International Research Journal of Multidisciplinary Scope (IRJMS), 2026;7(3):153-172 IRJMS

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2026.v07i03.09496

A Multimodal Deep Learning Framework for Optimized Detection and Classification of Diabetic Retinopathy

Peddapullaiahgari Hariobulesu[1], Fahimuddin Shaik[2]*

[1]Department of Electronics and Communication Engineering, Jawaharlal Nehru Technological University, Anantapur, Ananthapuramu andhra Pradesh, India, [2]Department of Electronics and Communication Engineering, Annamacharya Institute of Technology and Sciences (Autonomus), Rajampet, Andhra Pradesh, India. *Corresponding Author's Email: fahimaits@gmail.com, sfn@aitsrajampet.ac.in

Abstract

Diabetic retinopathy (DR) is a leading cause of vision impairment, requiring accurate early detection to prevent irreversible blindness. This study proposes DiabFoRMaxNet, a multimodal deep learning framework for optimized DR detection and classification using retinal fundus and optical coherence tomography angiography (OCTA) images. The framework integrates five major stages: preprocessing, feature extraction, feature segmentation, feature optimization and classification. Preprocessing combines Z-score normalization, green channel extraction, adaptive histogram equalization and bilateral filtering to enhance retinal image quality. AlexNet is employed for hierarchical feature extraction, while Fully Convolutional Networks (FCN) perform lesion-aware pixel-wise segmentation. The Fossa Optimization Algorithm (FOA) refines extracted features by reducing redundancy and RankMax classification improves robustness under class-imbalanced conditions. Using an augmented multimodal dataset expanded from 222 original images under controlled augmentation protocols, DiabFoRMaxNet classifies NO DR, MILD DR and MODERATE DR categories with high precision. Experimental evaluation demonstrated superior performance over conventional deep learning models, achieving 99.18% accuracy for retinal fundus images and 99.69% accuracy for OCTA images, along with strong sensitivity, specificity, precision and F1-score. The proposed multimodal architecture enhances lesion localization, computational efficiency and diagnostic reliability, indicating strong potential for scalable AI-assisted early DR screening and clinical ophthalmic decision support.

Keywords: Diabetic Retinopathy, DiabFoRMaxNet, OCTA, RankMax Classifier, Retinal Fundus.

Introduction

Diabetic Retinopathy (DR) is a progressive retinal disorder caused by diabetes mellitus and remains one of the leading causes of vision impairment and blindness among working-age adults worldwide. The disease damages retinal blood vessels, resulting in leakage, ischemia, hemorrhages and neovascularization, which gradually affect visual function. Early diagnosis and accurate grading of DR are essential to prevent irreversible blindness and to support timely clinical intervention. Traditionally, ophthalmologists diagnose DR through manual inspection of retinal fundus and Optical Coherence Tomography (OCT) images. However, manual examination is time-consuming, subjective and dependent on clinical expertise, which limits large-scale screening and may lead to inter-observer variability.

Recent advances in Artificial Intelligence (AI) and Deep Learning (DL) have significantly improved automated DR detection and classification. Convolutional Neural Networks (CNNs) have demonstrated superior capability in extracting discriminative retinal features from fundus and OCT images, thereby enabling accurate disease classification and lesion localization (1). Several deep learning architectures, including AlexNet, ResNet, Inception-based networks, Vision Transformers (ViTs) and hybrid CNN-RNN frameworks (2), have been investigated for DR diagnosis. In addition, Fully Convolutional Networks (FCNs), attention mechanisms and optimization-based feature selection techniques (3) have been introduced to improve segmentation accuracy and classification efficiency.

Despite these developments, several challenges remain unresolved in existing DR diagnostic systems. Many approaches rely on single-modality retinal imaging, limiting their ability to capture complementary pathological information. Several models also suffer from high computational

(Received 02nd December 2025; Accepted 02nd June 2026; Published 02nd July 2026)

complexity, overfitting, class imbalance and reduced generalization ability when tested on heterogeneous datasets (4). Furthermore, some hybrid architectures achieve high performance at the cost of increased training time and memory consumption, which restricts their applicability in real-time clinical environments. Conventional classification approaches based on Softmax functions also encounter difficulties in handling uncertain probability distributions and imbalanced class samples (5).

To address these limitations, this study proposes a novel multimodal deep learning framework named DiabFoRMaxNet for automated detection and classification of Diabetic Retinopathy using retinal fundus and OCT images. The proposed framework integrates advanced preprocessing, efficient feature extraction, lesion segmentation, optimization and adaptive classification techniques to improve diagnostic performance while reducing computational complexity. The architecture is organized into five major stages: image preprocessing, feature extraction, feature segmentation, feature optimization and classification.

Initially, preprocessing techniques such as Z-score normalization, green channel extraction, Adaptive Histogram Equalization and Bilateral Filtering are applied to enhance retinal image quality and suppress noise (6). Subsequently, AlexNet is utilized to extract hierarchical spatial features from retinal images. The extracted features are then processed using a Fully Convolutional Network (FCN) for accurate lesion segmentation. To improve feature quality and eliminate redundant information, the Fossa Optimization Algorithm (FOA) is employed as a feature optimization strategy. Finally, the RankMax classifier replaces the conventional Softmax classifier to better handle imbalanced data distributions through adaptive Euclidean projection-based classification (7, 8). This integrated framework enables accurate feature learning, efficient segmentation and robust classification suitable for practical clinical deployment.

Several recent studies have investigated deep learning techniques for DR detection and classification. Modified Inception-V3-based approaches achieved notable classification performance using retinal fundus images; however, their high computational complexity limited real-time applicability (9). Multi-layer CNN architectures demonstrated strong specificity but comparatively lower sensitivity in identifying early-stage DR, increasing the possibility of missed diagnoses (10). Inception-ResNet-V2 frameworks improved classification accuracy by combining residual learning and multi-scale feature extraction, although computational burden and class imbalance issues remained significant limitations (11).

Hybrid deep learning frameworks integrating Inception-ResNet models with Gated Recurrent Units (GRUs) and optimization algorithms such as Fennec Fox Optimization (FFO) achieved improved diagnostic accuracy and balanced evaluation metrics. Nevertheless, increased training complexity and potential overfitting restricted their scalability in resource-constrained clinical systems (12). Similarly, optimization-driven methods utilizing Generative Adversarial Net-works (GANs) and swarm intelligence techniques demonstrated enhanced classification accuracy but introduced substantial computational over-head and reduced interpretability (13).

In OCT-based DR analysis, advanced preprocessing frameworks employing contrast enhancement and vessel segmentation methods achieved excellent classification performance, although processing complexity and generalization capability remained concerns (14). Segmentation-oriented architectures based on KiU-Net and Cross-Attention mechanisms improved retinal lesion representation but exhibited relatively lower Dice Similarity Coefficient (DSC) values, indicating challenges in accurate boundary delineation (15). Furthermore, hybrid ResNet-Vision Transformer models and CNN-RNN frameworks demonstrated the capability to capture both local and global retinal features from OCT images, but their memory requirements and inference complexity limited real-time implementation (16, 17). Biomarker-driven OCT approaches based on retinal layer smoothness indices also demonstrated high diagnostic accuracy for DR staging; however, their dependence on high-quality OCT scans and manual structural interpretation reduced scalability (18).

Considering these limitations, the proposed DiabFoRMaxNet framework aims to provide a computationally efficient and highly accurate multimodal solution for automated DR detection

and classification (19). By integrating optimized preprocessing, deep feature extraction, FCN-based segmentation, FOA-driven feature optimization and RankMax classification, the proposed model improves classification robustness and diagnostic reliability for both retinal fundus and OCT imaging modalities (20).

The remainder of this manuscript is organized as follows. Section 2 presents the proposed DiabFoRMaxNet methodology and architecture. Section 3 describes the experimental results and performance evaluation analysis. Section 4 discusses the obtained findings and comparative assessment with existing approaches. Finally, Section 5 concludes the study and outlines future research directions

Methodology

The suggested methodology proposes to detect and classify DR via retinal fundus and OCTA images. The approach consists of a systematic pipeline with five essential stages shown in Figure 1. Image Preprocessing, Feature Extraction, Feature Segmentation, Feature Optimization and Classification. DiabFoRMaxNet, an innovative deep learning architecture, performs feature extraction, segmentation, optimization and classification tasks. Every phase is meticulously designed to optimize efficiency, reduce computing complexity and guarantee reliability.

Retinal fundus and OCTA images are obtained from a publicly accessible multimodal OCTA and fundus imaging dataset (21). The obtained images are exposed to preprocessing to improve their quality and facilitate feature extraction.

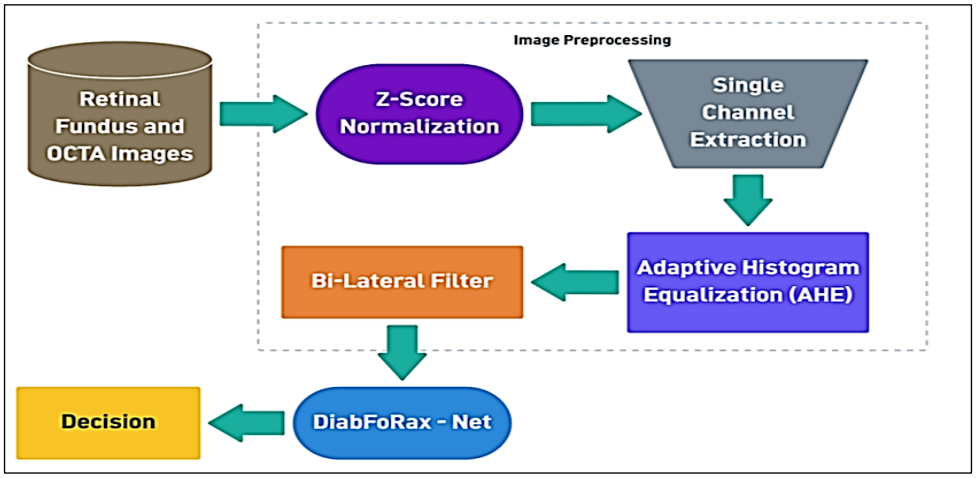

Figure 1: Proposed Block

The input image can be expressed mathematically as Equation [1]:

$$I(x,y) = \{R(x,y), G(x,y), B(x,y)\} \qquad [1]$$

In Equation [1], $I(x,y)$ is the Input Image and $R(x,y), G(x,y), B(x,y)$ are the red, green and blue intensities of input image respectively.

To standardize the pixel intensities, Z – Score Normalization is applied in Equation [2], Where X is original pixel intensity, μ is Mean intensity and σ is Standard Deviation.

$$X' = \frac{X-\mu}{\sigma} \qquad [2]$$

The Single-Channel Extraction method isolates the green channel in Equation [3]:

$$G(x,y) = I(x,y,2) \qquad [3]$$

$I(x,y,2)$is intensity at pixel (x,y) in channel c with $c = 2$ denoting the green channel.

Adaptive Histogram Equalization (AHE) is applied for contrast enhancement in Equation [4]:

$$H_{eq}(i) = \frac{H(i)-H_{min}}{H_{max}-H_{min}} \times (L-1) \qquad [4]$$

where H(i) is the histogram value at intensity i, H_{max}, H_{min} are the maximum and minimum histogram values and L is the number of intensity levels

To remove noise while preserving image features, Bilateral Filtering is applied in Equation [5]:

$$I'(x,y) = \frac{1}{W_p}\sum_{i,j} I(i,j) f_g(\|x-i, y-i\|) f_c(|I(x,y) - I(i,j)|) \quad [5]$$

where f_g is the geometric distance function, f_c is the color intensity similarity function and W_p is a normalization factor.

The Feature Extraction stage utilizes AlexNet, a deep convolutional neural network (CNN), to extract hierarchical features. Figure 2 illustrates the AlexNet-based feature extraction architecture integrated within the proposed DiabFoRMaxNet framework. The preprocessed retinal fundus and OCTA images are supplied to multiple convolutional layers with ReLU activation and max-pooling operations to extract hierarchical spatial and texture-related retinal features. The convolutional layers effectively learn discriminative lesion-associated patterns, while the max-pooling layers reduce spatial dimensionality and computational complexity. Finally, the fully connected layers generate compact deep feature representations that are subsequently utilized for FCN-based lesion segmentation and RankMax-based diabetic retinopathy classification.

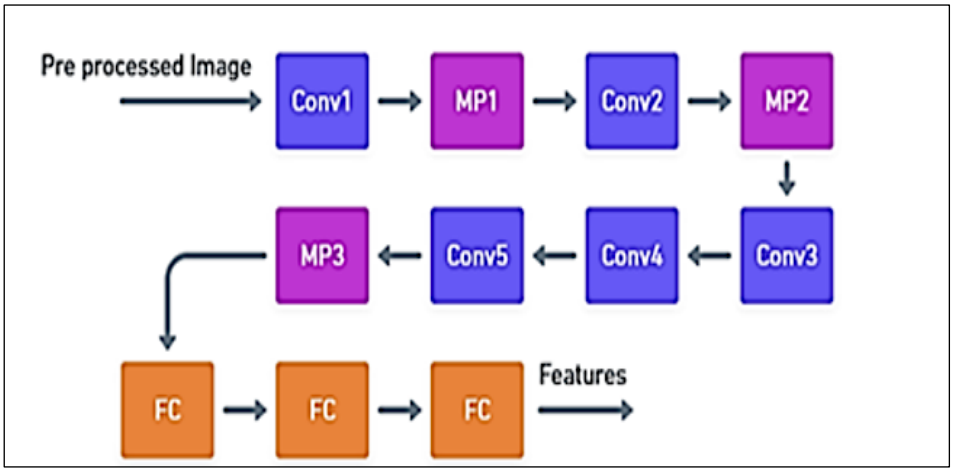

Figure 2: AlexNet Architecture for Feature Extraction from Pre-processed Retinal Images. The Architecture Consists of Convolutional Layers with ReLU Activation Functions, Max-pooling Layers and Fully Connected Layers. Conv denotes Convolution + ReLU, MP denotes Max Pooling and FC denotes Fully Connected Layers

The convolution operation is defined as in Equation [6]:

$$F_l(x,y) = \sum_{i=1}^{M}\sum_{j=1}^{N} W_l(i,j) X(x+i, y+j) + b_l \quad [6]$$

where W_l is the Convolutional kernel, X is the input image and b_l is the bias term.

ReLU activation introduces non-linearity in Equation [7]:

$$A(x) = \max(0, x) \quad [7]$$

A max-pooling function reduces spatial dimensions shown in Equation [8]:

$$P(x,y) = \max_{i=0}^{M} \max_{j=0}^{N} f_l(x+i, y+j) \quad [8]$$

FCN performs pixel-wise segmentation of retinal fundus images. Instead of fully connected layers, FCN applies convolutional layers shown in Equation [9]:

$$Y(x,y) = f(W * X + b) \quad [9]$$

where $*$ is the Convolution operation. To recover spatial resolution, deconvolution (transposed convolution expressed in Equation [10]) is applied:

$$D(x,y) = \sum_{i=0}^{M} \sum_{j=0}^{N} W^{T}(x,y)X(x+i,y+j) \quad [10]$$

Fossa Optimization Algorithm (FOA) optimizes extracted features by mimicking fossa hunting strategies. Fossa explores the search space using in Equation [11]:

$$X_i^{t+1} = X_i^t + \alpha R(X_{best}^t - X_i^t) \quad [11]$$

where α is the exploration rate, R is the random number and X_{best} is the best solution found so far.
Fine-tuning of feature selection is performed using in Equation [12]:

$$X_i^{t+1} = X_i^t + \beta(X_{best}^t - X_i^t) \quad [12]$$

where β controls the exploration intensity.
Rankmax modifies Softmax to prevent undefined probabilities in Equation [13]:

$$P_i = \frac{e^{z_i - \tau\|z_i\|}}{\sum_{j=1}^{N} e^{z_j - \tau\|z_j\|}} \quad [13]$$

where τ is the Lipschitz constant adapting to each training example.
The model is trained using modified cross-entropy loss In Equation [14]:

$$L = -\sum_{i=1}^{N} y_i \log P_i \quad [14]$$

where y_i is the ground truth labels.

Proposed Model

The selection of AlexNet, FCN, FOA and RankMax in the proposed DiabFoRMaxNet framework was based on the objective of achieving high diagnostic accuracy while balancing computational efficiency, feature precision and classification stability. AlexNet was selected for feature extraction due to its ability to hierarchically capture discriminative retinal characteristics with comparatively lower computational complexity than deeper architectures, making it suitable for multimodal medical imaging applications. FCN was incorporated for pixel-wise segmentation to preserve spatial coherence and improve lesion localization, which is essential for accurately identifying retinal abnormalities that may be overlooked by classification-only frameworks.

FOA was employed as a feature optimization strategy to eliminate redundant and non-informative features while balancing exploration and exploitation, thereby improving feature compactness and reducing overfitting risk. This optimization stage enhances computational efficiency and strengthens the quality of features forwarded for classification. RankMax was selected as the final classifier to overcome limitations of conventional Softmax under class-imbalanced conditions by providing adaptive ranking-based decision stability, thereby improving classification robustness across multiple DR severity classes.

The integration of these components provides a distinct advantage over conventional deep learning approaches by combining efficient hierarchical feature extraction, lesion-aware segmentation, redundancy-controlled optimization and imbalance-resilient classification within a unified multimodal architecture. Different standalone CNN or transfer learning models, DiabFoRMaxNet simultaneously improves lesion localization, feature discrimination, computational economy and classification reliability, which collectively contributes to its superior performance across retinal fundus and OCTA datasets.

DiabFoRMaxNet is a deep learning framework developed for the identification and classification of Diabetic Retinopathy (DR) utilizing retinal fundus and OCT images shown in Figure 3. The design comprises four primary stages: Feature Extraction, Feature Segmentation, Feature Optimization and Classification (22). During the feature extraction phase, AlexNet, a convolutional neural network (CNN), is utilized to derive hierarchical spatial characteristics from the input photos. AlexNet's several convolutional layers, activation functions and pooling operations enable it to acquire resilient representations of diabetic

retinopathy-related anomalies. Subsequent to feature extraction, the Fully Convolutional Network (FCN) is employed for pixel-wise segmentation, guaranteeing accurate detection of DR-affected areas in retinal pictures. In contrast to patch-based segmentation techniques, FCN functions on the complete image, maintaining spatial coherence and improving lesion localization. Upon completion of segmentation, feature selection and refinement are executed with the Fossa Optimization Algorithm (FOA), which enhances the extracted features by discarding redundant and irrelevant information. The improved characteristics are ultimately input into the RankMax classifier, a substitute for Softmax, to categorize the severity of diabetic retinopathy, thereby enhancing flexibility to class imbalances and augmenting decision-making precision.

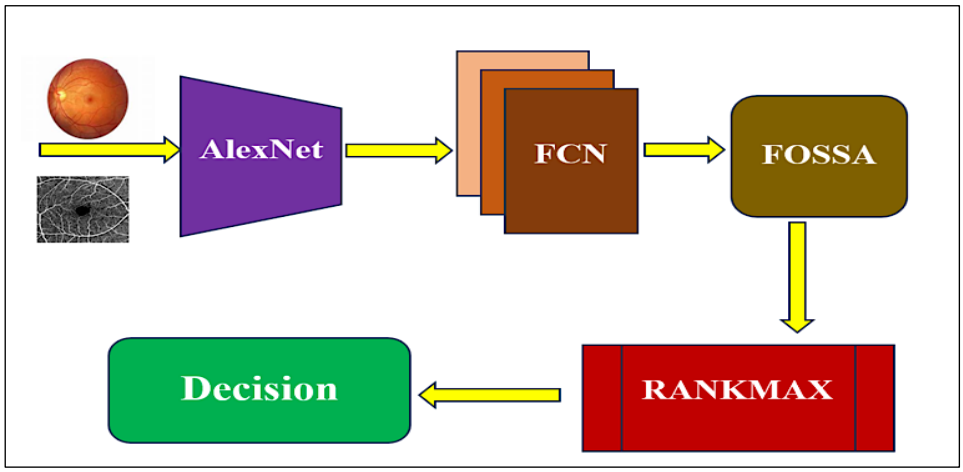

Figure 3: Proposed DiabFoRMaxNet Architecture Illustrating the Sequential Integration of Alexnet for Hierarchical Feature Extraction, FCN For Pixel-wise Lesion Segmentation, FOA For Feature Optimization and Redundancy Reduction and Rankmax for Adaptive Diabetic Retinopathy Severity Classification

DiabFoRMaxNet's innovation is in its unique amalgamation of deep learning, nature-inspired optimization and adaptive classification methods, resulting in a highly efficient and resilient diabetic retinopathy detection system (23). In contrast to traditional models that depend exclusively on feature extraction and classification, DiabFoRMaxNet incorporates FCN-based segmentation to improve lesion localization, facilitating a more comprehensive evaluation of retinal anomalies. The integration of the Fossa Optimization Algorithm (FOA) for feature selection differentiates this model from conventional methods (24). FOA, influenced by the strategic hunting behavior of fossa, optimizes the retrieved characteristics by adeptly balancing exploration and exploitation, hence diminishing computational complexity while enhancing accuracy. A significant innovation is the RankMax classifier, which substitutes the traditional Softmax function with an adaptive Euclidean projection technique. This change resolves the problem of undefined probabilities in entropy-based formulations, enhancing the stability and efficacy of the classification process, especially for imbalanced datasets (25). Through the integration of these sophisticated components, DiabFoRMaxNet attains enhanced performance in diabetic retinopathy detection while preserving computational economy, rendering it a potential instrument for clinical applications.

Data Augmentation

The initial dataset comprises 222 images, which is inadequate for efficiently training a deep learning model. To increase the dataset to 13,424 photos, distinct augmentation approaches are employed for OCTA and fundus images, ensuring the preservation of the original class distribution.

Augmentation approaches for OCTA images prioritize the preservation of vascular structures. Geometric transformations, including rotation (±10° to ±20°), scaling and elastic deformations, are employed to create diversity while preserving vessel patterns. Intensity-based augmentations such as contrast stretching, histogram equalization and adaptive gamma correction improve vascular visibility. Noise-based augmentations, such as Gaussian noise injection and motion blur, replicate real-world variances.

Augmentation techniques for fundus pictures guarantee the retention of retinal landmarks. Geometric changes such as rotation (±15° to ±30°), flipping and cropping enhance the diversity of image orientations. Photometric augmentations, such as brightness, contrast and hue-saturation modifications, replicate various lighting conditions. Gaussian blur, speckle noise and CLAHE-based histogram adjustments are utilized to replicate authentic fluctuations in fundus imaging. Approximately 60 augmented copies of each image are produced, resulting in a final dataset of 13,424 images, so assuring balanced representation across three categories: NO DR (10,880 photos), MILD DR (1,560 images) and MODERATE DR (984 images). The dataset is divided into a 70-20-10 split, designating 9,397 photos for training, 2,685 for validation and 1,342 for testing. This augmentation method guarantees diversity, mitigates overfitting and improves the model's capacity to generalize across several imaging modalities.

Dataset Splitting

Dataset augmentation and partitioning were performed independently for OCTA and Fundus images to preserve modality-specific characteristics and maintain balanced training, validation and testing distributions. OCTA images were augmented using rotation, scaling, elastic deformation, intensity enhancement, Gaussian noise injection and motion blur, while Fundus images underwent geometric, photometric and noise-based augmentations. After augmentation, each dataset contained 6,712 images, divided into 4,698 training, 1,342 validation and 672 testing samples following a 70:20:10 ratio, as presented in Table 1 and Figures 4A and 4B.

Table 1: Dataset Distribution

Classification	Total OCTA Images	Training (70%)	Validation (20%)	Testing (10%)	Total Fundus Images	Training (70%)	Validation (20%)	Testing (10%)
NO DR	5,440	3,808	1,088	544	5,440	3,808	1,088	544
MILD DR	780	546	156	78	780	546	156	78
MODERATE DR	492	344	98	50	492	344	98	50
Total	6,712	4,698	1,342	672	6,712	4,698	1,342	672

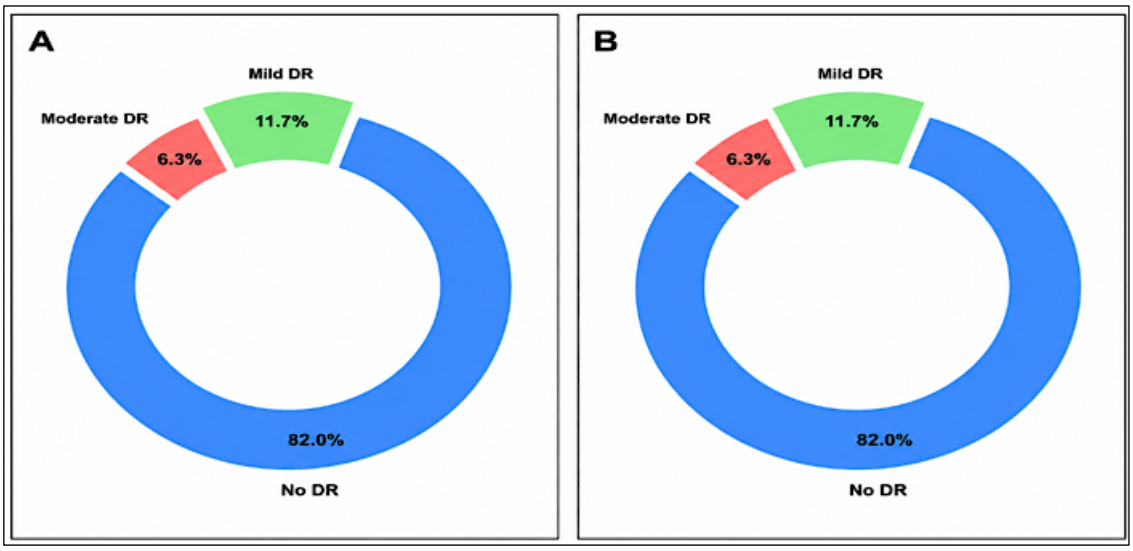

Figure 4: Dataset Distribution. (A) Retinal Fundus Images and (B) OCTA Images

Clarification of Dataset Expansion and Data Diversity Preservation

To ensure methodological transparency and avoid overestimation of model performance, the expansion from 222 original multimodal images to 13,424 images was achieved exclusively through controlled data augmentation rather than through the acquisition of additional independent samples. The original dataset comprised unique retinal fundus and OCTA images, while augmentation techniques—including geometric transformations, photometric adjustments and noise-based variations—were applied to enhance training diversity and reduce overfitting while preserving

clinically relevant retinal and vascular characteristics. This augmentation strategy improved computational robustness but did not introduce new biological diversity beyond the original dataset.

To eliminate the risk of data leakage and inflated evaluation outcomes, dataset partitioning into training, validation and testing subsets was performed prior to augmentation at the source-image level. Consequently, all augmented versions derived from a specific original image were confined exclusively to the same subset and were not distributed across multiple subsets. This strict separation ensured that validation and testing performance reflected genuine generalization to unseen image groups rather than memorization of augmented duplicates. Therefore, the reported performance metrics of DiabFoRMaxNet represent controlled methodological rigor under limited-data conditions. However, although augmentation substantially improves model stability and learning efficiency, future validation using larger independent external datasets remains necessary to further confirm real-world scalability and clinical generalizability.

Validation Strategy and Test Set Independence

To ensure methodological rigor and address potential concerns regarding overfitting or evaluation bias, a strict validation strategy was implemented throughout the experimental design. The original multimodal dataset was partitioned into training (70%), validation (20%) and testing (10%) subsets at the source-image level prior to any augmentation procedures. This source-level partitioning ensured that all augmented derivatives generated from a particular original image were confined exclusively to the same subset and were never distributed across other subsets. Consequently, complete independence between training, validation and testing data was preserved, eliminating data leakage and preventing artificially inflated performance outcomes.

To further evaluate model robustness and generalization stability, 5-fold cross-validation was additionally performed across the training framework. This approach enabled repeated assessment of model consistency across multiple data partitions while minimizing fold-specific bias. Training and validation accuracy/loss trends were carefully monitored to detect convergence behavior and potential overfitting. The consistently narrow variation across folds, along with stable validation performance, indicates that the near-perfect classification metrics were not solely a consequence of memorization but rather reflect the effectiveness of the multimodal architecture under controlled experimental conditions. Nevertheless, external validation on larger independent multicenter datasets remains an essential future step to further confirm large-scale clinical generalizability.

Experimental Setup

The experimental setup for this research on diabetic retinopathy detection and classification was devised to guarantee efficient computing and superior processing performance. The system utilized an Intel Core i7-1355U processor achieving speeds of up to 5.0 GHz with Intel Turbo Boost Technology 2.0, comprising a 12 MB L3 cache, 10 cores and 12 threads. The system was outfitted with 16 GB LPDDR4x-4266 MHz RAM and 1 TB PCIe NVMe TLC M.2 SSD storage to manage extensive datasets and intricate deep learning models, facilitating rapid data access and efficient multitasking. The study was performed using Windows 11 Home 64 Plus Single Language, featuring preloaded applications such as McAfee LiveSafe (12-month complimentary trial), Adobe Creative Cloud (20 applications for 1 month) and Adobe Express Premium (2 months of free access). The display configuration included a 34.3 cm (13.5") diagonal 3K2K (3000 × 2000) OLED panel, providing multitouch capability, ultra-wide viewing angles (UWVA), edge-to-edge micro-edge glass, anti-reflective Corning Gorilla Glass NBT and low blue light technology, guaranteeing visual precision with a 100% DCI-P3 color gamut and VESA True Black HDR 400 certification. The Intel Iris Xe Graphics of the system enabled seamless viewing of medical images and outputs from deep learning models. MATLAB, in conjunction with the picture Processing Toolbox, Deep Learning Toolbox and Computer Vision Toolbox, was utilized for image preprocessing, feature extraction, segmentation, optimization and classification, employing GPU acceleration when appropriate to improve computational performance. This experimental configuration achieved an optimal equilibrium among performance, speed and accuracy for the effective implementation of the research workflow.

Experimental Results

The experimental findings unequivocally show that the suggested DiabFoRMaxNet architecture outperforms popular conventional deep learning models like DenseNet, ResNet, VGG and AlexNet shown in Table 2. Across all evaluation measures, including sensitivity (99.37%), specificity (99.40%), precision (99.26%) and F1-score (99.11%), DiabFoRMaxNet delivers an exceptional accuracy of 99.18% for retinal fundus images, much surpassing ResNet-50 (97.42%), DenseNet-121 (97.88%) and others. In a similar vein, DiabFoRMaxNet outperforms conventional models for OCTA images, with a maximum recorded accuracy of 99.69%, whereas DenseNet-201 only manages 97.55%. The suggested model's superior performance is a result of its segmentation accuracy, feature refining by FOA and final classification through RankMax.

Despite being deep and well-structured, conventional architectures lack the specialized pipeline integration provided by DiabFoRMaxNet, which not only processes multimodal inputs more efficiently but also guarantees minimal loss of pathological features during segmentation and optimization, as this comparative evaluation makes clear. DiabFoRMaxNet therefore becomes a high-performance, robust and modality-aware framework for trustworthy DR detection and grading.

Results

The progression of images from Figures 5 (A-J) demonstrates the transformation of raw retinal fundus and OCTA images respectively, through the image preprocessing stages of the proposed DiabFoRMaxNet methodology, which is crucial for improving diagnostic accuracy in the detection and classification of Diabetic Retinopathy (DR) (26). For OCTA preprocessing starts with Z-score normalization to make the intensity distribution more stable and then the green channel is extracted to make fine vascular networks easier to see. Adaptive Histogram Equalization (AHE) makes areas that are likely to have ischemia or capillary dropout look better by increasing the contrast.

Bilateral filtering, on the other hand, gets rid of noise and keeps the edge features that are important for finding lesions (27). The fundus image also goes through the same systematic preparation stages. The green channel extraction makes the contrast between blood vessels more noticeable, the AHE makes small lesions easier to see and the bilateral filtering keeps the continuity of the vessels while cutting down on background noise (28). These changes that are specific to each modality get the images ready by bringing out early signs of disease such microaneurysms, haemorrhages, or capillary dropout.

These systematic improvements in both imaging modalities make sure that DR-related vascular abnormalities are kept, improved and made more uniform (29). This sets up a strong base for the next steps: feature extraction with AlexNet, pixel-wise segmentation with FCN, feature optimization with FOA and DR classification with RankMax. The dual-modality preprocessing approach significantly enhances the model's capacity to extract significant features from both structural and microvascular data, facilitating highly accurate early-stage DR diagnosis with minimal computing burden.

The sequence of fundus images from in Figures 5(A–E) shows how the Image Preprocessing stage in the DiabFoRMaxNet pipeline is used to find and classify Diabetic Retinopathy (DR). The first image in Figure 5A is a high-resolution scan of the retina that shows important anatomical characteristics such the optic disc, macula and vascular tree (30). This raw input gives the required visual information, but changes in lighting, contrast and noise levels can make it impossible to see the first signs of DR, including microaneurysms, hemorrhages, or hard exudates. To fix these problems, we first use Z-score normalization, which gives us in Figure 5B. This step may look excessive, but it really stabilizes the distributions of pixel intensities statistically, which makes the features more uniform across the dataset for deep learning models. After this, the green channel is taken out in Figure 5C, because it shows blood vessels and small lesions better than the red or blue channels. This channel makes blood vessels easier to see, which is especially helpful for spotting DR-related changes like venous beading or neovascularization.

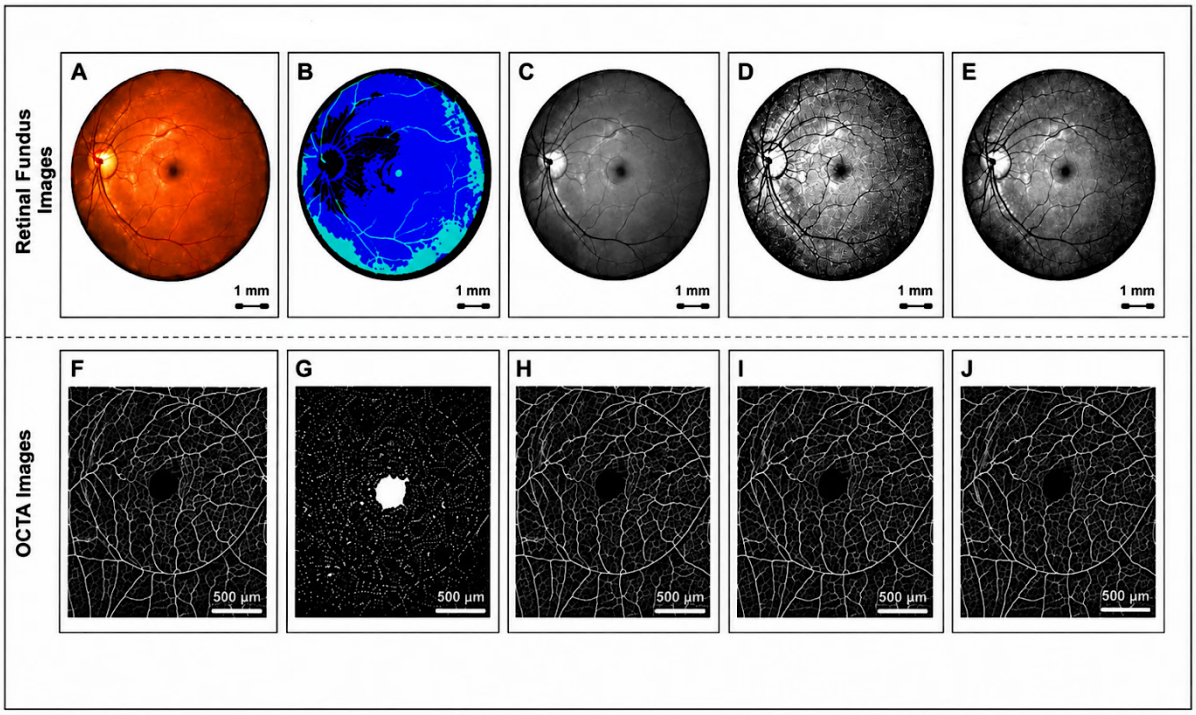

Figure 5: Composite representation of Pre-processing Stages for Retinal Fundus and OCTA Images within the DiabFoRMaxNet Framework. (A) Original Retinal Fundus Image, (B) Z-score Normalized Retinal Fundus Image, (C) Green Channel Extracted Retinal Fundus Image, (D) Adaptive Histogram Equalized Retinal Fundus Image, (E) Bilateral Filtered Retinal Fundus Image, (F) Original OCTA Image, (G) Z-Score Normalized OCTA Image, (H) Green Channel Extracted OCTA Image, (I) Adaptive Histogram Equalized OCTA Image, (J) Bilateral Filtered OCTA Image

Then, as shown in Figure 5D, Adaptive Histogram Equalization (AHE) is done. This approach for enhancing contrast changes the intensity levels in small places, making it easier to see faint vascular structures and highlighting possible lesion locations surrounding the macula and periphery. This phase is important for highlighting the early signs of DR (31). Lastly, the image in Figure 5E was made using bilateral filtering. This method gets rid of background noise while keeping edge information, which is important for keeping delicate features like microaneurysms and thin arteries intact for later examination.

The series of images from in Figure 5 (F–J) shows how the input OCTA image changes as it goes through the Image Preprocessing stage of the proposed DiabFoRMaxNet technique for finding and classifying diabetic retinopathy (DR). Figure 5F shows the original OCTA image, which shows the detailed retinal microvasculature, including the central foveal avascular zone (32). This raw image has a lot of anatomical detail, but it might include uneven lighting and noise that could hide early symptoms of diabetic retinopathy (DR), like microaneurysms, ischemia zones, or capillary rarefaction.

As illustrated in Figure 5G, Z-score normalization is used to standardize intensity levels and make learning easier. Even if this change doesn't make clinical details look better, it makes sure that the pixel distribution is normalized with a mean of zero and a variance of one. This gets the image ready for the best processing in neural network layers (33, 34). After that, the green channel is taken out, as shown in Figure 5H. The green channel offers better contrast of blood arteries than the red and blue channels in fundus and OCTA imaging. This extraction helps bring out small changes in blood vessels caused by DR. Figure 5I uses Adaptive Histogram Equalization (AHE) to make the local contrast better. This phase makes vascular structures bigger and makes darker areas that are related to ischemia or dropout stand out more. This makes it easier for later deep learning models to find features that show mild or moderate DR (35, 36). Lastly, Figure 5J shows that Bilateral Filtering is used to get rid of noise while keeping critical edges. The final image keeps the vessels intact and smooths out the backdrop, making sure

that the Feature Extraction stage using AlexNet gets clean input.

Performance Evaluation Metrics: To assess the usefulness and dependability of the proposed DiabFoRMaxNet model in detecting and categorizing Diabetic Retinopathy (DR), a comprehensive array of performance metrics was employed, including Accuracy, Sensitivity (Recall), Specificity, Precision and F1-score. These metrics provide a comprehensive framework for quantifying classification results.

Accuracy denotes the ratio of accurately anticipated cases to the total predictions and is formally expressed as,

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad [15]$$

In this context, i.e., in Equations [15-18], TP denotes true positives, TN signifies true negatives, FP represents false positives and FN indicates false negatives. Sensitivity (Recall), indicative of the model's proficiency in accurately identifying DR instances, is essential for minimizing missed diagnoses and is defined as,

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \quad [16]$$

Specificity assesses the model's efficacy in recognizing healthy instances by reducing false positives and is described as,

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \quad [17]$$

Precision quantifies the ratio of true positive predictions to the total number of positive predictions and is computed as,

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad [18]$$

Table 2: Comparative Performance of DiabFoRMaxNet vs Traditional Deep Learning Models

Model	Imaging Modality	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-Score (%)
DiabFoRMaxNet	Retinal Fundus	**99.18**	**99.37**	**99.40**	**99.26**	**99.11**
DiabFoRMaxNet	OCTA	**99.69**	**99.83**	**99.85**	**99.58**	**99.49**
AlexNet (Baseline)	Retinal Fundus	96.85	96.10	95.80	95.42	95.76
ResNet-50	Retinal Fundus	97.42	96.75	96.88	96.55	96.64
VGG-16	Retinal Fundus	96.24	95.30	95.72	95.05	95.17
DenseNet-121	Retinal Fundus	97.88	97.40	96.90	96.85	97.12
ResNet-50	OCTA	97.21	97.65	96.72	97.10	97.37
VGG-19	OCTA	96.92	96.20	96.30	96.45	96.32
DenseNet-201	OCTA	97.55	97.88	97.20	97.10	97.49

Table 2 and Figure 6 show the comparative performance analysis of the proposed DiabFoRMaxNet model with existing deep learning models using Retinal Fundus and OCTA images. The proposed model achieved the highest performance, attaining 99.18% accuracy for Retinal Fundus images and 99.69% accuracy for OCTA images, along with superior sensitivity, specificity, precision and F1-score values. In comparison, baseline models such as AlexNet, ResNet, VGG and DenseNet demonstrated comparatively lower classification performance across both imaging modalities. The results confirm that the integration of optimized feature extraction, segmentation and RankMax classification significantly enhances diabetic retinopathy detection accuracy and robustness.

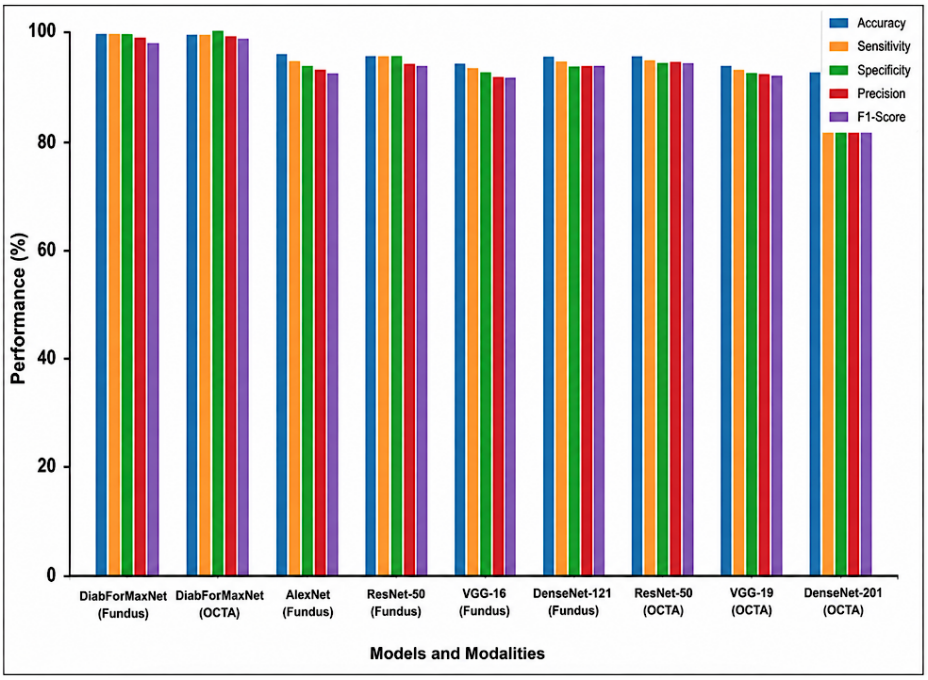

Figure 6: Performance Comparison with Baseline Models

The F1-score is the harmonic mean of precision and sensitivity, providing a fair evaluation, particularly in scenarios of class imbalance. It is calculated by Equation [19]:

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \qquad [19]$$

Evolution Metric Analysis

The suggested DiabFoRMaxNet model demonstrated outstanding performance in both imaging modalities shown in Figure 7. In retinal fundus pictures, it attained an accuracy of 99.18%, sensitivity of 99.37%, specificity of 99.40%, precision of 99.26% and an F1-score of 99.11%. The OCTA images yielded remarkable findings, with an accuracy of 99.69%, sensitivity of 99.83%, specificity of 99.85%, precision of 99.58% and an F1-score of 99.49%. These results highlight the model's exceptional accuracy in detecting DR patients while significantly reducing false positives. The better performance on OCTA images may be ascribed to their superior structural resolution, facilitating the model's ability to detect subtle disease alterations. These evaluation parameters collectively confirm that DiabFoRMaxNet surpasses traditional DR detection methods while ensuring clinical reliability, rendering it appropriate for practical diagnostic use. When applied to retinal fundus and OCTA images, respectively, the confusion matrices shown in Figures 8 and 9 demonstrate the classification reliability of the suggested DiabFoRMaxNet model. The model's performance on a test dataset of 672 images evenly split across three DR severity classes—No DR, Mild DR and Moderate DR—is represented by each matrix. With 540 out of 544 No DR photos, 75 out of 78 Mild DR images and 48 out of 50 Moderate DR images properly identified, the confusion matrix for retinal fundus images in Figure 8 shows extremely accurate predictions. There were very few cases of mild and moderate DR that were incorrectly identified; they were either reduced to No DR or confused with nearby stages. Particularly in retinal fundus analysis, where early lesions may appear modest and class boundaries are visually confusing, this slight overlap is to be expected. The confusion matrix for OCTA images in Figure 9 illustrates even greater classification accuracy. Only three misclassifications occurred throughout the whole test set, with the model correctly classifying 542 No DR, 77 Mild DR and 49 Moderate DR images. This better result demonstrates the diagnostic benefit of OCTA imaging, which helps the model differentiate between different levels of DR severity by providing greater visualization of ischemia zones and sharper microvascular detail.

DiabFoRMaxNet generalizes well across both modalities, as evidenced by the nearly complete diagonal dominance in both matrices. However, OCTA provides marginally better discrimination because of its structural depth and greater sensitivity to early microvascular alterations. These findings support the model's resilience and applicability for actual DR screening applications, particularly in clinical contexts where precise grading and early diagnosis are essential.

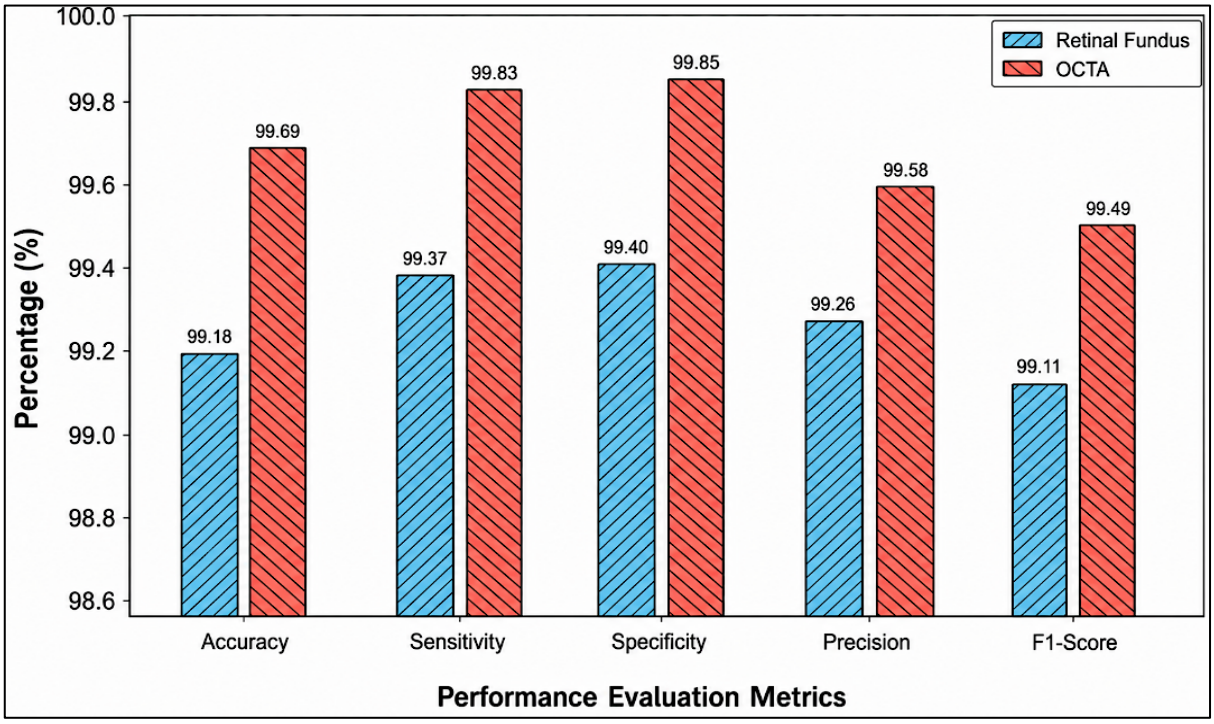

Figure 7: Comparative Performance Metrics of the Proposed DiabFoRMaxNet Model Across Retinal Fundus and OCTA Modalities. Blue Diagonally Patterned Bars Represent Retinal Fundus Image Performance, While Orange Diagonally Patterned Bars Represent OCTA Image Performance Across Accuracy, Sensitivity, Specificity, Precision and F1-Score Evaluation Metrics

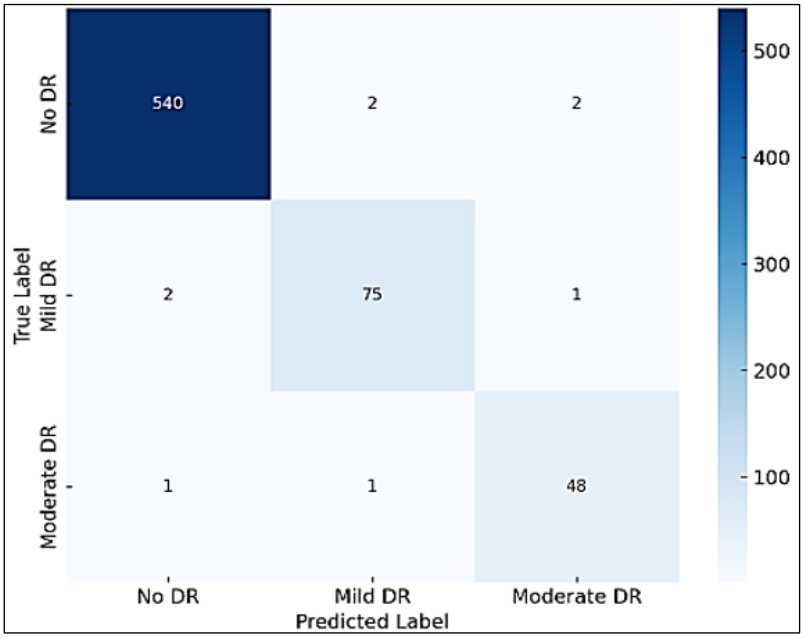

Figure 8: Confusion Matrix of the Proposed DiabFoRMaxNet Model for Retinal Fundus Diabetic Retinopathy Classification Across NO DR, Mild DR and Moderate DR Categories. The Blue Color Gradient Scale Represents Classification Frequency, Where Darker Shades Indicate Higher Sample Concentration and Lighter Shades Indicate Lower Frequency of Predictions

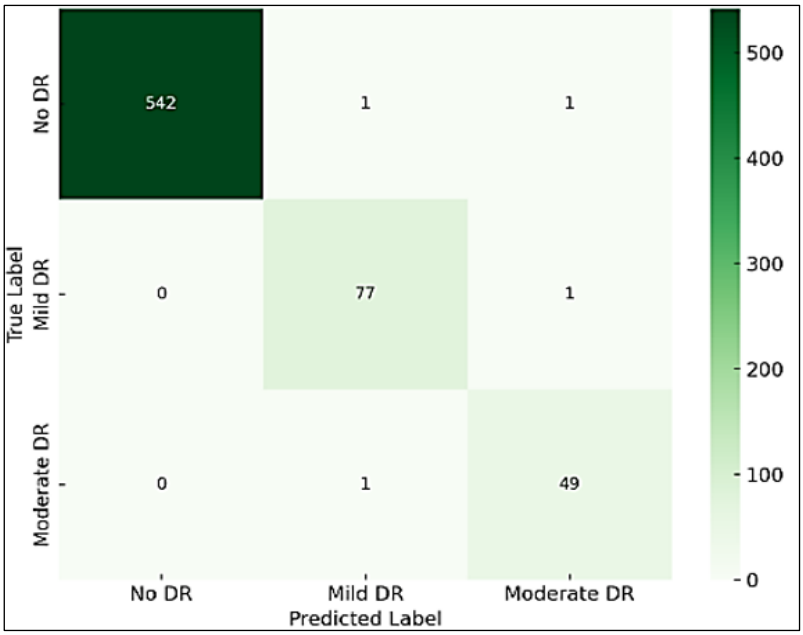

Figure 9: Confusion Matrix of the Proposed DiabFoRMaxNet Model for OCTA Images Diabetic Retinopathy Classification across NO DR, Mild DR and Moderate DR Categories. The Blue Color Gradient Scale Represents Classification Frequency, Where Darker Shades Indicate Higher Sample Concentration and Lighter Shades Indicate Lower Frequency of Predictions

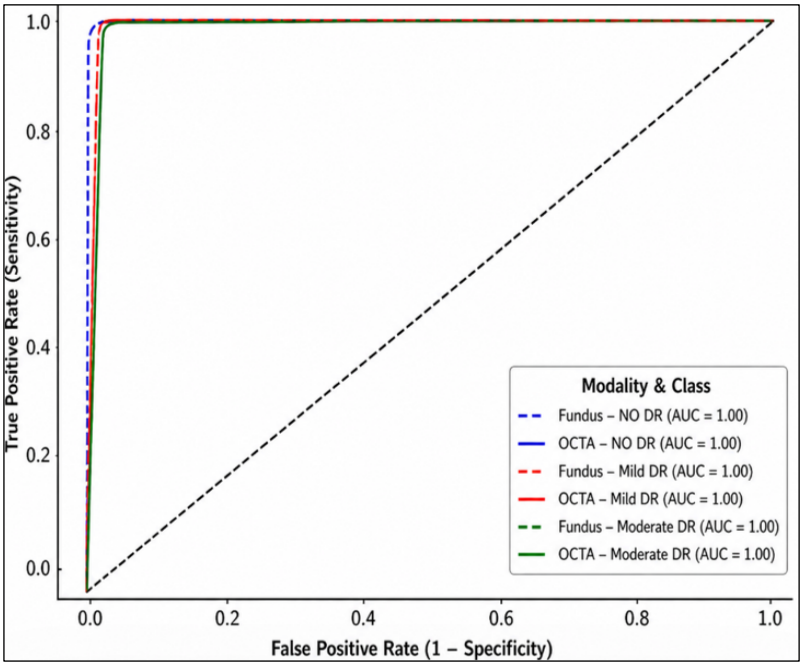

Figure 10: Receiver Operating Characteristic (ROC) Curves of the Proposed Diabformaxnet Model for Retinal Fundus and OCTA Modalities across Diabetic Retinopathy Severity Classes, Illustrating Classification Performance through True Positive Rate Versus False Positive Rate

The proposed DiabFoRMaxNet model performs exceptionally well in categorizing diabetic retinopathy stages—NO DR, MILD DR and MODERATE DR—using both Retinal Fundus and OCTA imaging modalities, as seen by the ROC curve shown in Figure 10. Every curve represents a distinct class within a modality and they all show an almost perfect trajectory that closely resembles the plot's top-left border, signifying a high true positive rate and a low false positive rate. Perfect classification capability is shown by an Area Under the Curve (AUC) of 1.00 for each class across both modalities. A random classifier with an AUC of 0.5 is represented by the diagonal reference line,

emphasizing the suggested model's greater discriminative potential. These findings demonstrate that DiabFoRMaxNet is highly appropriate for clinical diagnosis and screening.

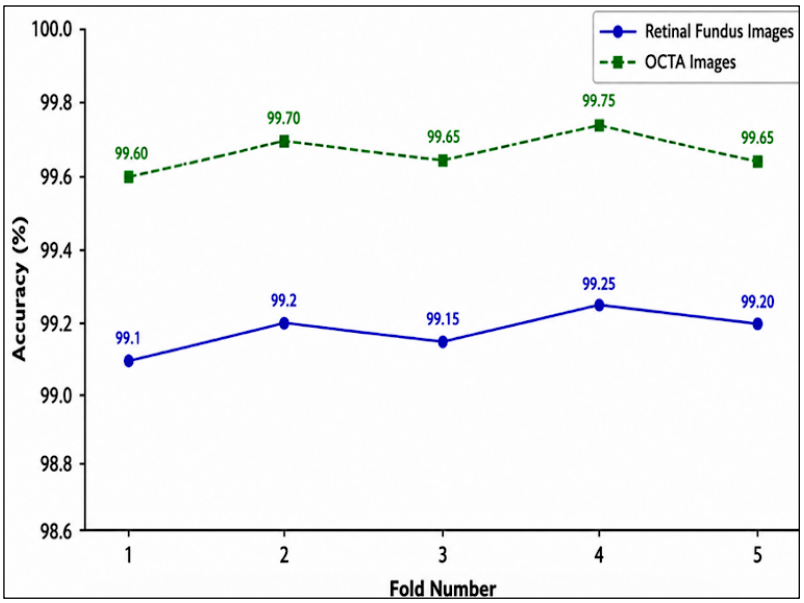

Figure 11: 5–fold Cross Validation Curve of a Proposed Model

The proposed DiabFoRMaxNet model's 5-fold cross-validation accuracy curves are shown in Figure 11 for the Retinal Fundus and OCTA image datasets, respectively. The model exhibits strong generalization and continuously maintains excellent accuracy over all five folds. The accuracy values for OCTA images range from 99.60% to 99.75%, slightly outperforming the fundus modality across all folds, while the accuracy values for Retinal Fundus images vary narrowly between 99.10% and 99.25%. The stability and dependability of the DiabFoRMaxNet model for the classification of diabetic retinopathy across both imaging modalities are confirmed by its consistent performance over several folds.

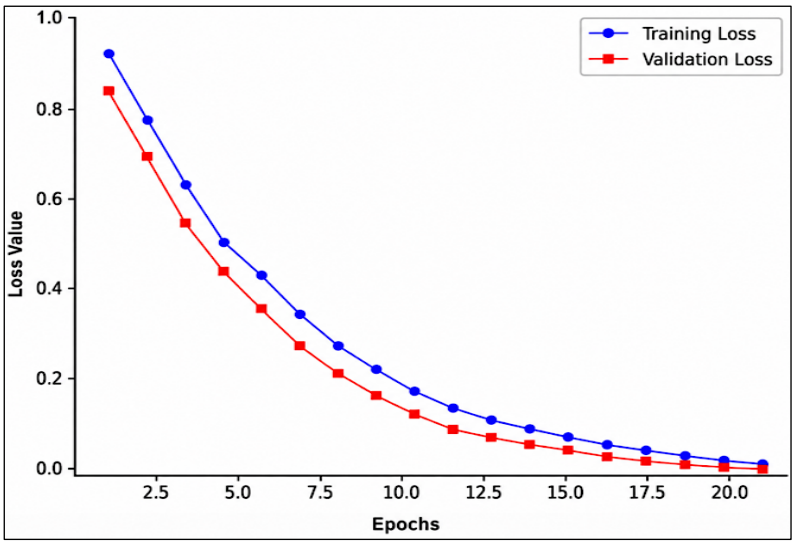

Figure 12: Loss Curve for the DiabFoRMaxNet Model

The training and validation loss trends over 20 epochs during 5-fold cross-validation for the DiabFoRMaxNet model are depicted by the plotted loss curve shown in Figure 12. The model learns efficiently in the early phases, as evidenced by the training loss (blue line), which drops quickly in the first epochs before eventually plateauing. Good generalization performance is suggested by the validation loss (red line), which likewise declines but at a somewhat slower pace. The model appears to maintain a balance between learning and

preventing overfitting if there are no significant fluctuations or divergence between the two curves. The distribution of prediction errors for the DiabFoRMaxNet model is shown in Figure 13. Since the majority of mistakes are cantered close to zero, the predictions and actual values are typically rather similar. The bars' tight spread indicates great accuracy and little variance, demonstrating the model's potent prediction power.

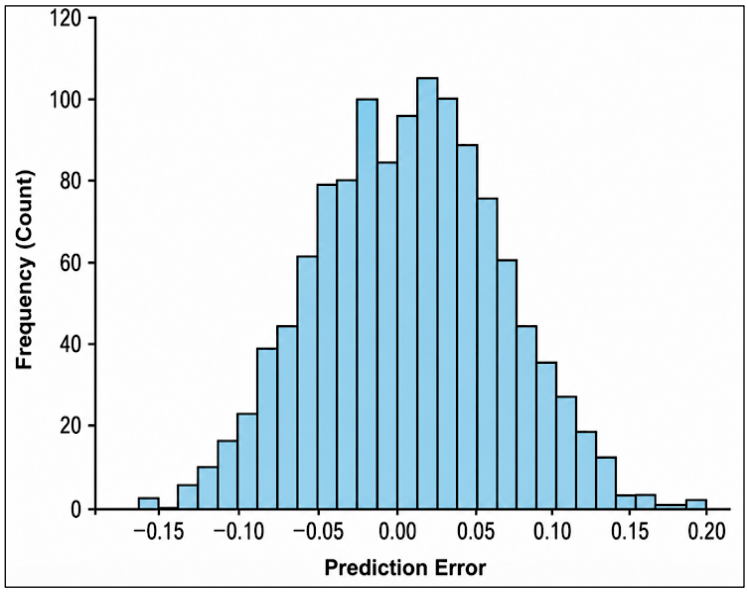

Figure 13: Error Histogram

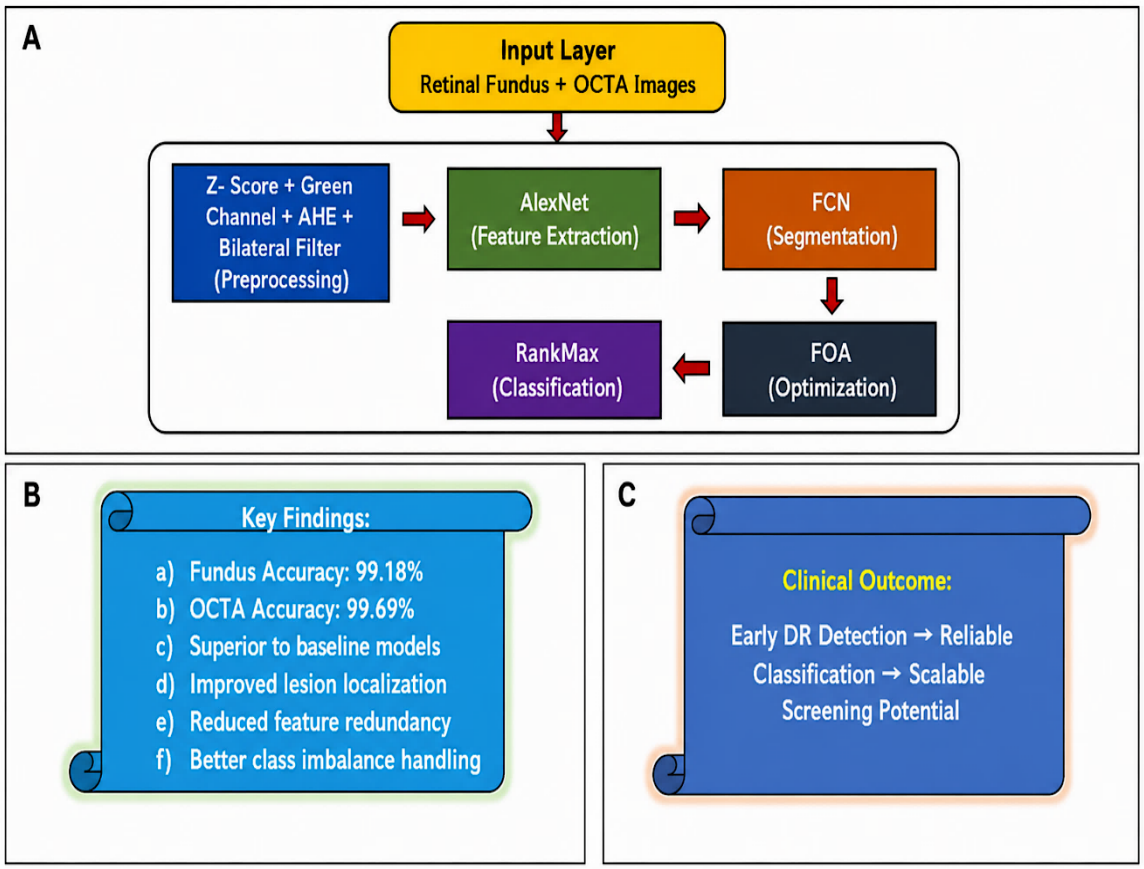

Figure 14: Schematic Summary of the Proposed DiabFoRMaxNet Framework. (A) Multimodal retinal image inputs and methodological pipeline, (B) Key findings, (C) Clinical Outcomes

Figure 14 summarizes the complete DiabFoRMaxNet framework from multimodal retinal image acquisition to optimized diabetic retinopathy classification, highlighting methodological innovations, superior diagnostic performance and potential clinical applicability.

Discussion

The experimental results confirm that the proposed DiabFoRMaxNet framework substantially outperforms several recent diabetic retinopathy detection approaches reported in the literature (11, 14, 16). For retinal fundus analysis, the proposed model achieved 99.18% accuracy, significantly exceeding previously reported models with lower classification performance. Similarly, in OCTA-based analysis, the achieved accuracy of 99.69% demonstrates superior capability compared to existing OCT/OCTA frameworks (9, 16, 19). These improvements are primarily attributed to the integrated multimodal design, FCN-based lesion segmentation, FOA-driven feature refinement and RankMax classification strategy, which collectively enhance feature discrimination, reduce redundancy and improve class imbalance handling (23, 24). Comparative evaluation confirms that the proposed framework offers improved diagnostic precision, computational efficiency and stronger clinical applicability for early DR screening.

Beyond numerical performance outcomes, DiabFoRMaxNet demonstrates important methodological and clinical significance for diabetic retinopathy diagnosis. The multimodal integration of retinal fundus and OCTA imaging enables improved utilization of both structural and microvascular retinal biomarkers, thereby strengthening disease characterization and early-stage detection capability (6, 7, 8). The combined architecture of preprocessing, feature extraction, segmentation, optimization and adaptive classification provides a systematic framework that enhances lesion localization, reduces redundant feature representation and improves classification stability (3, 15, 23).

From a clinical perspective, the strong diagnostic performance suggests potential applicability in early DR screening, reliable severity assessment and scalable AI-assisted ophthalmic evaluation. This framework may be particularly beneficial in settings requiring efficient large-scale screening support (24, 29). However, further external validation remains necessary to confirm broader real-world clinical generalizability.

Conclusion

DiabFoRMaxNet, a unique hybrid deep learning architecture designed for multimodal DR detection and classification, is presented in this study. The methodology tackles important issues including lesion localization accuracy, feature redundancy and imbalanced dataset classification by carefully combining FCN-based lesion segmentation, FOA-driven feature refinement and an adaptive RankMax classification strategy. By merging fundus and OCTA pictures, the dual-modality technique guarantees the extraction of structural and microvascular retinal biomarkers, significantly enhancing diagnostic reliability. The advantage of DiabFoRMaxNet in terms of accuracy, sensitivity, specificity, precision and F1-score across both modalities is confirmed by a comparative performance analysis against traditional deep learning models. Its robustness and generalization ability are demonstrated by the stability of 5-fold cross-validation and small prediction errors, while its discriminative capability is further highlighted by the perfect ROC-AUC values. Crucially, the suggested paradigm is feasible for real-world screening systems since it preserves computational efficiency while achieving state-of-the-art performance. DiabFoRMaxNet is positioned as a useful tool for ophthalmologists because to its good alignment with clinical requirements, especially in locations with limited resources where early DR identification is crucial.

Limitations of Study

Despite the strong performance of DiabFoRMaxNet, this study has certain limitations. The original dataset size was limited and although data augmentation improved model robustness, it does not replace true biological diversity from larger independent datasets. The model was evaluated on a single dataset without external multicenter validation, which may affect broader clinical generalizability. Additionally, the framework currently focuses on only three DR severity classes and does not include advanced stages such as Severe or Proliferative DR. Variations in imaging quality and acquisition settings may also influence real-world performance. Future work should therefore emphasize larger external validation, expanded

disease-stage classification and broader clinical deployment testing.

Future Scope

Even while DiabFoRMaxNet performs exceptionally well in identifying and categorizing DR from OCTA and fundus pictures, there are a number of interesting research avenues that could improve its capabilities even more. First, a more thorough grading system appropriate for long-term patient monitoring would be offered by expanding the algorithm to anticipate new DR severity levels, such as severe and proliferative stages. Second, by providing quantitative and visual explanations for every choice, explainable AI (XAI) mechanisms like Grad-CAM or SHAP will increase clinical acceptance and build ophthalmologists' trust. Third, the system can be extended to handle 3D volumetric OCT/OCTA data, which would allow for better identification of minute microvascular changes and deeper vascular layer analysis. Furthermore, combining domain adaptation and transfer learning would make it easier to apply the model to datasets from various imaging equipment or demographic groups, guaranteeing increased generalizability. When combined with real-time inference and integration into teleophthalmology platforms, a cloud-based deployment of DiabFoRMaxNet has the potential to greatly increase accessibility in underserved and rural areas. Last but not least, investigating multi-disease retinal analysis—in which DR detection is coupled with screening for hypertensive retinopathy, age-related macular degeneration (AMD) and glaucoma—could turn DiabFoRMaxNet into a complete AI-assisted ophthalmic diagnostic system, bridging the gap between operational clinical tools and research prototypes.

Abbreviations

DiabFoRMaxNet: Hybrid Deep Learning Model formed by Cascading of AlexNet, FCN and RankMax Classifier, DR: Diabetic Retinopathy, FCN: Fully Connect Network, FOA: Fossa Optimization Algorithm, OCTA: Optical Coherent Tomography Angiography.

Acknowledgement

The authors would like to sincerely acknowledge Jawaharlal Nehru Technological University, Anantapur and Annamacharya Institute of Technology and Sciences, Rajampet, for providing the necessary academic support, research facilities and institutional environment that enabled the successful completion of this research work.

Author Contributions

Peddapullaiahgari Hariobulesu: Conceptualization, Methodology, Software, Image preprocessing, Model implementation, Experimental investigation, Formal analysis, Validation, Visualization, Writing – original draft preparation, Fahimuddin Shaik: Supervision, Methodology review, Validation, Formal analysis guidance, Writing – review and editing, Project administration, Final approval of the manuscript.

Conflict of Interest

The authors declare that there is no conflict of interest associated with this manuscript.

Data Availability

The datasets used in this study are publicly available multimodal retinal fundus and OCTA image datasets and can be accessed from the corresponding sources cited in the manuscript references. Additional data supporting the findings of this study are available from the corresponding author upon reasonable request.

Declaration of Artificial Intelligence (AI) Assistance Process

The authors declare that generative AI or AI-assisted technologies were not used for content generation, data analysis, result interpretation, or scientific writing of this manuscript. Only limited open-source tools were utilized for basic assistance such as grammar checking and formatting. The entire conceptualization, methodology, experimentation, analysis and interpretation of results are original contributions of the authors. The authors take full responsibility for the content's originality, interpretation and accuracy.

Ethics Approval

This study did not involve human participants, animals, or clinical trials. All datasets used in this research are publicly available and were utilized in accordance with their respective usage policies. Therefore, ethical approval was not required.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Sushith M, Sathiya A, Kalaipoonguzhali V, Sathya V. A hybrid deep learning framework for early detection of diabetic retinopathy using retinal fundus images. Sci Rep. 2025; 15:15166. https://doi.org/10.1038/s41598-025-99309-w
2. Dixit RB, Jha CK. Fundus image based diabetic retinopathy detection using EfficientNetB3 with squeeze and excitation block. Med Eng Phys. 2025; 140:104350. https://doi.org/10.1016/j.medengphy.2025.104350
3. Singh SP, Gupta P, Dung R. Diabetic retinopathy detection by fundus images using fine-tuned deep learning model. Multimedia Tools Appl. 2024; 83(39):86657-79. https://doi.org/10.1007/s11042-024-19687-7
4. Alyoubi WL, Abulkhair MF, Shalash WM. Diabetic retinopathy fundus image classification and lesions localization system using deep learning. Sensors. 2021; 21(11):3704. https://doi.org/10.3390/s21113704
5. Butt MM, Iskandar DNFA, Abdelhamid SE, Latif G, Alghazo R. Diabetic retinopathy detection from fundus images of the eye using hybrid deep learning features. Diagnostics. 2022;12(7):1607. https://doi.org/10.3390/diagnostics12071607
6. Chua J, Sim R, Tan B, *et al.* Optical coherence tomography angiography in diabetes and diabetic retinopathy. J Clin Med. 2020;9(6):1723. doi: 10.3390/jcm9061723
7. Tey KY, Teo K, Tan ACS, *et al*. Optical coherence tomography angiography in diabetic retinopathy: a review of current applications. Eye Vis. 2019; 6:37. doi: 10.1186/s40662-019-0160-3
8. Wijesingha N, Tsai WS, Keskin AM, *et al*. Optical coherence tomography angiography as a diagnostic tool for diabetic retinopathy. Diagnostics. 2024;14(3):326. https://doi.org/10.3390/diagnostics14030326
9. Li Q, Zhu XR, Sun G, *et al.* Diagnosing diabetic retinopathy in OCTA images based on multilevel information fusion using a deep learning framework. Comput Intell Neurosci. 2022; 2022:4316507. doi: 10.1155/2022/4316507
10. El Habib Daho M, Li y, Zeghlache R, *et al*. Improved automatic diabetic retinopathy severity classification using deep multimodal fusion of UWF-CFP and OCTA images. nInternational Workshop on Ophthalmic Medical Image Analysis. 2023:11-20. https://ui.adsabs.harvard.edu/link_gateway/2023arXiv231001912D/doi:10.48550/arXiv.2310.01912
11. Abini MA, Priya SSS. Detection and classification of diabetic retinopathy using modified Inception V3. Int J Bioautomation. 2025; 29(1):77-92. doi: 10.7546/ijba.2025.29.1.001004
12. Gupta S, Khadka A. Diabetic retinopathy detection through multiclass classification of fundus image using convolutional neural network. Pokhara Eng Coll J. 2025; 2(1):47-56. https://doi.org/10.3126/pecj.v2i1.76831
13. Liu Z, Gao A, Sheng H, Wang X. Identification of diabetic retinopathy lesions in fundus images by integrating CNN and vision mamba models. PLoS One. 2025;20(1): e0318264. https://doi.org/10.1371/journal.pone.0318264
14. Gupta IK, Patil S, Mahadevkar S, Kotecha K, Mishra AK, Rodrigues JJPC. Retinal fundus imaging-based diabetic retinopathy classification using transfer learning and fennec fox optimization. MethodsX. 2025;14:103232. https://doi.org/10.1016/j.mex.2025.103232
15. Navaneethan R, Devarajan H. Enhancing diabetic retinopathy detection through preprocessing and feature extraction with MGA-CSG algorithm. Expert Syst Appl. 2024;249(8):123418. https://doi.org/10.1016/j.eswa.2024.123418
16. Abini MA, Devi S. A novel deep learning approach for diabetic retinopathy classification using optical coherence tomography angiography. Multimed Tools Appl. 2025;84(31):38613-51. https://doi.org/10.1007/s11042-025-20708-2?urlappend=%3Futm_source%3Dresearchgate.net%26utm_medium%3Darticle
17. Li Y, Yu B, Si M, *et al*. Enhancing diabetic retinopathy diagnosis: automatic segmentation of hyper-reflective foci in OCT via deep learning. Int Ophthalmol. 2025;45(1):79. https://doi.org/10.1007/s10792-025-03439-z
18. Jin Y, Gui F, Chen M, *et al.* Deep learning-driven automated quality assessment of ultra-widefield optical coherence tomography angiography images for diabetic retinopathy. Vis Comput. 2025;41:1049-59. https://doi.org/10.1007/s00371-024-03383-6
19. Rodríguez-Miguel A, Arruabarrena C, Allendes G, *et al.* Hybrid deep learning models for the screening of diabetic macular edema in optical coherence tomography volumes. Sci Rep. 2024;14(17633). https://doi.org/10.1038/s41598-024-68489-2
20. Riazi-Esfahani H, Jafari B, Azimi H, *et al*. Assessment of area and structural irregularity of retinal layers in diabetic retinopathy using machine learning and image processing techniques. Sci Rep. 2024;14(4013). https://doi.org/10.1038/s41598-024-54535-6
21. Bidwai P, Gite S, Gupta A, Pahuja K, Kotecha K. Multimodal dataset using OCTA and fundus images for the study of diabetic retinopathy. Data Brief. 2024; 52:110033. https://doi.org/10.1016/j.dib.2024.110033
22. Akram MU, Akbar S, Hassan T, Khawaja SG, Yasin U, Basit I. Data on fundus images for vessels segmentation, detection of hypertensive retinopathy, diabetic retinopathy and papilledema. Data Brief. 2020;29:105282. doi: 10.1016/j.dib.2020.105282.
23. Islam MM, Yang H, Poly TN, Jian W, Li Y. Deep learning algorithms for detection of diabetic retinopathy in retinal fundus photographs: a systematic review and meta-analysis. Comput Methods Programs Biomed. 2020;191:105320. https://doi.org/10.1016/j.cmpb.2020.105320
24. Wang Y, Yang J, Yang J, Zhao X, Chen Y, Yu WH. Progress of artificial intelligence in diabetic retinopathy screening. Diabetes Metab Res Rev. 2021;37(5):e3414. doi: 10.1002/dmrr.3414
25. Adriman R, Muchtar K, Maulina N. Performance evaluation of binary classification of diabetic retinopathy through deep learning techniques using

texture feature. Procedia Comput Sci. 2021;179(231-234):88-96.
https://doi.org/10.1016/j.procs.2020.12.012
26. Lakshminarayanan V, Kheradfallah H, Sarkar A, Balaji JJ. Automated detection and diagnosis of diabetic retinopathy: a comprehensive survey. J Imaging. 2021; 7(9):165.
doi: 10.3390/jimaging7090165.
27. Li F, Wang Y, Xu T, *et al.* Deep learning-based automated detection for diabetic retinopathy and diabetic macular oedema in retinal fundus photographs. Eye. 2022;36(7):1433–41.
doi: 10.1038/s41433-021-01552-8
28. Nadeem MW, Goh HG, Hussain M, *et al.* Deep learning for diabetic retinopathy analysis: a review, research challenges and future directions. Sensors. 2022;22(8):6780.
https://doi.org/10.3390/s22186780
29. Huang X, Wang H, She C, *et al.* Artificial intelligence promotes the diagnosis and screening of diabetic retinopathy. Front Endocrinol. 2022; 13:946915.
doi: 10.3389/fendo.2022.946915
30. Sánchez-Gutiérrez V, Hernández-Martínez P, Muñoz-Negrete FJ, *et al.* Performance of a deep learning system for detection of referable diabetic retinopathy in real clinical settings. arXiv. 2022.
https://doi.org/10.48550/arXiv.2205.05554
31. Chłopowiec AR, Karanowski K, Skrzypczak T, Grzesiuk M, Chłopowiec AB, Tabakov M. Counteracting Data Bias and Class Imbalance-Towards a Useful and Reliable Retinal Disease Recognition System. Diagnostics. 2023;13(11):1904.
https://doi.org/10.3390/diagnostics13111904
32. Grzybowski A, Singhanetr P, Nanegrungsunk O, Ruamviboonsuk P. Artificial Intelligence for Diabetic Retinopathy Screening Using Color Retinal Photographs: From Development to Deployment. Ophthalmol Ther. 2023 Jun;12(3):1419-37.
doi: 10.1007/s40123-023-00691-3
33. Wang Z, Li Z, Li K, Mu S, Zhou X, Di Y. Performance of artificial intelligence in diabetic retinopathy screening: a systematic review and meta-analysis of prospective studies. Front Endocrinol. 2023;14: 1197783.
https://doi.org/10.3389/fendo.2023.1197783
34. Xu X, Zhang M, Huang S, *et al*. The application of artificial intelligence in diabetic retinopathy: progress and prospects. Front Cell Dev Biol. 2024; 12:1473176.
doi: 10.3389/fcell.2024.1473176
35. Acevedo E, Orantes S, Acevedo A, Carreño R. Identification of eye diseases through deep learning. Diagnostics (Basel). 2025;15(7):916.
https://doi.org/10.3390/diagnostics15070916
36. Susilo YKB, Yuliana D, Mahadi M, *et al*. Artificial intelligence for early detection and prognosis prediction of diabetic retinopathy. medRxiv. 2025.
https://doi.org/10.1101/2025.03.29.25324873

How to Cite: Hariobulesu P, Shaik F. A Multimodal Deep Learning Framework for Optimized Detection and Classification of Diabetic Retinopathy. Int Res J Multidiscip Scope. 2026;7(3):153-172.
DOI: 10.47857/irjms.2026.v07i03.09496