



A Convnext And Resnet50-Driven Modified SLIC Approach For Diabetic Retinopathy Detection

¹Ankita Sinha, ²Dr. Manjusha Pandey, ³Dr. Siddharth Swarup Rautaray

¹Assistant Professor, ^{2,3}Associate Professor

¹Computer Science and Engineering,

¹Siliguri Institute of Technology, Siliguri, India

^{2,3}School of Computer Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar, India

¹ethosankita@gmail.com, ²manjushafcs@kiit.ac.in, ³siddharthfcs@kiit.ac.in

Abstract: Accurate segmentation and measurement of retinal fundus images are essential for diagnosing and treating eye diseases at earliest. However, this task is difficult due to the wide variability in the size, location, and shape of the fluid, along with their complex, curved boundaries accurate classification in minimalist time. To overcome these challenges, we propose a new modified simple linear iterative clustering (mSLIC) along with ConvNeXt, designed specifically for retinal fundus detection. This study begins by enhancing the images using an adaptive fuzzy filter, followed by segmenting them with a Modified Simple Linear Iterative Clustering (mSLIC) algorithm. This module captures long-range dependencies and non-local similarities while maintaining attention to local features. Additionally, it has spatial region-awareness, allowing it to adaptively concentrate on lesion areas. To extract important features, the segmented images are fed into the ResNet-50 deep learning model. Then, Particle Swarm Optimization (PSO) is used in a Selective Feature Pooling layer to enhance these features. The improved features are classified using ConvNeXt classifiers. Finally, the model's performance is tested using 10-fold cross-validation. When combined with PSO for feature selection and ConvNeXt for classification, the method achieved a top accuracy of 98.25% using 10-fold cross-validation. The inclusion of segmentation and optimization steps notably improved classification performance. Experimental results show that our proposed methodology surpasses existing literature study techniques in retinal detection performance.

Index Terms – Diabetic Retinopathy, fundus images, adaptive fuzzy, Simple Linear Iterative Clustering, ResNet-50, ConvNeXt.

Introduction

Diabetic retinopathy (DR) is a leading cause of vision loss globally and ranks among the most frequent causes of blindness in working-age adults (Wu et al, 2017). Early-stage screening plays a crucial role in controlling and potentially preventing vision loss caused by DR. Fundus photography serves as a key imaging technique to capture the back portion (fundus) of the eye, allowing visualization of critical anatomical features such as the retina, macula, and optic disc.

Radiologists analyze fundus images to identify early signs of diabetic retinopathy (DR), such as haemorrhages, microaneurysms, soft exudates, and hard exudates. Haemorrhages can manifest in various forms—dots, blots, or flame-shaped—depending on their location within the retinal layers (Basu et al., 2017). Microaneurysms are small, round bulges in the capillaries and are typically the earliest detectable signs of DR. They often appear as isolated red dots or small clusters and generally do not affect vision in the initial stages. Categorize haemorrhages and microaneurysms as red lesions, while soft and hard exudates are classified as bright lesions. Soft exudates, also known as cotton wool spots, are pale, greyish-white patches with blurred margins located in the nerve fiber layer. In contrast, hard exudates are yellow-white deposits within the retina. These waxy

accumulations result from lipid leakage caused by damaged capillaries (De et al., 2018). In advanced cases, hard exudates can arrange into circular patterns, known as circinate formations, which can severely impair vision.

Innovations in neural network design have been key to progress in representation learning. Convolutional neural networks (ConvNets) have greatly influenced computer vision by enabling automatic feature learning for many visual tasks, replacing manual feature design (Schmidt et al, 2020). More recently, transformers originally used in image processing have become popular due to their ability to scale well with bigger models and datasets. The ConvNeXt model has updated ConvNets to show that convolutional models can also scale effectively. Despite these advances, most architecture designs are still tested mainly by measuring their performance on the ImageNet dataset. Recently, CNNs learning approaches have demonstrated superior performance in DR classification and segmentation tasks compared to conventional machine learning methods (Li et al, 2019). These learning models perform pixel-level segmentation in fundus images, which enhances the accuracy of DR severity assessment. Retinopathy examination plays a vital role in identifying DR, which is essential for accurate diagnosis and effective treatment planning (fang et al, 2019). Manually analyzing these lesions requires significant time and a high level of precision from radiologists due to the variability in lesion morphology, size, shape, and color. To address this, Computer-Aided Diagnosis (CAD), (Rahimy et al, 2018) systems offer a reliable solution for fast and accurate DR detection, enabling timely medical intervention. In a CAD system, segmenting lesions is a fundamental step (Son et al, 2020). It allows the system to identify precise regions of interest, which helps in evaluating the severity of DR. Traditional segmentation methods that relied on basic image features have now given way to more advanced machine learning techniques, which offer greater accuracy (Chalakkat et al, 2019).

Faces difficulties in manual segmentation due to the wide variability in the size, location, and shape of the fluid, along with their complex, curved boundaries, hence accurate classification in minimalist time is not possible. To reduce the risk of misclassification, CAD systems assist radiologists by offering automated image assessments. With the advancement of Artificial Intelligence (AI), the accuracy and efficiency of analyzing fundus images have significantly improved. This study introduces a deep learning-modified variant of the Simple Linear Iterative Clustering (mSLIC) segmentation algorithm, tailored for the precise detection of retinal lesions, specifically microaneurysms and hard exudates. The key contributions of this research are outlined below:

- An adaptive fuzzy filter is the first step in the creation of pipeline for fundus images detection model followed by modified SLIC segmentation method known as mSLIC is used.
- After segmentation, images are subjected to the ResNet-50 model for feature extraction.
- Next, these extracted features are then refined using Particle Swarm Optimization (PSO) within a selective feature pooling layer, enhancing the most relevant attributes.
- Finally, Finally, the classification stage employs ConvNeXt classifiers to categorize the images. To ensure detection model robustness and reliability performance of the ConvNeXt classifier is evaluated using 10-fold cross-validation in different scenario.

The remaining part of this work is structured as follows: Section 2 is all about discussion on recently developed methodology for detecting diabetic retinopathy. The proposed methodology is outlined in Section 3. Also, provide an explanation of the classification procedure, the deep-weighted averaging feature fusion technique, and the selective feature pooling layer that uses PSO optimization approach. Section 4 provides a comparative analysis of the outcomes. Finally, Section 5 encapsulates the principal findings and offers suggestions for further study avenues.

I. LITERATURE REVIEW

Diabetic retinopathy is a diabetes-related condition that harms the retina and can cause vision loss (Serikbaeva et al, 2024), so early detection is important. Media haze (MH) is cloudiness in the cornea or lens and may signal eye problems (Ahmed et al, 2020). Optic disc cupping (ODC) helps detect glaucoma, which damages the optic nerve (Rebolleda et al, 2020). Hu et al. (Hu et al, 2018), developed a segmentation model called sASPP that uses special dilated (spread-out) convolutions. This method helps the model capture important details at different scales, improving accuracy and lowering the chance of overfitting. Feng et al. (Feng et al, 2020) introduced CPFNet, a model with two pyramid modules that combine information from different scales. This helped it perform well in tasks like segmenting the retina. Also utilizing the RFMID dataset, Shitole et al. (Shitole et al., 2020) employed deep learning to identify early indicators of twenty eye

disorders. They used ensemble approaches to assess models such as XceptionNet, EfficientNet-B7, ResNet125V2, EfficientNet-V2M, MobileNetV2, and EfficientNet-201. While classifying risks, their model achieved an accuracy of 96.08% and a multiple disease detection accuracy of 74.92%. In study (Alfian et al, 2020), researchers developed a method to predict diabetic retinopathy (DR) using individual risk factors. They combined a DNN with recursive feature elimination (RFE) to remove unnecessary data. Their approach, tested on a dataset of 115,000 patients and another smaller dataset from NHISS Korea, showed very high accuracy. Using stratified 10-fold cross-validation, they achieved 96.5% AUC. Although they compared their method with other machine learning models, they didn't specify which ones. Overall, their DNN with RFE approach outperformed others in predicting early DR.

Liu et al. (Liu et al, 2021), used attention gates to better handle the features passed through skip connections and added a special loss function to fix mistakes when different fluid areas in the retina are merged incorrectly.

Xing et al. (Xing et al, 2022), created a curvature loss function that uses knowledge about the shape of the fluid. This improved how well the model could outline the shapes and edges of fluid regions. A model was developed by Ho et al. (Ho et al, 2022) to identify 45 different eye illnesses employing color fundus images taken from the RFIDD dataset. The ensemble model for distinguishing between healthy and diseased eyes acquired an AUROC score of 0.9613, thanks to the usage of a multi-label convolutional neural network (ML-CNN). Among specific disorders, SE-ResNeXt had the best single model with a score of 0.9586 and an average AUROC of 0.9295.

Patel et al. (Patel et al, 2024), focused on detecting glaucoma using fundus images. Researchers used special features called GKDG within a 2D wavelet transform method (FAWT). Their model performed better than existing methods, achieving 95.84% accuracy, 97.17% specificity, and 94.55% sensitivity using tenfold cross-validation. Another study Rahman et al. (Rahman et al, 2024), looked at using hybrid machine learning methods to detect diabetic retinopathy. They tested k-Nearest Neighbors (kNN), Support Vector Machine (SVM), and Histogram-based Gradient Boosting (HGB) on the APTOS dataset. SVM performed best with 96.9% accuracy, while kNN and HGB followed with 95.6% accuracy. A fuzzy rule-based system for classifying retinal diseases was proposed in (Orlova et al, 2024), using two classifiers: rule-based (RB) and Stochastic Gradient Boosting (SGB). Their performance on unbalanced multi-class data was measured using MAUC and MMCC, showing a 5%–8% accuracy improvement over traditional methods. A method using extended fuzzy logic was proposed to detect diabetic retinopathy (Ahmed et al, 2023). Instead of just looking for obvious symptoms, it focused on subtle retinal changes seen in both diabetic and non-diabetic people. It used an 'f-validity' score based on risk factors and symptoms, which were turned into values using a membership function. These were combined using the OWA operator. The method showed good results, with 90% accuracy, 92.2% precision, and 75% sensitivity.

Researches shows that deep learning can help automate retinal image analysis, but current studies often use small, similar datasets that lack diversity. These limitations, along with little real-world testing, lead to problems like overfitting and poor clinical usefulness. Most models focus only on retinal diseases, ignoring how other eye conditions or image quality might affect results. This raises concerns about how well these models work in real-world settings.

To address these issues, this study developed a transfer learning model designed for better real-world use. Instead of clinical trials, the model was tested using strong evaluation methods, with a focus on recall and precision. Techniques like data preprocessing and augmentation were used to improve reliability and accuracy in identifying actual cases. The goal is to improve how well the model generalizes and performs outside of controlled research settings. Table 1 summarizes recent key studies.

Table 1 Literature review summary

Ref. & Year	Datasets	Classifiers	Results	Limitations
(Ho et al, 2019)	RFMID	SE-ResNeXt (CNN)	Au-roc = 96.13%.	Accuracy is impacted due to the limited dataset instances.
(Shitole et al, 2020)	RFMID	MobileNetV2, EfficientNetV2/B7, ResNet125V2, DenseNet-201, XceptionNet	Multi-disease detection - accuracy 76.92% and 96.98% for risk classification.	Lacks discussion on model limitations or biases.

(Muller et al, 2021)	RFMID	EfficientNetB, DenseNet201, InceptionV3, ResNet151, Logistic regression	Logistic regression presents highest Au-roc: 99.90%,	-
(Hu et al, 2022)	AI Challenger	Stochastic atrous spatial pyramid pooling (sASPP) segment SRF and PED lesions, ResNet50 (feature extraction)	SRF (DSC-0.8759, TPVF-0.8799, PPV-0.8881) PED (DSC-0.7371, TPVF-0.8341, PPV-0.6975)	DSC score can be improved.
(Ahmed et al, 2023)	Private dataset	Extended fuzzy logic	Accuracy- 90%, Precision-92.2%, Sensitivity- 75%.	Solely on a single disease, without using any segmentation methods.
(Patel et al, 2024)	APTOS	Flexible analytical wavelet transforms along with modified GKDG filter, and LS-SVM	Accuracy- 95.84 %, specificity- 97.17 %, sensitivity- 94.55 %.	Solely on a single disease, glaucoma, without using any segmentation methods.
(Rahman et al, 2024)	APTOS	SVM, KNN, HGB, Transfer learning (features selection)	SVM outperformed KNN and HGB with accuracy: 96.9%	Did not utilized any segmentation methods.
(Orlova et al, 2024)	Unbalanced dataset	Fuzzy logic, RB, SGB	MAUC (SGB- 91.22% and RB- 87.88%).	Restricted and class-imbalanced dataset.

II. PROPOSED METHODOLOGY

This section provides a detailed overview of the resources and methods used for classifying fundus images. Figure 1. illustrates the methodological framework followed in this study.

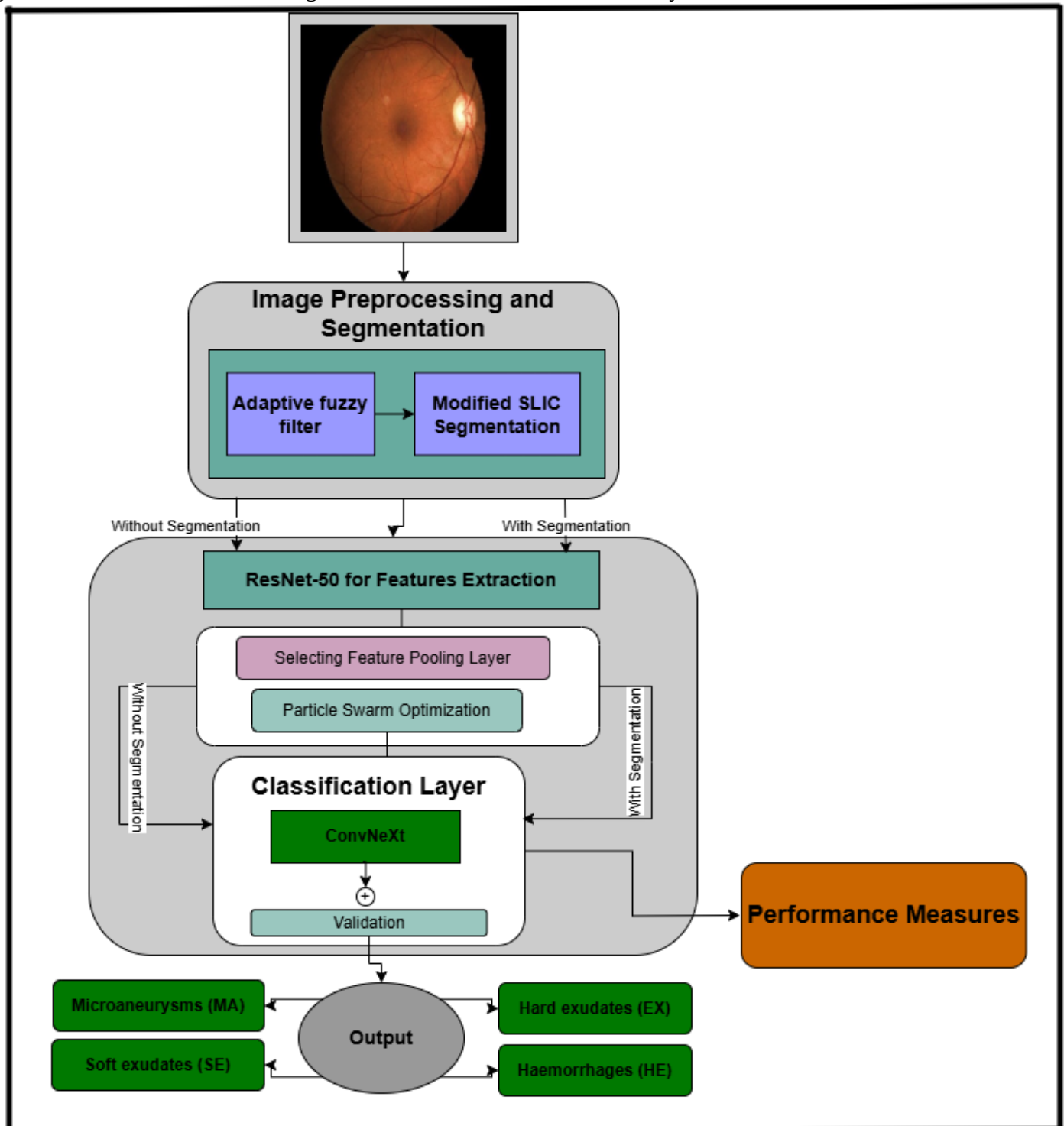


Figure. 2 Proposed Methodology

3.1 Dataset

The IDRiD dataset, which is hosted at "<https://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid>" and contains 516 high-resolution images measuring 4288×2848 pixels, were utilized in the proposed methodology. These images were taken by retinal specialists at an eye clinic in Nanded, Maharashtra, India (Porwal et al, 2018). Binary masks are used to annotate pixels in certain color fundus pictures for common DR abnormalities such as microaneurysms (MA), soft exudates (SE), hard exudates (EX), and hemorrhages (HE). Also, for each of the 516 images, you can get the DR severity grade.

3.2 Image Preprocessing

Analyzing retinal fundus images helps identify tumor types and predict patient outcomes. However, these images often suffer from complex textures, varying sizes, and noise. This study applies an adaptive fuzzy filter, (Kushwaha et al, 2004), to resized images (224 × 224) to improve clarity. The filter uses fuzzy

logic, (Van et al, 2003), to assess pixel intensity, contrast, and neighborhood variation, then updates each pixel based on a weighted average using fuzzy rules. This process removes noise and enhances image quality, making it easier to accurately identify and segment retinal key regions.

3.3 Segmentation: Modified SLIC (mSLIC) Algorithm

An image segmentation is made easier with the help of superpixels, which are groups of neighbouring pixels that have comparable brightness, color, and texture properties (Achanta et al, 2012). The SLIC algorithm initializes by uniformly distributing M seed points throughout the image. Given N total pixels and target of M superpixels, each superpixel approximately encompasses $\sqrt{N/M}$ pixels. Consequently, the seed points are spaced at an interval of $S = \sqrt{N/M}$ to ensure even coverage across the image (Csillik et al, 2017). Each superpixel center is defined by a feature vector $C_i = [l_i, a_i, b_i, x_i, y_i]^T$, which combines color values in the CIELAB space and pixel location. To improve segmentation, SLIC shifts each seed to the lowest gradient point in its 3×3 neighborhood. Then, it repeatedly assigns nearby pixels (within a $2S \times 2S$ area) to the closest center based on color and spatial similarity. This study presents a simplified mSLIC algorithm that makes calculations faster by using a 3D feature vector made up of spatial coordinates (x, y) and the grayscale value (g_s). The following equations are used to update the distance estimate between each pixel and the cluster center based on these features:

$$d_{(gs)}(i, j) = \sqrt{(g_i - g_j)^2} \quad (3.1)$$

$$d'_{(ts)}(i, j) = \sqrt{(d_s^2 - \alpha d_{gs})^2} \quad (3.2)$$

Here, d_{gs} denotes the total similarity of the grayscale values, including the spatial distance, while d_{ts} measures the degree of similarity between a pixel and the cluster center. The mSLIC algorithm then uses these values to segment the image into superpixels. The mSLIC algorithm is explained step-by-step below:

1. Convert the microscopic color image into a grayscale format and randomly divide it into k segments. Use an entropy-based evaluation to determine optimal threshold values that best separate the grayscale histogram. The goal is to maximize the entropy for ideal segmentation.
2. Partition the grayscale image into $K + 1$ distinct intensity intervals using the most suitable threshold values. For each interval, apply a contrast-enhancing transformation function that is convex and then concave with a turning point computed using the least squares method. The transformation function $f(x) = ax^r + b$ improves contrast, and the varying of r controls the enhancement strength.
3. Initialize clustering by placing superpixel centers on a grid, where each grid cell size is N/M , N being the number of pixels, and M the desired number of superpixels.
4. Refine each superpixel center by moving it to the lowest gradient position within its 3×3 neighborhood, ensuring that it is positioned on a relatively flat region.
5. For each pixel within a radius S of a center, compute a similarity distance considering both color and spatial proximity.
6. Assign each pixel to the nearest center if the new similarity distance is smaller than its current assignment, updating its label accordingly.
7. Repeat the refinement and assignment process until the cluster assignments stabilize. Recalculate the mean grayscale and position for each superpixel to update the centers.
8. Finally, merge small or isolated superpixels with nearby ones to improve overall smoothness and visual consistency as shown in Fig 2.

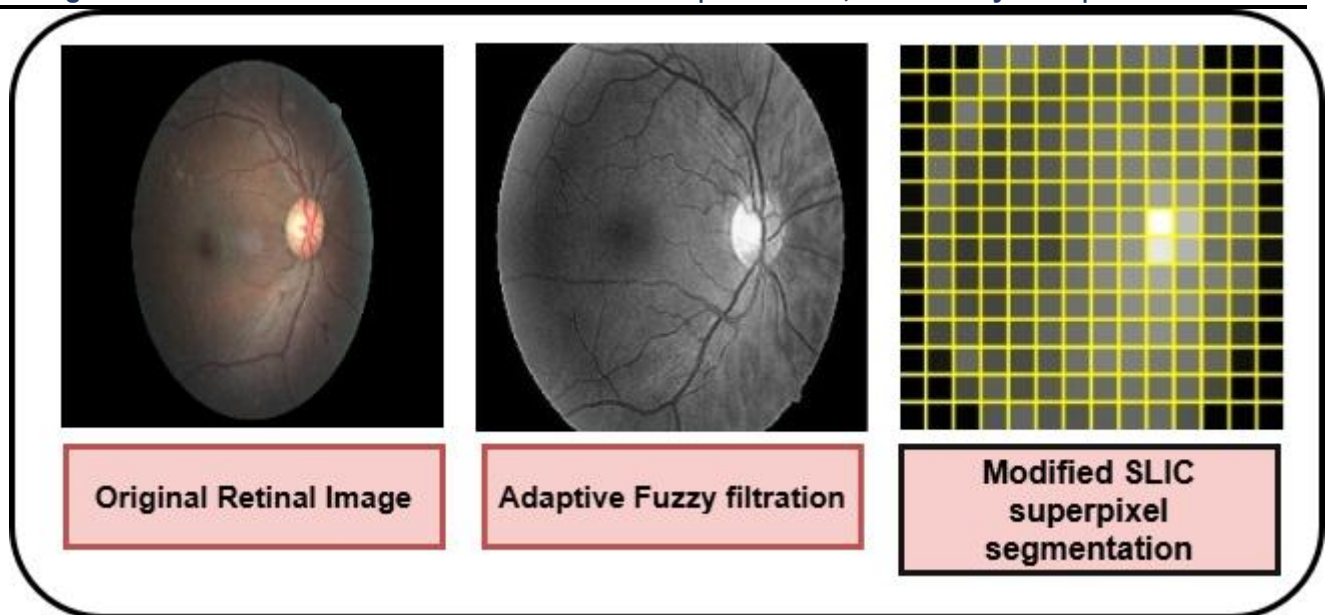


Figure 2. Image Evolution: (a) original retinal image (b) adaptive fuzzy filtered image, (c) modified SLIC superpixel segmentation.

3.4 ResNet-50

Learning networks are highly effective, but they frequently encounter challenges such as saturation, reduced accuracy, performance degradation, and issues like vanishing or exploding gradients. ResNet-50 (Ray et al., 2018) addresses these problems through the use of residual learning and identity mapping. By incorporating shortcut connections, it significantly minimizes the vanishing gradient issue and reduces the risk of overfitting. As illustrated in the fig 4, the mapping function is expressed as:

$$W(x) = F(x) + x \quad (3.4.1)$$

where the network learns the residual $F(x)$ and adds it to the original input x . ResNet architectures stack different stages across versions to produce a 2048-dimensional 1D feature vector per image.

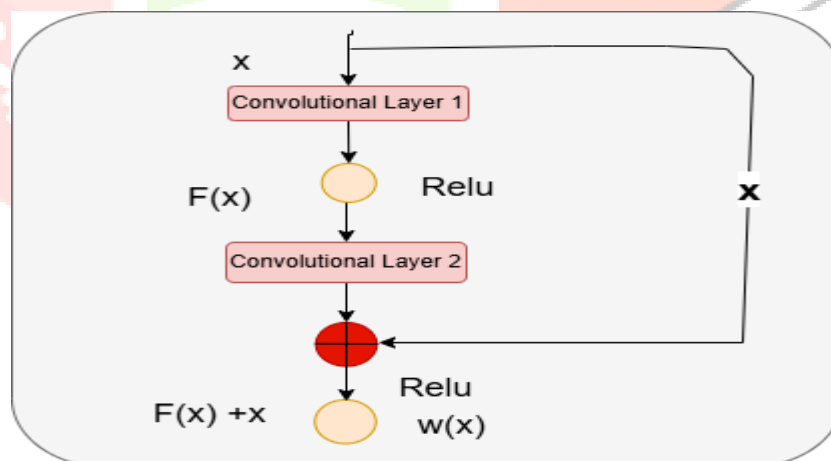


Figure 3: ResNet Architecture

3.5 Feature Statistical Analysis

Deeper analysis using statistical methods improves fundus diabetic retinopathy classification accuracy with less and relevant features. Features extracted from ResNet-50 are assessed using metrics such as mean, variance, skewness, kurtosis, Pearson Correlation Coefficient (PCC), Dice coefficient, and Canonical Correlation Analysis (CCA). These measures help evaluate how effectively the features capture the unique patterns of the retinal fundus associated with hemorrhages, microaneurysms, soft exudates, and hard exudates. The statistical outcomes for ResNet-50 in these scenarios are presented in Table 2.

Table 2: Statistical metrics evaluation for extracted features using ResNet-50

Parameters	Classes			
	Hemorrhages	Microaneurysms	Soft exudates	Hard exudates
Mean	0.3348	0.3409	0.3518	0.4537
Variance	0.3324	0.3489	0.3847	0.4405
Skewness	5.6525	6.1248	3.7689	4.4668
Kurtosis	48.3895	53.2777	31.1486	39.4781
PCC	0.5179	0.4994	0.5158	0.9168
Dice Coefficient	0.9125	0.9098	0.8943	0.8892
CCA	0.8811			

3.6 Selective feature pooling layer: Particle Swarm optimization (PSO)

For improved generalizability and classifier performance the selective feature pooling layer compresses the features of the retinal image into small vectors. In the diagnosis of diabetic retinopathy, these techniques increase precision by using bioinspired optimization algorithms like PSO (Liu et al, 2011). Particle Swarm Optimization (PSO) was first proposed by Kennedy and Eberhart in 1995, modeling it after the behavior of bird flocks to solve optimization problems. The process starts by initializing particles and setting key parameters. How PSO updates feature selection width, particle position, and velocity is shown below:

- **Initialization:** Randomly set each particle's position and velocity. Set each particle's best position (***pbest***), the overall best position (***gbest***), and the acceleration coefficient (c_1 , c_2).
- **Iteration Loop:** For each iteration:
 - Calculate the inertia weight(w_i).
 - Revise the velocity of each particle based on ***pbest*** and ***gbest***.
 - Revise the position of each particle utilizing the updated velocity.
 - When the new position of a particle is better, it is necessary to update its ***pbest***.
 - Update ***gbest*** if and only if the new ***pbest*** is better than the previous ***gbest***.
- **Output:** The ideal solution should be output as ***gbest*** after the maximum number of iterations has been reached.

Parameters in this study are adjusted experimentally where random values both r_1, r_2 are set to 0.85 for iterations from 100 to 1000: inertia weight (w_i) between 0.45 and 0.9, and components like cognitive (c_1) and social (c_2) between 1.0 and 2.0.

3.7 Statistical Analysis based on Fuzzy Entropy

Entropy is commonly used for feature selection in retinopathy. It measures how much information each feature provides in predicting outcomes. In this study, features from all abnormal cases are evaluated using Fuzzy Entropy to assess their importance. Fuzzy entropy is a valuable method for detecting diabetic retinopathy as it measures the uncertainty or complexity in retinal images. By evaluating the variations in pixel intensity and texture, fuzzy entropy helps identify subtle abnormalities such as microaneurysms or hemorrhages. This makes it effective in distinguishing between healthy and diseased retinal tissues, improving the accuracy of early detection and aiding in timely treatment of diabetic retinopathy.

Table 3. Fuzzy Entropy based statistical measure for PSO.

Measure	Particle Swarm Optimization			
	Hemorrhages	Microaneurysms	Soft exudates	Hard exudates
Fuzzy Entropy	0.6812	0.6235	0.5864	0.4231

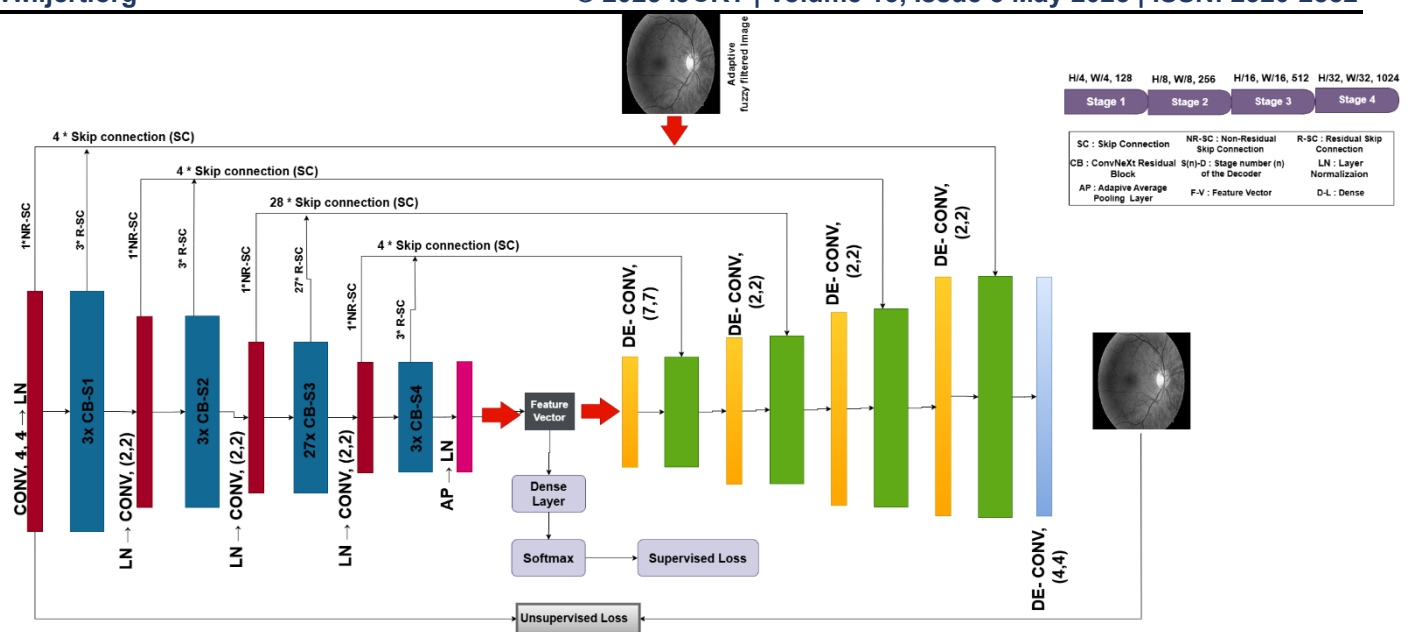


Figure 4. Deep Autoencoder with ConvNeXt-Base and Skip-Connected Decoder for Training.

Table 3 shows the feature selection techniques PSO based on fuzzy entropy-based statistical measure. Fuzzy Entropy, which reflects the complexity and uncertainty within a fuzzy system, is higher for hemorrhages under PSO (0.6812), suggesting greater ambiguity in feature relationships. Similarly, the entropy value for microaneurysms (0.6235) under PSO points to a more structured dataset with reduced variability in feature representation. For soft and hard exudates, PSO yields a lower entropy value (0.5864 for soft and 0.4231 for hard exudates), indicating more clearly defined and less uncertain feature interactions.

3.8 ConvNeXt Layer

This section explains the classification approach used in our research. Figure 4, shows an overview of the proposed ConvNeXt architecture. The ConvNeXt encoder-decoder process is discussed in subsection.

3.8.1 ConvNeXt Encoder

ConvNeXt, (Benchallal et al, 2024), is a type of neural network made entirely of convolutional layers. It was developed by updating traditional ResNet models to include ideas from transformer models, while still using only convolutional parts. The update process involved several steps:

1. The way computation is spread across different stages was changed by adjusting how many blocks are used in each stage. Each stage handles feature maps of different resolutions.
2. The starting layer of ResNet was replaced with a 4×4 convolution layer called the patchify stem, inspired by how transformers handle image patches.
3. Depth-wise convolution was added, which works on each channel separately, mixing only spatial information, shown in Fig. 5. Combined with 1×1 convolutions, this separates spatial and channel mixing similar to how transformers work.
4. A design element from transformers, the inverted bottleneck (where the middle layer is wider than the input and output), was added.
5. A large 7×7 depth-wise convolution was placed before the 1×1 convolution, mimicking the order of layers in vision transformers.
6. The ReLU activation function was replaced with GELU, and batch normalization was replaced with layer normalization. Only one activation and one normalization layer are used per block, just like in transformers.
7. Downsampling layers were separated from the main blocks, with layer normalization added before them. The patchify stem also has layer normalization after it.

ConvNeXt now matches vision transformers in both scalability and performance. In our autoencoder model, we used the ConvNeXt-Base version as the encoder. It uses a compute ratio of (3, 3, 27, 3) across its four stages. As the encoder goes deeper, the number of channels increases, leading to more computation in the deeper blocks, as shown in Fig 6.

3.8.2 ConvNeXt Decoder

The decoder in our autoencoder also uses a multi-stage design, allowing it to use feature maps from all encoder stages. It includes 40 skip-connections: 36 from residual blocks and 4 from downsampling layers, including the patchify stem, as shown in Fig 7. These connections help capture detailed information and maintain symmetry between encoder and decoder. Each decoder stage starts with an upsampling layer to support reconstruction and integrate skip-connection data effectively (Yang et al, 2024).

Figure 8, shows each decoder stage includes deconvolution blocks, matching the number of skip-connections from the encoder. Each block receives two inputs: one from a skip-connection and one from either a previous block or an upsampling layer. These inputs are combined by adding them together, then passed through a 1×1 deconvolution layer, followed by layer normalization and an activation function. This setup mixes encoder and decoder information without changing the feature dimensions.

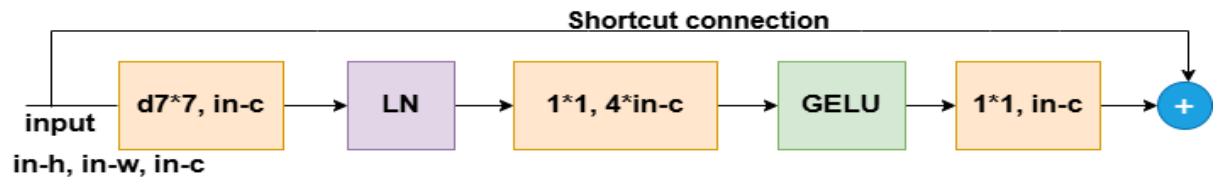


Figure 5. ConvNeXt Residual Block Architecture with Depth-Wise and 1×1 Convolutions for Enhanced Feature Representation.

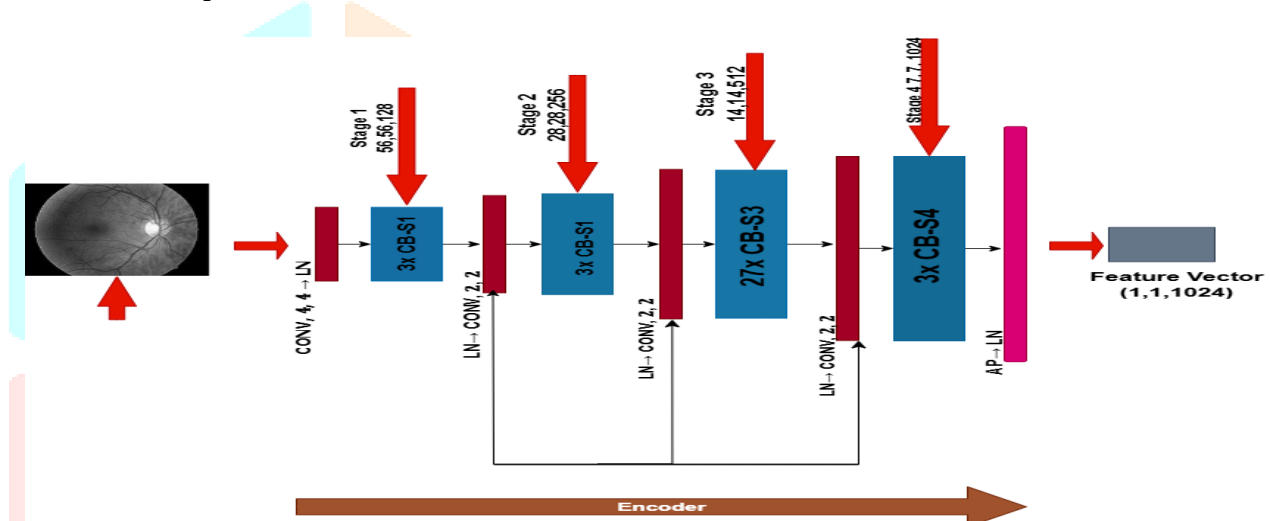


Figure 6. ConvNeXt Encoder with Stage-Wise Residual Block Distribution and Skip Connections for Decoder Integration.

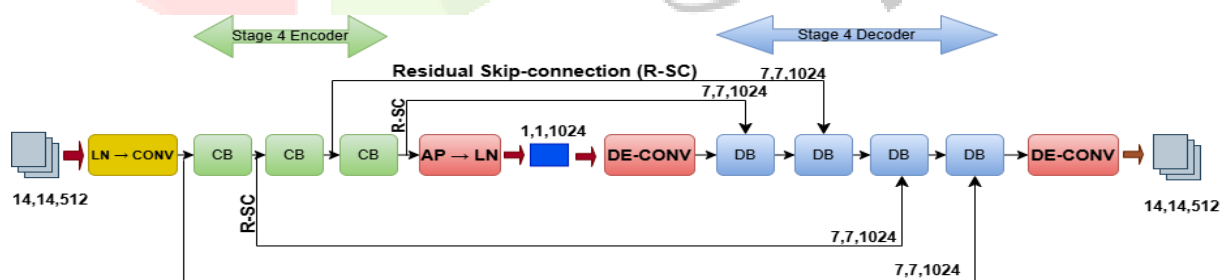


Figure 7. Skip-Connection Architecture Between Encoder and Decoder Stages with Normalized Downsampling and Upsampling Layers.

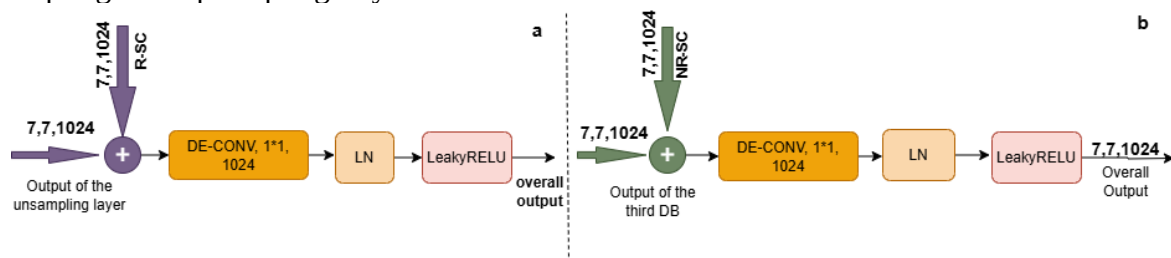


Figure 8. In-Depth View of Deconvolution Blocks (a) and (b) in Stage 4 of the Decoder.

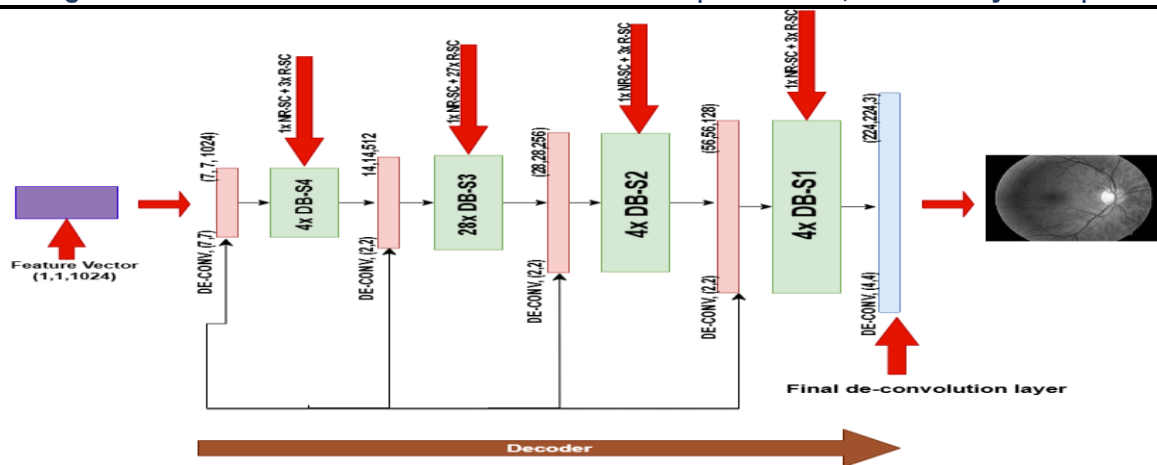


Figure 9. Multi-Stage Decoder Network with Symmetric Autoencoder Design and Integrated Skip Connections.

In the decoder, as shown in Fig 9, we used LeakyReLU for stages 4, 3, and 2 to avoid issues with ReLU's zero output. Stage 1 uses ELU to help with vanishing gradients and faster learning. At the end, a 4×4 deconvolution layer with 3 kernels and stride 4 is added, followed by a sigmoid function. The sigmoid is used because input image pixel values are scaled between 0 and 1 by dividing each channel by 255.

In the semi-supervised setup, we added a Dense Layer with units matching the number of target classes. It takes the ConvNeXt encoder's output and applies a SoftMax function to predict class labels. During training, we added Gaussian noise to input images. This helps the model learn to reconstruct from noisy data, based on the idea that small changes to inputs shouldn't affect the predicted labels.

III. EXPERIMENTAL RESULTS

A confusion matrix is used in this study to evaluate transfer learning models. During the evaluation process, the models are trained with 90% of the input features, while the remaining 10% are reserved for testing. The confusion matrix components are as follows: True Positive (TP), False Positive (FP), True Negative (TN), False Negative (FN). The proposed model performance is evaluated by using several statistical metrics, including Accuracy, Error Rate, F1 Score, Matthews Correlation Coefficient (MCC), and Jaccard Index. The AUC is usually determined by summing the areas of consecutive trapezoids formed beneath the ROC curve. The mathematical definitions of these metrics are provided in Table 4.

Table 4: Metrics mathematical definitions.

Metrics	Equation
Accuracy (%)	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100$
Error Rate (%)	$Err = \frac{FP + FN}{TP + TN + FP + FN} \times 100$
F1 Score (%)	$F1 = \frac{2TP}{2TP + FP + FN} \times 100$
MCC	$MCC = \frac{TN \times TP - FN \times FP}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$
Jaccard Index (%)	$Jaccard = \frac{TP}{TP + FP + FN} \times 100$

4.1 Performance Analysis

Table 5 shows the results of the study's evaluation of the ConvNeT classifiers' performance for retinal image classification across a range of K values of 10. First, using just a ConvNet model and a ResNet-X based feature fusion approach, we were able to attain a 94.610% accuracy rate at K=10 in the absence of segmentation. For the second case, we used segmentation and a combination of ResNet-50 based feature extraction and the ConvNet model to attain an accuracy of 96.920% at K=10. The third scenario involved segmentation and feature selection. Using PSO for feature selection led to feature extraction based on ResNet-50. The convNet achieved an accuracy of 98.250% at K=10.

Table 5: Performance analysis of the ConvNet Classifiers for all three cases and k=10.

Cases	Metrics				
	Accuracy	Error Rate	F1 score	MCC	Jaccard Index
Without Segmentation and FS	94.610	5.370	94.367	0.701	0.700
With modified SLIC only	96.920	3.910	96.906	0.980	0.980
With modified SLIC and FS along PSO	98.250	2.120	98.240	0.544	0.544

4.2 Complementary experiments for DR detection

Table 6 displays the detailed performance results of all classes evaluated on ConvNext, including modified SLIC, FS, and PSO, along with their collective performance metrics. The performance measurements for the detection of individual classes and their aggregated performance.

Table 6: Performance of all classes on the test data

Classes	Accuracy	F1-score	Jaccard Index	AUC
Hemorrhages	95.711	93.049	0.532	93.369
Microaneurysms	94.606	89.631	0.685	97.775
Soft exudates	96.360	95.047	0.659	95.542
Hard exudates	98.942	95.594	0.432	99.559
Aggregated	98.250	98.240	0.544	99.548

4.3 Comparison with Literature Study

The proposed methodology has been comprehensively compared with many conventional methods from the literature, recognizing the existence of diverse retinal illnesses and the adaptation of different methodologies to particular conditions. The chosen approaches were assessed according to essential criteria, including the incorporation of several retinal disease categories and the application of deep learning, namely transfer learning. The proposed approach exhibited superior results across these benchmarks. The comparative outcomes are illustrated in the Fig 10.

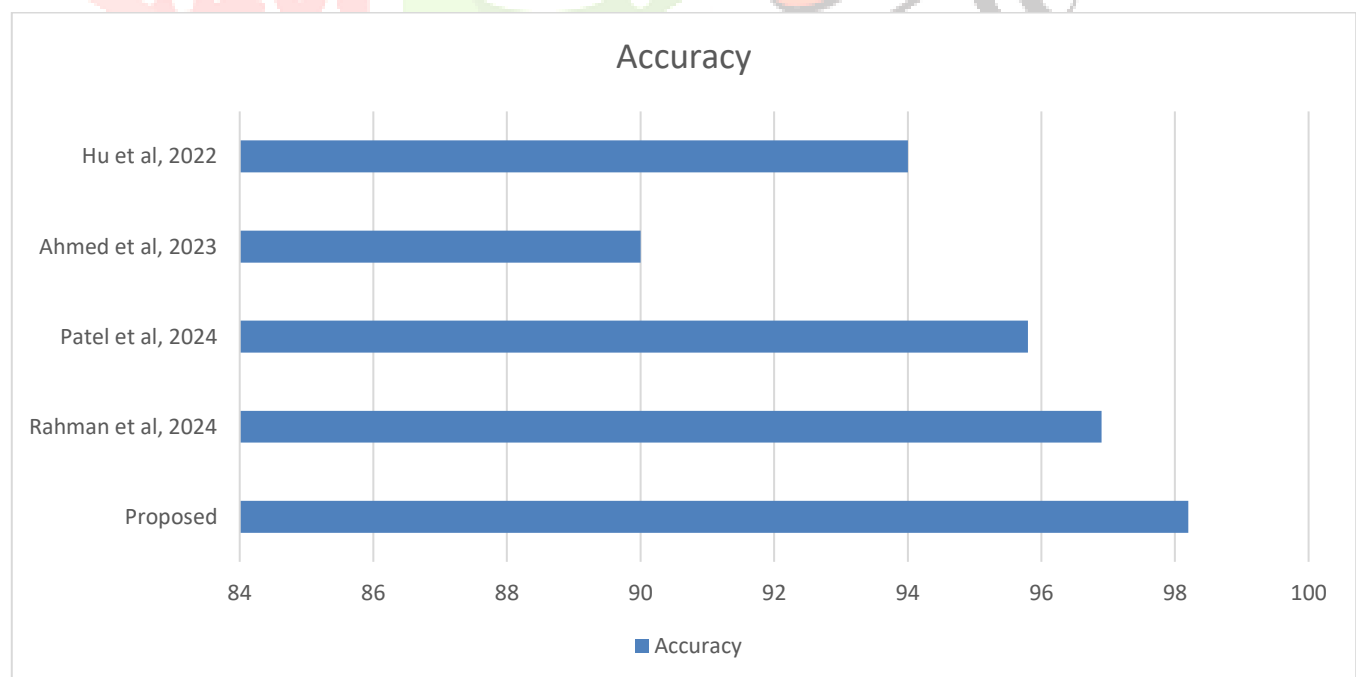


Figure 10. Comparison with Literature studies.

IV. CONCLUSION

In this paper, we presented a supervised learning approach aimed at developing deep learning models that are both robust and generalizable for accurate DR detection identification from retinal fundus images in real-world conditions. Our method addresses the common challenge of accurate retinal detection using fundus image by effectively leveraging labelled samples during training. To achieve this, we designed a novel deep learning architecture featuring an adaptive fuzzy filter, followed by segmentation using the modified SLIC algorithm. The ConvNeXt-Base as the encoder, which is a modern and efficient backbone, paired with a carefully constructed decoder. This architecture enables the successful integration of consistency regularization a key technique in supervised learning allowing the Resnet-50 model to learn relevant features even from image data.

Extensive experiments demonstrated that our method maintains strong and consistent classification performance, even when trained on a small amount of labelled data. In contrast, traditional supervised models showed a significant drop in performance when faced without ResNet50 and PSO. Our results also outperformed other comparable supervised learning approaches, highlighting the strength and reliability of the proposed method.

Looking ahead, we aim to further enhance the generalization capabilities of our proposed model and explore the adaptation of our approach to other challenging tasks in computer vision. By doing so, we hope to contribute to more efficient and scalable deep learning solutions in environments where labelled data is scarce or difficult to obtain.

REFERENCES

- [1] Wu, M., Chen, Q., He, X., Li, P., Fan, W., Yuan, S., & Park, H. (2017). Automatic subretinal fluid segmentation of retinal SD-OCT images with neurosensory retinal detachment guided by enface fundus imaging. *IEEE Transactions on Biomedical Engineering*, 65(1), 87-95.
- [2] Basu, Keenan, T. D., Chakravarthy, U., Loewenstein, A., Chew, E. Y., & Schmidt-Erfurth, U. (2021). Automated quantitative assessment of retinal fluid volumes as important biomarkers in neovascular age-related macular degeneration. *American journal of ophthalmology*, 224, 267-281.
- [3] De Fauw, J., Ledsam, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., ... & Ronneberger, O. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature medicine*, 24(9), 1342-1350.
- [4] Schmidt-Erfurth, U., Riedl, S., Michl, M., & Bogunović, H. (2020). Artificial Intelligence in retinal vascular imaging. *Retinal Vascular Disease*, 133-145.
- [5] Li, F., Chen, H., Liu, Z., Zhang, X. D., Jiang, M. S., Wu, Z. Z., & Zhou, K. Q. (2019). Deep learning-based automated detection of retinal diseases using optical coherence tomography images. *Biomedical optics express*, 10(12), 6204-6226.
- [6] Fang, L., Wang, C., Li, S., Rabbani, H., Chen, X., & Liu, Z. (2019). Attention to lesion: Lesion-aware convolutional neural network for retinal optical coherence tomography image classification. *IEEE transactions on medical imaging*, 38(8), 1959-1970.
- [7] Rahimy, E. (2018). Deep learning applications in ophthalmology. *Current opinion in ophthalmology*, 29(3), 254-260.
- [8] Son, J., Shin, J. Y., Kim, H. D., Jung, K. H., Park, K. H., & Park, S. J. (2020). Development and validation of deep learning models for screening multiple abnormal findings in retinal fundus images. *Ophthalmology*, 127(1), 85-94.
- [9] Chalakkal, R. J., Abdulla, W. H., & Thulaseedharan, S. S. (2019). Quality and content analysis of fundus images using deep learning. *Computers in biology and medicine*, 108, 317-331.
- [10] Serikbaeva, A., Li, Y., Ma, S., Yi, D., & Kazlauskas, A. (2024). Resilience to diabetic retinopathy. *Progress in retinal and eye research*, 101271.
- [11] Ahmed, M. S., Rahman, A., Alhabboub, Y., Alzahrani, K., Baragbah, H., Altaha, B., ... & Bakry, A. (2025). Diagnosing Retinal Eye Diseases: A Novel Transfer Learning Approach.
- [12] Rebolleda, G., Noval, S., Contreras, I., Arnalich-Montiel, F., García-Perez, J. L., & Muñoz-Negrete, F. J. (2009). Optic disc cupping after optic neuritis evaluated with optic coherence tomography. *Eye*, 23(4), 890-894.
- [16] Hu, J., Chen, Y., & Yi, Z. (2019). Automated segmentation of macular edema in OCT using deep neural networks. *Medical image analysis*, 55, 216-227.

- [17] Feng, S., Zhao, H., Shi, F., Cheng, X., Wang, M., Ma, Y., ... & Chen, X. (2020). CPFNet: Context pyramid fusion network for medical image segmentation. *IEEE transactions on medical imaging*, 39(10), 3008-3018.
- [15] Alfian, G., Syafrudin, M., Fitriyani, N. L., Anshari, M., Stasa, P., Svub, J., & Rhee, J. (2020). Deep neural network for predicting diabetic retinopathy from risk factors. *Mathematics*, 8(9), 1620.
- [] Liu, X., Wang, S., Zhang, Y., Liu, D., & Hu, W. (2021). Automatic fluid segmentation in retinal optical coherence tomography images using attention based deep learning. *Neurocomputing*, 452, 576-591.
- [18] Xing, G., Chen, L., Wang, H., Zhang, J., Sun, D., Xu, F., ... & Xu, X. (2022). Multi-scale pathological fluid segmentation in OCT with a novel curvature loss in convolutional neural network. *IEEE Transactions on Medical Imaging*, 41(6), 1547-1559.
- [13] Ho, E., Wang, E., Youn, S., Sivajohan, A., Lane, K., Chun, J., & Hutnik, C. M. (2022). Deep ensemble learning for retinal image classification. *Translational Vision Science & Technology*, 11(10), 39-39.
- [14] Shitole, A., Kenchappagol, A., Jangle, R., Shinde, Y., & Chadha, A. S. Enhancing Retinal Scan Classification: A Comparative Study of Transfer Learning and Ensemble Techniques.
- [22] Ahmed, M. I. B. (2023). Early Detection of Diabetic Retinopathy Utilizing Advanced Fuzzy Logic Techniques. *Mathematical Modelling of Engineering Problems*, 10(6).
- [19] Patel, R. K., Chouhan, S. S., Lamkuche, H. S., & Pranjali, P. (2024). Glaucoma diagnosis from fundus images using modified Gauss-Kuzmin-distribution-based Gabor features in 2D-FAWT. *Computers and Electrical Engineering*, 119, 109538.
- [20] Rahman, A., Youldash, M., Alshammari, G., Sebiany, A., Alzayat, J., Alsayed, M., ... & Aljishi, N. (2024). Diabetic Retinopathy Detection: A Hybrid Intelligent Approach. *Computers, Materials & Continua*, 80(3).
- [21] Orlova, E. V. (2024). Artificial Intelligence-Based System for Retinal Disease Diagnosis. *Algorithms*, 17(7), 315.
- [23] Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., & Meriaudeau, F. (2018). Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research. *Data*, 3(3), 25.
- [24] Van De Ville, D., Nachtegaal, M., Van der Weken, D., Kerre, E. E., Philips, W., & Lemahieu, I. (2003). Noise reduction by fuzzy image filtering. *IEEE transactions on fuzzy systems*, 11(4), 429-436.
- [25] Kushwaha, S. (2024). An effective adaptive fuzzy filter for speckle noise reduction. *Multimedia Tools and Applications*, 83(12), 34963-34978.
- [26] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11), 2274-2282.
- [27] Csillik, O. (2017). Fast segmentation and classification of very high resolution remote sensing data using SLIC superpixels. *Remote Sensing*, 9(3), 243.
- [28] Ray, S. (2018). Disease classification within dermoscopic images using features extracted by resnet50 and classification through deep forest. *arXiv preprint arXiv:1807.05711*.
- [29] Liu, Y., Wang, G., Chen, H., Dong, H., Zhu, X., & Wang, S. (2011). An improved particle swarm optimization for feature selection. *Journal of Bionic Engineering*, 8(2), 191-200.
- [30] Benchallal, F., Hafiane, A., Ragot, N., & Canals, R. (2024). ConvNeXt based semi-supervised approach with consistency regularization for weeds classification. *Expert Systems with Applications*, 239, 122222.
- [31] Yang, Y., Yao, S., Xue, Y., Zhao, W., & Wu, C. (2024). A variable fidelity approach for predicting aerodynamic wall quantities of hypersonic vehicles using the ConvNeXt encoder-decoder framework. *Aerospace Science and Technology*, 155, 109605.