Received: 9 November 2024 | Revised: 10 March 2025 | Accepted: 3 April 2025 | Published online: 14 May 2025

RESEARCH ARTICLE

Journal of Computational and Cognitive Engineering
2026, Vol. 5(2) 233–245
DOI: 10.47852/bonviewJCCE52024763

BON VIEW PUBLISHING

Early-Stage Diabetic Retinopathy Diagnosis with Feature Pyramid Networks and Spatial Pyramid Pooling Utilizing Full-Field Optical Coherence Tomography (FF-OCT)

Anitha Jaikumar[1,*] and Sreenivasa Chakravarthi Sangapu[1]

[1]*Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham-Chennai, India*

Abstract: Diabetes is a disease that is growing at the highest rate globally, and it has several impacts. One of such complications is diabetic retinopathy (DR), which causes damage to the retina and vision impairment. DR can be mild, and it can also be severe, and there are many shades in between. Therefore, it is possible to avoid vision loss if DR is diagnosed and treated in its early stages. Diagnosis of DR today is still a tedious process as only ophthalmologists can diagnose it using digitized colorful retinal fundus images. This article describes a new strategy for the early identification of DR that combines full-field optical coherence tomography (FF-OCT) with sophisticated deep learning (DL) algorithms, notably Feature Pyramid Networks (FPNs) linked with spatial pyramid pooling (SPP). FF-OCT produces high-resolution, cross-sectional pictures of the retina, revealing precise structural changes that occur before apparent symptoms of DR. We created a multi-scale DL framework that takes advantage of the capabilities of FPNs and SPP to efficiently process and evaluate the fine features of FF-OCT images. The FPN design detects changes in the retinal structure at different scales, which improves the network's performance in detecting early signs of DR. On the other hand, the SPP module collects contextual information from several sub-regions of the image to provide a stable and accurate feature representation regardless of the size and location of the lesion. These models were trained and validated on this dataset using performance indicators such as sensitivity, specificity, and area under the ROC curve (AUC). The findings of this paper suggest that the FPNs with the SPP model are superior to the traditional image analysis methods and the standard Convolutional Neural Network in diagnosing early-stage DR.

Keywords: diabetic retinopathy (DR), early-stage diabetic, deep learning (DL), full-field optical coherence tomography (FF-OCT), Feature Pyramid Networks (FPNs), spatial pyramid pooling

1. Introduction

Blood sugar regulation is compromised in people with diabetes. Elevated blood sugar levels impact the body's tissues and organs. Diabetes becomes more evident and poses a greater risk when blood sugar levels reach a maximum [1]. A common cause of diabetic retinopathy (DR), a condition that affects vision, is diabetes mellitus. Due to varying degrees of severity, early detection aids in management [2]. The retina's purpose is to convert light into electrical signals for proper image processing. It relies on nearby blood vessels for nourishment and oxygen. Diabetes can damage these blood vessels, reducing blood flow to the retina, compromising its functionality, and ultimately resulting in decreased vision. DR is a disorder characterized by retinal abnormalities that, if left untreated, can cause blindness. Currently, DR is diagnosed through a labor-intensive and time-consuming analysis of digital color retinal fundus images by qualified ophthalmologists. Any diabetic who develops chronic hyperglycemia is at risk of developing diabetic retinal disruptions (DR) due to progressive vascular interruptions in the retina [3–6].

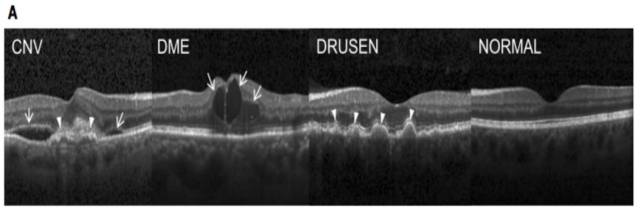
DR is usually classified into two types: proliferative DR (PDR) and non-proliferative DR (NPDR). Increased vascular permeability and capillary blockage inside the retinal blood vessels are two characteristics of PDR, commonly referred to as advanced DR. Here, neovascularization is thought to represent PDR, a more developed stage of DR [7]. Throughout this period, neuronal retinal degeneration and medically undetected microvascular abnormalities progress since DR is primarily asymptomatic in its early stages. Clinically, fundus examination is used to identify the first stage of DR [8]. These days, functional differences in the electroretinogram (ERG), retinal blood vessel diameter, and retinal circulation are used by researchers to identify DR using a variety of techniques. Because fundus images are noninvasive, quick, and well-tolerated, it is a commonly used technique for assessing the degree of DR. To accurately diagnose and measure the severity of diabetic eye disease, ophthalmologists use high-resolution fundus images to track retinal abnormalities [9]. The majority of fundus screens rely on the medical expertise of ophthalmologists and involve the manual examination of fundus images.

***Corresponding author:** Anitha Jaikumar, Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham-Chennai, India. Email: j_anitha@ch.students.amrita.edu

Thus, from a practical standpoint, automating the diagnosis and detection of DR is essential. Moreover, lesion segmentation is crucial for both diagnosing and detecting DR, and it is also utilized for evaluating the severity of DR later on [10].

Since it lessens the chance that NPDR will develop into PDR, early detection of DR is important and is used to avoid the worst effects in later stages. DR is often divided into four stages. Mild nonproliferative retinopathy is the first of the four phases in this case. The initial stage is distinguished by microaneurysms. Intermediate nonproliferative retinopathy is the disease's second stage. It occurs when the retina's blood arteries clog, reducing the quantity of blood that reaches the various retinal regions. Proliferative retinopathy, the last stage of the disease, results in vision impairment [11]. The detection of DR has been on the rise in recent years due to the application of different artificial intelligence (AI) methods, which in turn enhances the computational power required to develop new DL techniques. Figure 1 shows the optical coherence tomography image.

Figure 1
The optical coherence tomography image



The imaging technique known as retinal OCT is utilized to obtain high-resolution cross-sections of living patients' retinas. An estimated 30 million OCT scans are carried out annually, and a substantial amount of time is spent on picture analysis and interpretation. (A) The neovascular membrane (white arrowheads) and related subretinal fluid (arrows) are visible in the far left choroidal neovascularization (CNV). Diabetic macular edema (DME) with intraretinal fluid linked to retinal thickening (middle left) (arrows). (Center right) In early age-related macular degeneration (AMD), there are many drusen (arrowheads). (Long right) Normal retina without any retinal fluid or edema, with the foveal contour retained.

Based on full-field optical coherence tomography (FF-OCT) images, the study proposes a new approach for the early identification of DR by integrating spatial pyramid pooling (SPP) and Feature Pyramid Network (FPNs). This method combines the advantages of both methods to progress the presentation of the retinal image analysis. FPNs are capable of extracting multi-scale features from the retinal images, which helps in representing the various pathological signs at different resolutions. Furthermore, SPP enhances this ability because the model operates with data from several scales at once, allowing it to identify pathologies of any size and location in the retina. The addition of FF-OCT, a high-resolution imaging technique that provides cross-sectional images of the retinal architecture to this approach, has made it possible to improve the screening of DR at an early stage. By using these methods, clinicians are able to identify the presence of retinal abnormalities with greater sensitivity and at an earlier stage than would be possible without the use of these methods, which may prevent the development of severe vision loss.

1.1. Motivation of this research

The objective of this work is to increase the identification of DR in its early stages, which is among the major causes of blindness. Specifically, it uses FPN with SPP to diagnose DR at an early stage with the assistance of FF-OCT. The current methods of detection usually incorporate signs that are observable in the advanced stages of the disease, which means that there are missed opportunities for early diagnosis. The goal of this work is to use FPNs with SPP to improve the accuracy and efficiency of the processing of the FF-OCT data in order to find retinal structural alterations that might indicate the early stages of DR. This strategy is not only directed at improving the medical imaging technology but also has the possibility of altering the screening practices by providing timely therapy for the loss of vision in diabetes patients.

1.2. Research gaps

1) Many of the current models may not be very representative of the diverse populations and deviations that are present in real-life scenarios since the datasets used in their training were relatively small and homogeneous. This restricts the detection systems' capacity to be used generally and may influence their performance for various racial and ethnic groups.
2) DR can be more accurately identified by using multimodal data, such as OCT, fluorescein angiography, and patient-specific clinical information. The ideal methods for multimodal integration are still not well understood, and research on how to optimally integrate different data types with conventional fundus imaging is still in its early stages.

1.3. The main contribution of this research

1) To develop a novel technique for the early identification of DR that combines FF-OCT with sophisticated deep learning (DL) algorithms, notably FPNs linked with SPP.
2) FF-OCT enables precise cross-sectional imaging of the retina, indicating structural alterations that may develop before DR symptoms appear.
3) We created a multi-scale DL framework that takes advantage of the capabilities of FPNs and SPP to efficiently process and evaluate the fine features of FF-OCT images.
4) Finally, the FPN design detects retinal changes at numerous scales, improving the network's capacity to spot subtle indicators of early DR.

The organization is as follows: Section 2 discusses earlier work in DR detection and categorization, whereas Section 3 demonstrates how the proposed technique works. The results of the established work are presented in Section 4, and the paper is concluded in Section 5.

2. Literature Survey

DR, the main cause of vision loss in diabetics, must be identified and treated as quickly as possible. Recent developments in DL are beneficial for enhancing the early-stage DR detection's accuracy and speed. Conventional diagnostic methods, though effective, often involve the use of fundus images and are based on the observer's interpretation, which may not identify early signs of the disease. The use of Convolutional Neural Networks (CNNs) and other Machine Learning (ML) techniques may improve the early identification of DR among patients. This literature review is concerned with the enhancement and utilization of these methods and their effects on patients' outcomes, early identification, and diagnostic accuracy.

As the incidence of DR has risen, many techniques for screening this disease have been developed. Shankar et al. [12] first introduced DNN-MSO, a method for detecting and categorizing

DR images. This method comprises several steps, which include segmentation, preprocessing, classification, and feature extraction. contrast limited adaptive histogram equalization (CLAHE) was applied first to enhance the difference of the DR images. Following this preprocessing step, the histogram method was applied to segment the DR images. Last but not least, the DNN-MSO classifier was used to classify the extracted feature vectors.

In their study, Bhardwaj et al. [13] developed a hierarchical grading system for DR utilizing two datasets, MESSIDOR and IDRiD. They extracted textural features from the MESSIDOR dataset, including form, intensity, and elements of the Gray-Level Co-occurrence Matrix. The accuracy of the K-nearest neighbor (KNN) and support vector machine (SVM) algorithms was 95.30% and 92%, respectively. The percentage of positive responses to the questions was 60% and 70%, respectively, after the features were implemented. The KNN classifier achieved an accuracy of 94%. The accuracy achieved for the given datasets is 98% for the DIARETDB1 dataset and 00% for the IDRiD dataset. Nevertheless, their suggested technique would not work well for a large and difficult dataset like Kaggle-EyePACS.

Luo et al. [14] proposed a Self-supervised Fuzzy Clustering Network (SFCN) to address the challenge of labeling retinal dataset images. This approach significantly enhances performance as it does not need the time-consuming human annotation. The SFCN methodology is useful, but there is still room for enhancement in the DR detection and classification system, which may one day surpass the current state of the art in supervised learning.

Wu et al. [15] suggested that the input data be classified into two types: between DR-affected images and non-DR images. This is done using a coarse network. A coarse network is a network that is made up of a large number of nodes. Integrating attention gate modules into the CNN design offers two benefits: improving the contrast of the lesion and reducing the amount of interference from other areas of the image. The fine network then classifies the DR classifications provided by the coarse network into mild, moderate, severe, and proliferative phases of DR. Two datasets are used in this method: EyePACS, which contains photos of DR taken with various cameras from both the left and right eyes, and the IDRiD, which consists of images of DR taken at an Indian eye clinic utilizing a Kowa VX-10α digital fundus camera. In the IDRiD dataset and the EyePACS dataset, the best accuracy of the model was 56.19% and 83.1%, respectively.

Two DL architectures were created by Mohanty et al. [16] in order to diagnose and classify DR: the DenseNet 121 network and a novel model that combines the VGG16 architecture with an XGBoost classifier. Retinal images from the APTOS 2019 Blindness Detection Kaggle dataset were used to assess these models. Appropriate balancing procedures were used to correct the dataset's imbalance of image classes.

To enhance the early diagnosis of DR, AbdelMaksoud et al. [17] proposed the E-DenseNet hybrid model. This model was intended to overpower the complexities associated with distinguishing DR from retinal images employing CNNs. It is possible that conventional CNNs may not be able to precisely differentiate among dissimilar kinds of lesions with different features. To overcome this, the Eyenet and DenseNet models were fused to form the E-DenseNet, which is a novel architecture. The E-DenseNet model proved to be highly efficient with a computation time of only 3.5 minutes. In addition to achieving a near-perfect average accuracy of 91.2%, it also demonstrated 69% specificity, 96% sensitivity, a 92.45% Dice similarity coefficient, and a 0.883 quadratic kappa score.

Khan et al. [11] have discussed the reduction in model training and convergence time for DR categorization. They used a network-in-network, SPP layer, and VGG16. The fundus retina images were subjected to data preparation techniques such as scaling, cropping, normalization, and augmentation. Based on the VGG16 design, the network-in-network and SPP layers were put into practice. Using the SPP layer, the last convolutional layer of VGG16 was linked to the first fully connected layer in order to avoid cropping and information corruption. Using the Xavier technique, the network-in-network layer was initialized. The final fully linked layers of VGG16 were then modified to learn nonlinear features. This approach was superior to other methods with an AUC of 0.95 in the Kaggle dataset and a 52% parameter reduction.

2.1. Problem identification of existing systems

1) Diabetes frequently results in DR, a problem that is difficult to identify and treat before major vision loss occurs since it often advances without symptoms in its early stages.
2) Many diabetic patients may not have access to routine eye exams, resulting in delayed diagnosis and treatment of DR, a condition that could have been prevented with earlier identification.
3) Because DL and contemporary imaging technologies are not fully integrated, there are few opportunities for early intervention and management of DR.
4) Diagnosing early-stage DR can be challenging since it might present with modest or overlapping visual symptoms, making it difficult to differentiate from other eye disorders.

3. Proposed Model

This section introduces a novel approach to diagnose DR at an early stage by integrating FPNs combined with SPP with FF-OCT. FF-OCT offers high-resolution cross-sectional images of the retina and thus can detect structural changes in DR before clinical signs appear. To make the best of the FPNs and SPP, we proposed a multi-scale DL framework for the FF-OCT image analysis. The FPN design identifies retinal alterations at multiple resolutions, enhancing the network's ability to identify early signs of DR. However, the SPP module compiles contextual information from several sub-regions of the picture to provide a stable and accurate feature representation independent of lesion size and location. The dataset for this study included FF-OCT scans of normal people and diabetic patients at various retinopathic phases. Figure 2 shows the block diagram for the proposed FF-OCT approach.

3.1. Dataset description

This dataset is a collection of OCT images, primarily used for the diagnosis and classification of retinal diseases. The dataset includes a comprehensive set of grayscale OCT scans categorized into four distinct classes: Normal, CNV, DME, and Drusen. It is derived from clinical settings, ensuring high-quality imaging standards for robust analysis and model training. This dataset is particularly useful for training ML models to detect pathological conditions in retinal images, making it a valuable tool in developing automated diagnostic systems. With over 84,495 images split into training, validation, and test sets, the dataset provides ample data for supervised learning tasks, including image classification and segmentation. The images are organized into folders by category, simplifying preprocessing and enabling efficient model pipeline development. Figure 3 displays the sample images for dataset.

Figure 2
Block diagram of FF-OCT proposed method

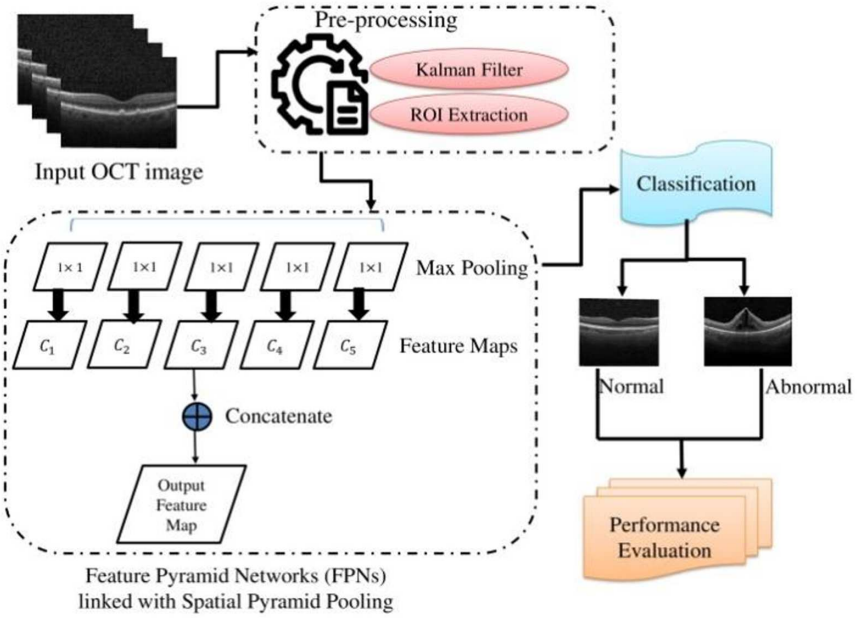
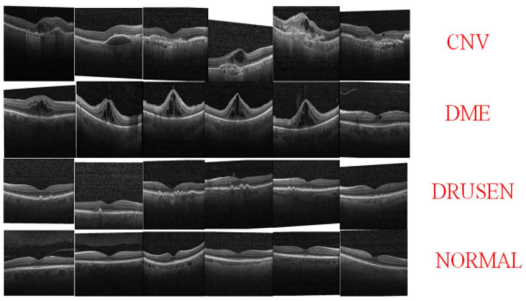


Figure 3
Sample images for dataset



3.2. Preprocessing

The input image is taken from a dataset, and noise and outliers are reduced using a Kalman filter. This leads to a significant improvement in the input image's quality [18].

3.2.1. Kalman filter

Considering its high computational efficiency, the Kalman filter is the most frequently used version of the Bayes filter. The definition of the Kalman filter in this case for the input image is:

$$\psi(f_w) \approx \wp(f_w; \lambda_w, \sum_w) \quad (1)$$

$\sum_w$ stands for estimate uncertainty, and a wider distribution is produced by a bigger covariance. λ_w stands for mean distribution $\wp(f_w; \lambda_w, \sum_w)$ also represents the likelihood of f_w for a Gaussian distribution with a covariance of $\sum_w$ and a mean of λ_w.

3.2.2. ROI extraction

The computational hurdles are greatly minimized by extracting a specific portion of an image using the Region of Interest (ROI) extraction [19] approach. The ROI is computed by accounting for the preprocessed image's pixel intensity values. To locate crucial areas in a noisy image, the ROI extraction technique is used. The scattered component is segmented into non-responsive regions to provide an efficient classification strategy. ROI, or pixel intensity value, defines the border for the extracted portion of an input image. To eliminate exterior misrepresentations, external components have been removed from the preprocessed images. The ROI extraction result is shown as Figure 2.

3.3. Full-field optical coherence tomography

FF-OCT is a high-resolution imaging technique utilized to gain detailed cross-sectional images of biological tissues. FF-OCT provides better imaging speed and resolution since it captures the whole field of vision simultaneously, in contrast to standard OCT, which uses a scanning beam [20].

Low-coherence interferometry forms the basis of FF-OCT. This technique requires a light source, a beam splitter, a reference mirror, and an example in its most basic version. The source's light is split into two pathways: one illuminates the sample, while the other proceeds toward the reference mirror. A camera records the interference pattern produced when the light reflected from both pathways is combined again.

3.3.1. Interference and coherence

The interplay of light waves replicated from the example and the reference mirror results in an interference pattern.

$$I(x,y) = I_s(x,y) + I_r + 2\sqrt{I_s(x,y)I_r}\cos(\phi(x,y) - \phi_r) \quad (2)$$

Where:

1) The sample's reflected light intensity at position (x,y) is measured at $I_s(x,y)$.
2) I_r is the reference mirror's light intensity reflection.

3) The phase of light reflected by the example is indicated by $\varphi_s(x, y)$.
4) φ_r represents the quantity of light that the reference mirror reflects.

3.3.2. Depth resolution

The Δ_z depth resolution in FF-OCT is defined by the light source's coherence length Lc.

$$\Delta_z = \frac{2In2}{\pi} - \frac{\lambda_0^2}{\Delta\lambda} \tag{3}$$

Where:

1) The light source's central wavelength is represented by λ_0.
2) The light source's spectral bandwidth is represented by $\Delta\lambda$.

3.3.3. Imaging process

1) Illumination: The sample is illuminated by a light source that emits a wide range of light.
2) Interference: Light reflected from the example and the reference mirror interferes, creating an interference pattern.
3) Detection: A camera captures the interference pattern across the entire field of view
4) Reconstruction: An exhaustive, high-resolution cross-sectional image of the example is produced using the data gathered.

3.3.4. Signal processing

To obtain depth information, Fourier transform algorithms are applied to the signals collected by the camera. The signal that was found over time t, indicated as $I(x, y, t)$, is:

$$I(x, y, t) - I_s(x, y) + I_s + \{2\sqrt{I_s(x, y)}I_r \cos(2kz + \varnothing_s(x, y)) - \varnothing_r\} \tag{4}$$

The sample depth profile can be obtained from the temporal signal by applying the Fourier transform.

$$\begin{aligned} F\{I(x, y, t)\} = {} & F\{I_s(x, y)\} + F\{I_r\} \\ & + F\{2\sqrt{I_s(x, y)}I_r \cos(2kz + \emptyset_s(x, y)) - \emptyset_r\} \end{aligned} \tag{5}$$

The Fourier transform helps in the differentiation of the various frequencies, which enables the determination of depth information.

3.3.5. Applications

In biomedical imaging, particularly for detailed pictures of biological samples and tissues, including the skin and retina, FF-OCT is routinely used. In clinical diagnosis and research, it is valuable as it provides structural information. Using low-coherence interferometry and Fourier transform signal processing, FF-OCT is a novel imaging method producing high-resolution depth-sectioning pictures of biological tissues.

3.4. Feature Pyramid Networks (FPNs)

FPNs integrate lateral connections with top-down architecture to improve object identification. This approach generates a feature pyramid by combining high-level traits, which are more semantically rich, with low-level qualities, which are exact in their geographical location. This lets the network spot objects at different diameters [21].

3.4.1. Architecture

Backbone Network: A standard CNN serves as the backbone, providing feature maps at dissimilar levels. Let $\{C_1, C_2, C_3, C_4, C_5\}$ be the feature maps from dissimilar stages of the backbone network, where C_i is the feature map from the ith stage.

Top-Down Pathway: The top-down approach increases the resolution of lower-level feature maps by integrating them with higher-level feature maps through lateral connections. Let P_i denote the feature map at level iii of the pyramid. The upsampling process is defined as:

$$P_i^{up} = Upsample(P_{i+1}) \tag{6}$$

where *Upsample* is typically done using bilinear interpolation.

Lateral Connections: To improve spatial characteristics, each upsampled feature map P_i^{up} is combined with the lateral feature map C_i from the backbone.

$$P_1 = Conv(P_i^{up} + C_i) \tag{7}$$

Here, *Conv* represents a convolutional layer used to merge and refine the feature.

Feature Pyramid: The last feature maps $\{P_1, P_2, P_3, P_4, P_5\}$ form the feature pyramid, which is used for object detection. These maps contain information at different scales, allowing for the detection of objects ranging from small to large sizes.

Finally, FPNs construct a multi-scale feature pyramid by combining high-level and low-level feature maps through top-down and lateral connections. This results in more robust object detection across varying scales. The key operations involve upsampling, addition, and convolution.

$$Upsampling : P_i^{up} = Upsample(P_{i+1}) \tag{8}$$

$$Feature\ Fusion\ : P_i = Conv\left(P_i^{up} + C_i\right) \tag{9}$$

3.5. Spatial pyramid pooling (SPP)

SPP is used to handle varying object sizes and aspect ratios by pooling features from different spatial bins. The key idea is to pool features over multiple spatial scales and then concatenate them into a fixed-size representation.

Pooling: Given a feature map F of size $H \times W$, the SPP layer divides F into multiple spatial bins. For each bin configuration (e.g., 1×1, 2×2, 4×4), it performs max pooling or average pooling to extract features.

Let's denote the pooling operation for a configuration with $k \times k$ bins as $SPP_k(F)$. For each bin, pooling extracts:

$$SPP_k(F) = [pool(F)_{1\times1}, pool(F)_{2\times2}, pool(F)_{4\times4}] \tag{10}$$

where $pool(F)_{1\times1}$ represents pooling over the entire feature map, $pool(F)_{2\times2}$ represents pooling over 2×2 bins, and so forth.

Concatenation: The pooled features from different bin configurations are concatenated to form a fixed-size representation.

$$SPP(F) = concat(SPP_{1\times1}(F), SPP_{2\times2}(F), SPP_{4\times4}(F)) \tag{11}$$

4. Result

4.1. Experimental setup

The FF-OCT was applied using the Python 3.9 programming language and the TensorFlow software framework. TensorFlow was chosen for the purpose of building and training an FPN because of its ease of use. Moreover, Keras, a neural network framework built on top of TensorFlow, may be used as a network that has already been trained. An Intel Core i7-7500U laptop with 16 GB of RAM and a laptop motherboard was used as the gear for developing and training the FPN. In this research, the existing models used are CNN-DenseNet [22], ResNet50 [23], Principal Component Analysis-Color Fundus Photography (PCA-CFP) [24], and Deep Restricted Boltzmann Machine (DRBM) [25]. The experimental setup settings are described in detail in Table 1.

Training a model for the proposed method FF-OCT involves preprocessing high-resolution retinal images to reduce noise and enhance quality, followed by data augmentation to increase dataset diversity. A DL architecture, such as FPN with SPP, is used to extract multi-scale features and detect subtle retinal changes. Key hyperparameters like learning rate, batch size, and epochs are optimized for performance, while loss functions and regularization techniques prevent overfitting. The model is trained on high-end GPUs, with evaluation metrics like accuracy and Dice coefficient ensuring robust generalization for tasks like early-stage DR diagnosis.

4.2. Performance metrics

The performance measures include accuracy, Matthews Correlation Coefficient (MCC), Negative Predictive Value (NPV), False Positive Rate (FPR), and False Negative Rate (FNR), among others. The following is the suggested mathematical model:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$FPR = \frac{FP}{FP + TN} \tag{13}$$

$$FNR = \frac{FN}{FN + TP} \tag{14}$$

$$MCC = \frac{TP*TN - FP*FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{15}$$

$$NPV = \frac{TN}{TN + FN} \tag{16}$$

4.2.1. Accuracy analysis

The accuracy of the FF-OCT method is associated with different methods in Figure 4 and Table 2. The graph shows how the DL approach effectively improves accuracy. For example, the FF-OCT model has an accuracy of 91.45% for 100 data points, whereas the CNN-DenseNet, ResNet50, PCA-CFP, and DRBM

Figure 4
Accuracy analysis for FF-OCT method

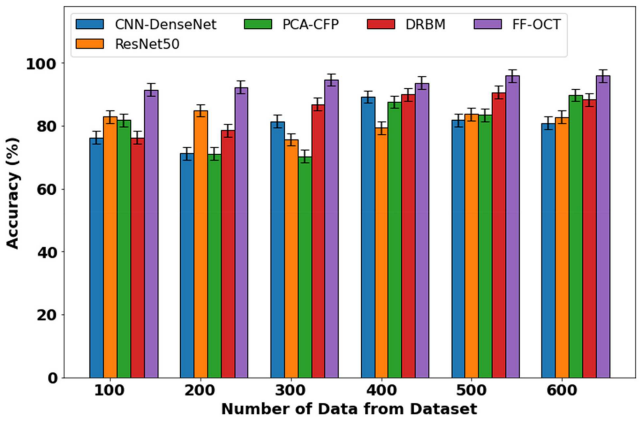


Table 1
Experimental setup for the proposed method FF-OCT

Hyperparameter	Description	Typical values	Significance in FF-OCT implementation
Learning Rate	Controls the step size during gradient descent optimization.	10–4 to 10–5	Determines how quickly the model learns. Too high can cause instability; too low can slow convergence.
Batch Size	Number of FF-OCT images processed in one forward/backward pass during training.	8 or 16	Larger batch sizes require more memory but provide stable gradients. Smaller sizes introduce noise.
Number of Epochs	Number of complete passes through the training dataset.	100–500 (with early stopping)	Prevents underfitting (too few epochs) or overfitting (too many epochs).
Optimizer	Algorithm used to update model weights during training.	Adam, SGD with momentum	Adam is preferred for adaptive learning rates; SGD with momentum is used for fine-tuning.
FPN Pyramid Levels	Number of levels in the Feature Pyramid Network for multi-scale feature extraction.	3–5	Determines the granularity of feature extraction for detecting subtle retinal changes in FF-OCT images.
SPP Bin Sizes	Spatial scales at which features are pooled in spatial pyramid pooling.	8 × 8	Captures contextual information at different scales, improving lesion detection in FF-OCT images.
Backbone Network	Base architecture for feature extraction in FPN	ResNet, EfficientNet	Pre-trained backbones (e.g., ResNet-50) are fine-tuned on FF-OCT data for efficient feature extraction.

Table 2
Accuracy analysis for FF-OCT technique

Number of data from dataset	CNN-DenseNet	ResNet50	PCA-CFP	DRBM	FF-OCT
100	76.28	82.89	81.83	76.31	91.45
200	71.27	84.87	71.11	78.55	92.33
300	81.45	75.67	70.32	86.81	94.65
400	89.24	79.41	87.56	89.98	93.66
500	81.76	83.66	83.45	90.67	95.89
600	80.89	82.81	89.76	88.34	95.91

models have accuracy values of 76.28%, 82.89%, 81.83%, and 76.31%, respectively. Additionally, the FF-OCT model has demonstrated its effectiveness across various data sizes. For instance, the accuracy for FF-OCT with 600 data points is 95.91%, whereas for CNN-DenseNet, ResNet50, PCA-CFP, and DRBM, the corresponding accuracy values are 80.89%, 82.81%, 89.76%, and 88.34%, respectively.

4.2.2. MCC analysis

The MCC of the FF-OCT methodology is contrasted with alternative methods in Figure 5 and Table 3. The effectiveness of the DL method in enhancing MCC is depicted in the graph. For 100 data points, the MCC values of the FF-OCT model are 87.18%, whereas the MCC values of the CNN-DenseNet, ResNet50, PCA-CFP, and DRBM models are 60.67%, 61.899%, 63.45%, and 65.81%, respectively. Nevertheless, the FF-OCT model has demonstrated its effectiveness across various data sizes. In this regard, the MCC for FF-OCT with 600 data points is 95.91%, whereas for CNN-DenseNet, ResNet50, PCA-CFP, and DRBM, the corresponding MCC values are 80.78%, 82.56%, 84.11%, and 85.45%, respectively.

4.2.3. FPR analysis

An FPR evaluation of the FF-OCT technique is displayed in Figure 6 and Table 4 in comparison to alternative approaches. The graph shows that even with low FPR, DL technology performs better. On the other hand, the FPR values for the CNN-DenseNet, ResNet50, PCA-CFP, and DRBM models are 50.81%, 44.17%, 43.18%, and 33.56%, respectively. Meanwhile, the FF-OCT model exhibits an FPR of 21.98% with 100 data points. The FF-OCT model, nevertheless, performs best across various datasets, yielding low FPR values. The FPR value for the FF-OCT model with 600 data points is 30.32%, whereas the CNN-DenseNet, ResNet50, PCA-CFP, and DRBM models have respective FPR values of 53.67%, 49.99%, 43.87%, and 42.46%.

Figure 5
MCC analysis for FF-OCT method

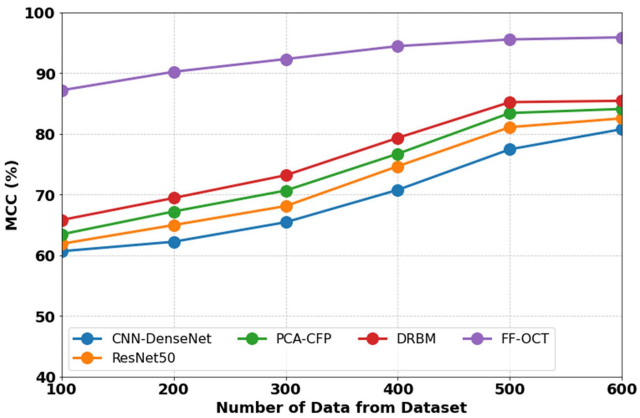


Figure 6
FPR analysis for FF-OCT method

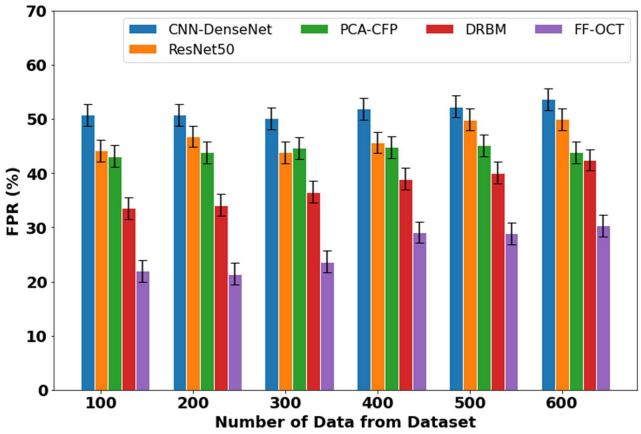


Table 3
MCC analysis for FF-OCT technique

Number of data from dataset	CNN-DenseNet	ResNet50	PCA-CFP	DRBM	FF-OCT
100	60.67	61.89	63.45	65.81	87.18
200	62.23	64.98	67.22	69.44	90.23
300	65.45	68.11	70.67	73.19	92.33
400	70.78	74.67	76.73	79.34	94.46
500	77.46	81.11	83.44	85.23	95.56
600	80.78	82.56	84.11	85.45	95.91

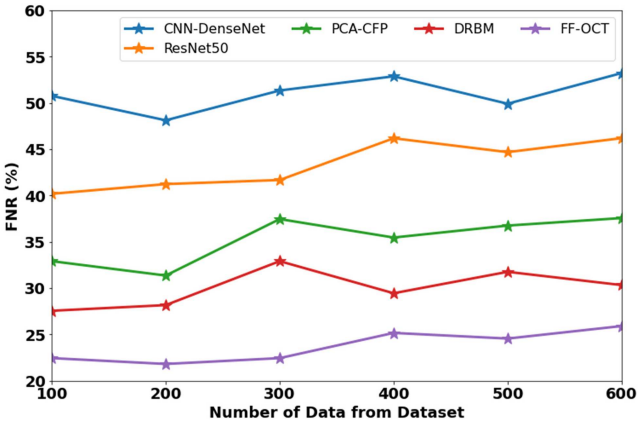
Table 4
FPR analysis for FF-OCT technique

Number of data from dataset	CNN-DenseNet	ResNet50	PCA-CFP	DRBM	FF-OCT
100	50.81	44.17	43.18	33.56	21.98
200	50.76	46.81	43.89	34.11	21.44
300	50.11	43.89	44.67	36.56	23.65
400	51.89	45.67	44.81	38.98	29.11
500	52.34	49.91	45.17	40.11	28.87
600	53.67	49.99	43.87	42.46	30.32

4.2.4. FNR analysis

Comparing the FF-OCT methodology to alternative approaches, an FNR assessment is presented in Figure 7 and Table 5. The DL technology performs better with a low FNR, as the graph shows. The CNN-DenseNet, ResNet50, PCA-CFP, and DRBM models had FPR values of 50.76%, 40.18%, 32.91%, and 27.56%, respectively. Meanwhile, the FF-OCT model exhibits an FPR of 22.45% with 100 data points. The FF-OCT model, nevertheless, performs best across numerous datasets, yielding low FNR values. The FNR value for the FF-OCT model with 600 data points is 25.91%, whereas the CNN-DenseNet, ResNet50, PCA-CFP, and DRBM models have respective FNR values of 53.22%, 46.19%, 37.56%, and 30.33%.

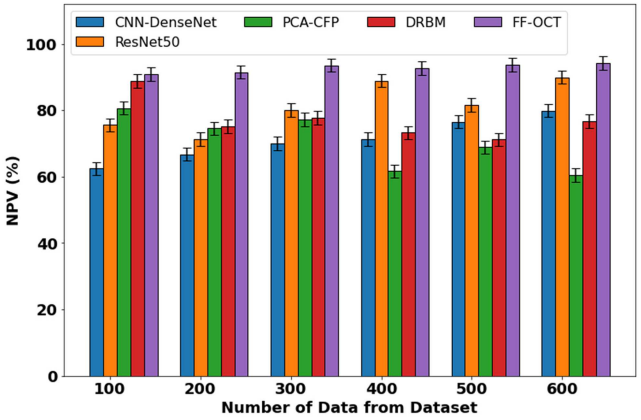
Figure 7
FNR analysis for FF-OCT method



4.2.5. NPV analysis

The NPV of the FF-OCT methodology is contrasted with alternative methods in Figure 8 and Table 6. The DL approach increases NPV, as shown in the graph. For example, the NPV values for

Figure 8
NPV analysis for FF-OCT method



the CNN-DenseNet, ResNet50, PCA-CFP, and DRBM models are 62.45%, 75.56%, 80.67%, and 88.81%, respectively, while the NPV for the FF-OCT model is 87.18% for 100 data points. Nevertheless, the FF-OCT model has demonstrated its effectiveness across various data sizes. In this regard, the NPV for FF-OCT with 600 data points is 94.17%, whereas for CNN-DenseNet, ResNet50, PCA-CFP, and DRBM.

4.2.6. Execution time analysis

Table 7 and Figure 9 compare the execution times of the proposed FF-OCT methodology to those of existing methodologies. FF-OCT has outperformed all other approaches. For example, the proposed FF-OCT technique takes only 1.675 ms to execute 100 data points. Other present approaches, such as CNN-DenseNet, ResNet50, PCA-CFP, and DRBM, required 12.567 ms, 11.224 ms, 8.876 ms, and 6.187 ms of execution time, respectively. Similarly, the proposed FF-OCT strategy executes 600 data points in 5.918 ms, whereas existing techniques such as CNN-DenseNet, ResNet50,

Table 5
FNR analysis for FF-OCT technique

Number of data from dataset	CNN-DenseNet	ResNet50	PCA-CFP	DRBM	FF-OCT
100	50.76	40.18	32.91	27.56	22.45
200	48.12	41.23	31.36	28.17	21.82
300	51.34	41.67	37.45	32.91	22.45
400	52.87	46.18	35.46	29.45	25.17
500	49.91	44.68	36.75	31.76	24.56
600	53.22	46.19	37.56	30.33	25.91

Table 6
NPV analysis for FF-OCT method

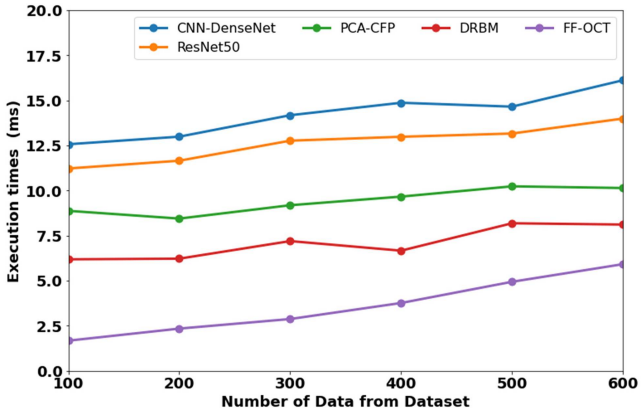
Number of data from dataset	CNN-DenseNet	ResNet50	PCA-CFP	DRBM	FF-OCT
100	62.45	75.56	80.67	88.81	90.91
200	66.78	71.34	74.56	75.18	91.45
300	69.98	79.98	77.21	77.78	93.55
400	71.22	88.87	61.65	73.24	92.65
500	76.56	81.54	68.87	71.18	93.67
600	79.91	89.95	60.44	76.65	94.17

Table 7
Execution time analysis for FF-OCT method

Number of data from dataset	CNN-DenseNet	ResNet50	PCA-CFP	DRBM	FF-OCT
100	12.567	11.224	8.876	6.187	1.675
200	12.987	11.654	8.446	6.223	2.345
300	14.176	12.765	9.186	7.198	2.876
400	14.867	12.981	9.665	6.667	3.765
500	14.651	13.161	10.234	8.186	4.941
600	16.115	13.991	10.141	8.115	5.918

PCA-CFP, and DRBM require 16.115 ms, 13.991 ms, 10.141 ms, and 8.115 ms, respectively.

Figure 9
Execution time analysis for FF-OCT technique



The FF-OCT method's training and validation losses on an 80:20 data split are displayed in Figure 10. The training loss measures the difference between the starting values and the expected performance within the training set. On the other hand, the validation loss evaluates the effectiveness of the FF-OCT approach on the validation set. The findings demonstrate the improved performance and classification accuracy of the FF-OCT system, as evidenced by the decreasing training and validation losses as the number of epochs increases. The method's enhanced capacity to recognize relationships and patterns is highlighted by the reduction in loss values.

Using an 80:20 ratio between the training and validation sets, Figure 10 illustrates the accuracy of the FF-OCT system during training and validation. When evaluating training accuracy, the training dataset is used, whereas validation accuracy is assessed using a separate testing dataset. The results show that the accuracy increases as the epoch counts increase for both training and validation. Therefore, as more epochs are added, the FF-OCT technique is more accurate on both datasets.

4.2.7. Comparative analysis

In Figure 11 and Table 8, the comparative analysis of the FF-OCT method is given with reference to other methods. For example, the FF-OCT values of accuracy, MCC, FPR, FNR, and NPV are 95. 91%, 95. 91%, 30. 32%, 25. 91%, and 94. 17%, respectively.

4.3. Ablation study

In an ablation study conducted with FPNs with SPP for FF-OCT-based early-stage DR detection, the effect of several components of the model architecture is analyzed. By removing or modifying certain parts of the model, the study assesses the effects of certain components like the FPN or the SPP layers on the model's robustness and precision in detecting early-stage DR. The model's performance in identifying pathological changes that may indicate early DR is enhanced by understanding how each component contributes to multi-scale features and spatial information from FF-OCT images.

4.3.1. Influence of the FPN

FPNs have significantly contributed to the early-stage DR detection since they enhance the models' performance in terms of identifying fine-grained information in the retinal images. FPN enhances the detection performance by using the multi-scale features that help the system to detect the early signs of DR. This capability is significant in early diagnosis and intervention since it increases the sensitivity and specificity of the automated systems in identifying early changes in the retina that may not be detected by other methods.

4.3.2. Influence of the K-fold cross-validation

In FF-OCT-based early-stage DR detection, the application of 10-fold cross-validation considerably increases model robustness and reliability. Ten segments, or subsets, of the dataset are divided

Figure 10
Analysis of accuracy and loss during training and validation

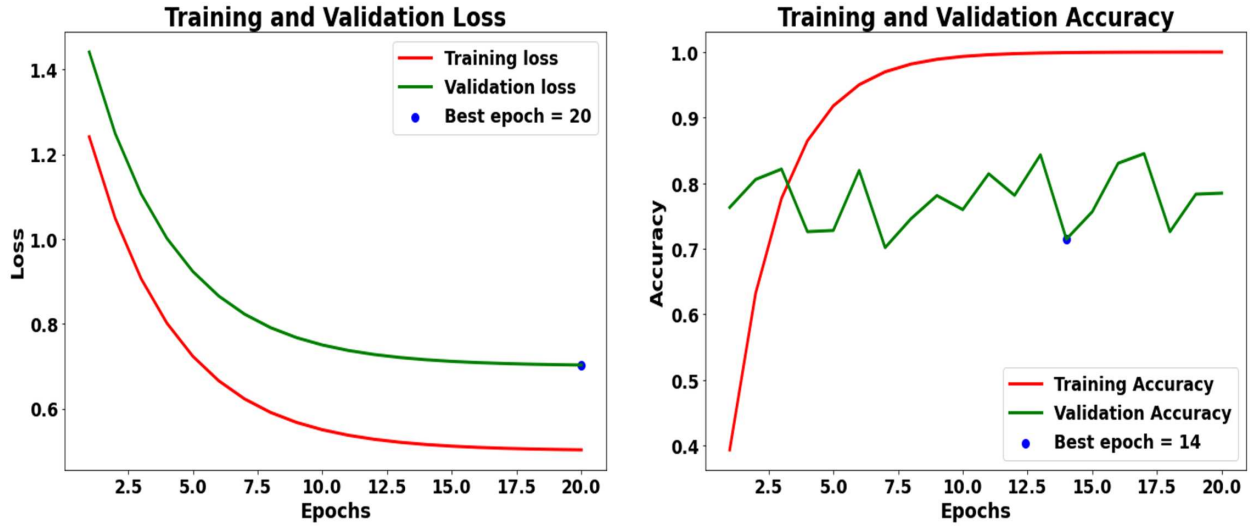


Figure 11
Comparative analysis for proposed model FF-OCT

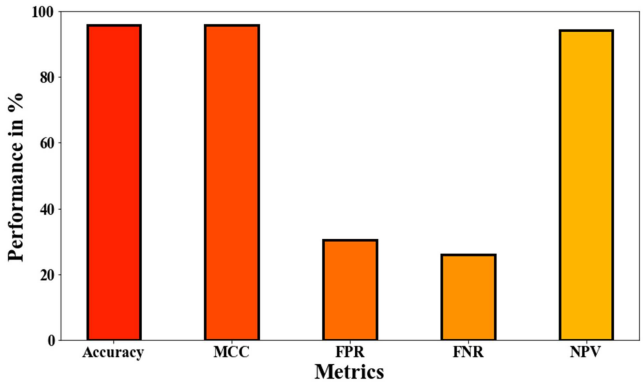


Table 8
Comparative analysis for proposed model FF-OCT

Methods	FF-OCT
Accuracy	95.91
MCC	95.91
FPR	30.32
FNR	25.91
NPV	94.17

Table 9
10-fold cross-validation

K-folds	Accuracy
1	0.93
2	0.92
3	0.95
4	0.95
5	0.96
6	0.94
7	0.93
8	0.96
9	0.94
10	0.95
10-Fold Mean	0.95

Table 10
A comparison of the proposed model to innovative validation analysis methodologies

Models	Evaluation methods	Accuracy (%)
SVM	10-fold cross-validation	81.33
KNN	10-fold cross-validation	81.33
DT	10-fold cross-validation	78.17
NB	10-fold cross-validation	77.77
Proposed	10-fold cross-validation	95.91

for the statistical method known as cross-validation. The remaining nine subsets are used for training, and one is used for validation. By employing 10-fold cross-validation, the suggested model, FF-OCT, obtained a 95.91% improvement in presentation for our input data. In association, the existing Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree (DT), and Naïve Bayes (NB) models obtained accuracy presentations of 81.46%, 81.43%, 77.22%, and 77.75%, respectively, as shown in Tables 9 and 10.

Figure 12 and Table 11 present a comparative analysis of the FF-OCT method and its performance on various datasets in relation to other existing methods. For instance, when evaluated on the DIARETDB1 dataset, the FF-OCT method achieved an accuracy of 95.91%.

Figure 12
Comparative analysis of proposed technique on dissimilar approaches and datasets

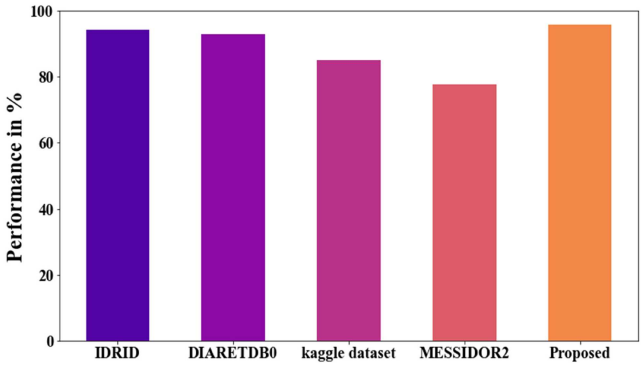


Table 11
Comparative investigation of the proposed technique using different approaches and datasets

Author	Dataset	Accuracy
Goel et al. [26]	IDRID	94.3
Bilal et al. [27]	DIARETDB0	93
Khan et al. [11]	Kaggle dataset	85
Luo et al. [28]	MESSIDOR2	77.75
Proposed	DIARETDB1	95.91

5. Conclusion

In this work, we used FF-OCT images to investigate the efficacy of FPN in conjunction with SPP for the identification of early-stage DR. The proposed method FF-OCT faces several limitations. First, even using sophisticated feature extraction methods like FPN and SPP, it can be difficult to identify the subtle clinical alterations typical of early-stage DR, which could result in a missed diagnosis. Furthermore, the need for sizable, excellent annotated datasets is a major limitation because it is challenging and time-consuming to get such datasets for early-stage DR. Another issue is computational complexity, which could impede real-time clinical application since the merging of FPN and SPP raises resource requirements. The system is resource-intensive as the mix of FPN and SPP raises processing demands. This may limit its application in real-time healthcare contexts or in locations with limited access to high-performance calculation infrastructure. The proposed methodology leverages FPN's prowess in collecting multi-scale characteristics and SPP's adaptability to different spatial resolutions, both of which are essential for detecting subtle pathological alterations indicative of early-stage DR. When compared to conventional procedures, the combination of these methodologies demonstrated a considerable improvement in both detection accuracy and sensitivity. FPN and SPP's improved feature extraction and pooling techniques allowed a more sophisticated analysis of FF-OCT images and helped to more precisely and quickly identify DR in its early stages. This advancement could lead to improved patient outcomes by significantly enhancing early diagnosis and treatment. Further research may explore the use of innovative AI methods to recognize the two types of exudates in retinal images from a variety of datasets.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/datasets/paultimothymooney/kermany2018.

Author Contribution Statement

Anitha Jaikumar: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Visualization. **Sreenivasa Chakravarthi Sangapu:** Conceptualization, Formal analysis, Investigation, Resources, Writing – review & editing, Supervision, Project administration.

References

[1] Kalyani, G., Janakiramaiah, B., Karuna, A., & Prasad, L. V. N. (2023). Diabetic retinopathy detection and classification using capsule networks. *Complex & Intelligent Systems*, *9*(3), 2651–2664. https://doi.org/10.1007/s40747-021-00318-9

[2] Kulyabin, M., Zhdanov, A., Nikiforova, A., Stepichev, A., Kuznetsova, A., Ronkin, M., . . . , & Maier, A. (2024). OCTDL: Optical coherence tomography dataset for image-based deep learning methods. *Scientific Data*, *11*(1), 365. https://doi.org/10.1038/s41597-024-03182-7

[3] Tsiknakis, N., Theodoropoulos, D., Manikis, G., Ktistakis, E., Boutsora, O., Berto, A., ..., & Marias, K. (2021). Deep learning for diabetic retinopathy detection and classification based on fundus images: A review. *Computers in Biology and Medicine*, *135*, 104599. https://doi.org/10.1016/j.compbiomed.2021.104599

[4] Wang, W., Pei, Y., Wang, S.-H., Manuel Gorriz, J., & Zhang, Y.-D. (2023). PSTCNN: Explainable COVID-19 diagnosis using PSO-guided self-tuning CNN. *Biocell*, *47*(2), 373–384. https://doi.org/10.32604/biocell.2023.025905

[5] Wang, W., Zhang, X., Wang, S.-H., & Zhang, Y.-D. (2022). Covid-19 diagnosis by WE-SAJ. *Systems Science & Control Engineering*, *10*(1), 325–335. https://doi.org/10.1080/21642583.2022.2045645

[6] Zhang, C., Chen, P., & Lei, T. (2023). Multi-point attention-based semi-supervised learning for diabetic retinopathy classification. *Biomedical Signal Processing and Control*, *80*, 104412. https://doi.org/10.1016/j.bspc.2022.104412

[7] Farag, M. M., Fouad, M., & Abdel-Hamid, A. T. (2022). Automatic severity classification of diabetic retinopathy based on DenseNet and convolutional block attention module. *IEEE Access*, *10*, 38299–38308. https://doi.org/10.1109/ACCESS.2022.3165193

[8] Özbay, E. (2023). An active deep learning method for diabetic retinopathy detection in segmented fundus images using artificial bee colony algorithm. *Artificial Intelligence Review*, *56*(4), 3291–3318. https://doi.org/10.1007/s10462-022-10231-3

[9] Agarwal, S., & Bhat, A. (2023). A survey on recent developments in Diabetic Retinopathy detection through integration of deep learning. *Multimedia Tools and Applications*, *82*(11), 17321–17351. https://doi.org/10.1007/s11042-022-13837-5

[10] Aswini, K. R. N., & Vijayaraghavan, S. (2022). Differentiation and identification of retinopathy of prematurity stages using DnCNN algorithm. *NeuroQuantology*, *20*(10), 11481–11494.

[11] Khan, Z., Khan, F. G., Khan, A., Rehman, Z. U., Shah, S., Qummar, S., ..., & Pack, S. (2021). Diabetic retinopathy detection using VGG-NIN a deep learning architecture. *IEEE Access*, *9*, 61408–61416. https://doi.org/10.1109/ACCESS.2021.3074422

[12] Shankar, K., Perumal, E., & Vidhyavathi, R. M. (2020). Deep neural network with moth search optimization algorithm based detection and classification of diabetic retinopathy images. *SN Applied Sciences*, *2*(4), 748. https://doi.org/10.1007/s42452-020-2568-8

[13] Bhardwaj, C., Jain, S., & Sood, M. (2021). Hierarchical severity grade classification of non-proliferative diabetic retinopathy. *Journal of Ambient Intelligence and Humanized Computing*, *12*(2), 2649–2670. https://doi.org/10.1007/s12652-020-02426-9

[14] Luo, Y., Pan, J., Fan, S., Du, Z., & Zhang, G. (2020). Retinal image classification by self-supervised fuzzy clustering network. *IEEE Access*, *8*, 92352–92362. https://doi.org/10.1109/ACCESS.2020.2994047

[15] Wu, Z., Shi, G., Chen, Y., Shi, F., Chen, X., Coatrieux, G., ..., & Li, S. (2020). Coarse-to-fine classification for diabetic retinopathy grading using convolutional neural network. *Artificial Intelligence in Medicine*, *108*, 101936. https://doi.org/10.1016/j.artmed.2020.101936

[16] Mohanty, C., Mahapatra, S., Acharya, B., Kokkoras, F., Gerogiannis, V. C., Karamitsos, I., & Kanavos, A. (2023). Using deep learning architectures for detection and classification of diabetic retinopathy. *Sensors*, *23*(12), 5726. https://doi.org/10.3390/s23125726

[17] AbdelMaksoud, E., Barakat, S., & Elmogy, M. (2022). A computer-aided diagnosis system for detecting various diabetic retinopathy grades based on a hybrid deep learning technique. *Medical & Biological Engineering & Computing*, *60*(7), 2015–2038. https://doi.org/10.1007/s11517-022-02564-6

[18] Memari, N., Ramli, A. R., Saripan, M. I. B., Mashohor, S., & Moghbel, M. (2019). Retinal blood vessel segmentation by using matched filtering and fuzzy c-means clustering with integrated level set method for diabetic retinopathy assessment. *Journal of Medical and Biological Engineering*, *39*(5), 713–731. https://doi.org/10.1007/s40846-018-0454-2

[19] Renukalatha, S., & Suresh, K. V. (2017). Automatic ROI extraction in noisy medical images. *ICTACT Journal on Image and Video Processing*, *7*(4), 1505–1514. https://doi.org/10.21917/ijivp.2017.0215

[20] Thouvenin, O., Apelian, C., Nahas, A., Fink, M., & Boccara, C. (2017). Full-field optical coherence tomography as a diagnosis tool: Recent progress with multimodal imaging. *Applied Sciences*, *7*(3), 236. https://doi.org/10.3390/app7030236

[21] Dang, J., Tang, X., & Li, S. (2023). HA-FPN: Hierarchical attention feature pyramid network for object detection. *Sensors*, *23*(9), 4508. https://doi.org/10.3390/s23094508

[22] Yasashvini, R., Vergin Raja Sarobin, M., Panjanathan, R., Jasmine, S. G., & Anbarasi, L. J. (2022). Diabetic retinopathy classification using CNN and hybrid deep convolutional neural networks. *Symmetry*, *14*(9), 1932. https://doi.org/10.3390/sym14091932

[23] Ali, G., Dastgir, A., Iqbal, M. W., Anwar, M., & Faheem, M. (2023). A hybrid convolutional neural network model for automatic Diabetic Retinopathy classification from fundus images. *IEEE Journal of Translational Engineering in Health and Medicine*, *11*, 341–350. https://doi.org/10.1109/JTEHM.2023.3282104

[24] Usman, T. M., Saheed, Y. K., Ignace, D., & Nsang, A. (2023). Diabetic Retinopathy detection using principal component analysis multi-label feature extraction and classification. *International Journal of Cognitive Computing in Engineering*, *4*, 78–88. https://doi.org/10.1016/j.ijcce.2023.02.002

[25] Shakibania, H., Raoufi, S., Pourafkham, B., Khotanlou, H., & Mansoorizadeh, M. (2024). Dual branch deep learning network for detection and stage grading of diabetic retinopathy. *Biomedical Signal Processing and Control*, *93*, 106168. https://doi.org/10.1016/j.bspc.2024.106168

[26] Goel, S., Gupta, S., Panwar, A., Kumar, S., Verma, M., Bourouis, S., & Ullah, M. A. (2021). Deep learning approach for stages of severity classification in diabetic retinopathy using color fundus retinal images. *Mathematical Problems in Engineering*, *2021*(1), 7627566. https://doi.org/10.1155/2021/7627566

[27] Bilal, A., Sun, G., Mazhar, S., Imran, A., & Latif, J. (2022). A transfer learning and U-Net-based automatic detection of diabetic retinopathy from fundus images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, *10*(6), 663–674. https://doi.org/10.1080/21681163.2021.2021111

[28] Luo, X., Pu, Z., Xu, Y., Wong, W. K., Su, J., Dou, X., ..., & Mou, L. (2021). MVDRNet: Multi-view diabetic retinopathy detection by combining DCNNs and attention mechanisms. *Pattern Recognition*, *120*, 108104. https://doi.org/10.1016/j.patcog.2021.108104

How to Cite: Jaikumar, A., & Sangapu, S. C. (2026). Early-Stage Diabetic Retinopathy Diagnosis with Feature Pyramid Networks and Spatial Pyramid Pooling Utilizing Full-Field Optical Coherence Tomography (FF-OCT). *Journal of Computational and Cognitive Engineering*, *5*(2), 233–245. https://doi.org/10.47852/bonviewJCCE52024763

Appendix

Abbreviations Description

DL	Deep Learning
DR	Diabetic Retinopathy
PDR	Proliferative DR
NPDR	Non-Proliferative DR
FPN	Feature Pyramid Networks
SPP	Spatial Pyramid Pooling
FA	Fluorescein Angiography
OCT	Optical Coherence Tomography
A/V	Artery/Vein
DMN	Diabetic Retinopathy Network
LBO	Ladybug Beetle Optimization
HPO	Hunter-Prey Optimizer
GLCM	Gray-Level Co-occurrence Matrix
KNN	K-Nearest Neighbor
DNN-MSO	Deep Neural Network-based Moth Search Optimization
SVM	Support Vector Machine
IDRiD	Indian Diabetic Retinopathy Image Dataset
HRF	High-Resolution Fundus
CNN	Convolutional Neural Network
SFCN	Self-supervised Fuzzy Clustering Network