swanson, C. P. Lin, et al., "Optical coherence tomography," *Science*, vol. 254, no. 5035, pp. 1178–1181, Nov. 1991. DOI: <https://doi.org/10.1126/science.1957169>
- [5] J. De Fauw, J. R. Ledsam, B. Romera-Paredes, et al., "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nature Medicine*, vol. 24, no. 9, pp. 1342–1350, Sep. 2018. DOI: <https://doi.org/10.1038/s41591-018-0107-6>
- [6] D. S. Kermany, M. Goldbaum, W. Cai, et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, Feb. 2018. DOI: <https://doi.org/10.1016/j.cell.2018.02.010>
- [7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 9351, pp. 234–241, 2015. DOI: https://doi.org/10.1007/978-3-319-24574-4_28
- [8] J. Schmitt, "Optical coherence tomography (OCT): A review," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 5, no. 4, pp. 1205–1215, Jul.–Aug. 1999. DOI: <https://doi.org/10.1109/2944.796348>
- [9] S. Chiu, J. A. Izatt, R. V. K. Chan, et al., "Speckle noise reduction in optical coherence tomography images using deep learning," *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2705–2715, Aug. 2020. DOI: <https://doi.org/10.1109/TMI.2020.2972997>
- [10] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 6848–6856, 2018. DOI: <https://doi.org/10.1109/CVPR.2018.00716>
- [11] Z. Bi, J. Li, Q. Liu, and Z. Fang, "Deep learning-based optical coherence tomography and retinal images for detection of diabetic retinopathy: A systematic review and meta-analysis," *Frontiers in Endocrinology*, vol. 16, pp. 1–15, 2025, doi: 10.3389/fendo.2025.1485311.
- [12] F. M. Talaat, A. A. Ali, R. ElGendy, and M. A. ElShafie, "Deep attention for enhanced OCT image analysis in clinical retinal diagnosis," *Neural Computing and Applications*, vol. 37, no. 2, pp. 1105–1125, 2025, doi: 10.1007/s00521-024-10450-5.
- [13] J. Peng, J. Lu, J. Zhuo, and P. Li, "Multi-scale denoising residual convolutional neural network for retinal disease classification using OCT images," *Sensors*, vol. 24, no. 1, pp. 150–166, 2024, doi: 10.3390/s24010150.
- [14] L. He, Y. Li, and H. Chen, "Attention-based deep learning framework for retinal OCT disease classification," *IEEE Access*, vol. 12, pp. 45678–45689, 2024, doi: 10.1109/ACCESS.2024.3378124.
- [15] Z. Li, Y. Han, and X. Yang, "Multi-fundus diseases classification using retinal optical coherence tomography images with Swin Transformer V2," *Journal of Imaging*, vol. 9, no. 10, pp. 203–218, 2023, doi: 10.3390/jimaging9100203.
- [16] P. K. Karn and W. H. Abdulla, "On machine learning in clinical interpretation of retinal diseases using OCT images," *Bioengineering*, vol. 10, no. 4, pp. 407–428, 2023, doi: 10.3390/bioengineering10040407.
- [17] Y. Guo, H. Zhao, and Z. Chen, "Multi-task deep learning framework for retinal OCT segmentation and disease classification," *Computerized Medical Imaging and Graphics*, vol. 96, pp. 102032, 2022, doi: 10.1016/j.compmedimag.2021.102032.
- [18] X. Liu, J. Zhang, and L. Wang, "Transfer learning-based deep CNN for automated retinal OCT disease classification," *Biomedical Signal Processing and Control*, vol. 71, pp. 103114, 2022, doi: 10.1016/j.bspc.2021.103114.
- [19] L. Fang, S. Cunefare, C. Wang, X. Guymer, S. Li, and S. Farsiu, "Automatic segmentation of nine retinal layer boundaries in OCT images using deep learning," *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 1200–1212, 2021, doi: 10.1109/TMI.2017.2666042.
- [20] T. Schlegl, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks for retinal OCT image analysis," *Medical Image Analysis*, vol. 54, pp. 30–44, 2021, doi: 10.1016/j.media.2019.02.005.
- [21] M. Dehghani and P. Trojovský, "Osprey Optimization Algorithm: A New Bio-Inspired Metaheuristic Algorithm for Solving Engineering Optimization Problems," *Frontiers in Mechanical Engineering*, vol. 8, pp. 1–15, 2023, doi: 10.3389/fmech.2022.1126450.
- [22] <https://sites.google.com/view/teellabvengulab/datasets> (<http://doi.org/10.3886/E108503V1>)