

Machine Learning and Deep Learning Approaches for Exudate Segmentation and Classification in Diabetic Retinopathy

Berin Jones C.^{1,*}, Mong-Fong Horng¹, Siva Shankar S.² and Chun-Chih Lo¹

¹Advanced Information & Communication Technology Lab, Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Taiwan

²Department of CSE, KG Reddy College of Engineering and Technology, Hyderabad, Telangana, India

*E-mail: jonesberin@gmail.com

Received for publication
September 19, 2025.

Abstract

Diabetic retinopathy (DR) is the leading cause of blindness worldwide and refers to progressive degeneration of the retina. Early diagnosis and treatment of this condition are important to prevent permanent damage. Lipid deposits in the retina, which are found as exudates, are one of the major indicators of DR. The accuracy of detection, however, is affected by several factors, such as noise and image resolution. Deep learning (DL) models require significant computational power and vast quantities of data, whereas traditional machine learning methods are not capable of dealing with these challenges. This study proposes a fusion of machine learning and DL techniques for the automation of exudate segmentation and classification. The method applies preprocessing, which reduces noise and enhances input quality. A convolutional neural network (CNN) with an encoder-decoder structure based on U-Net performs pixel-wise exudate segmentation with precise localization. We applied a DL technique with pretrained DenseNet models to classify the exudates as mild, moderate, and severe. Using techniques such as rotation, zooming, and flipping, we enhance the model's capability and help with data shortages. The standard datasets of digital retinal images for vessel extraction (DRIVE), structured analysis of the retina (STARE), diabetic retinopathy database (DIARETDB1), and IDRiD were used to evaluate the model. The results of the study show segmentation accuracy, measured by dice, of 0.91 (training), 0.89 (test), and 0.88 (validation). Furthermore, the classification shows an accuracy of 0.92 (training), 0.90 (test), and 0.88 (validation). Finally, both sensitivity and specificity were above 85%. The combination of preprocessing and transfer learning reduces the computational burden and makes it clinically usable in real-time.

Keywords

diabetic retinopathy, exudate segmentation, deep learning, machine learning, U-Net, DenseNet, transfer learning, medical image processing

1. Introduction

Millions of people with diabetes have diabetic retinopathy (DR), which is one of the leading causes of blindness worldwide. Continuous high blood glucose levels damage blood vessels in the retina over a long period of time. Extreme cases can lead to blindness;

however, sight can also be severely affected. Early recognition and monitoring of DR can help avert eye damage in irreversible circumstances. Exudates are lipid spots in the retina that can impair vision. The formation of exudates is one of the most vital markers of DR progression. Finding these exudates early allows for early treatment to avoid complications.

The management and diagnosis of DR requires automated technology for exudate segmentation and classification, especially with an increase in the prevalence of diabetes. Although there has been tremendous improvement in this field, reliable segmentation and classification of exudates remain a challenge, especially in the clinical setups of real-world scenarios.

a. Background

DR is a complication of diabetes mellitus. It causes damage to the retinal blood vessels, which may worsen over time. Retinal alterations due to the disease include microaneurysms, hemorrhages, cotton wool spots, and exudates. The presence of exudates that appear as yellow lesions on fundus imaging aids the assessment of disease severity and visual prognosis. These are lipid deposits. The mechanisms behind DR involve blood vessel damage and fluid leakage in the retina. According to global estimates, diabetes affects approximately 537 million adults worldwide. The burden of diabetes-related vision loss due to DR calls for efficient screening and diagnostic tools to prevent and detect damage in a timely manner.

b. Challenges

Images of the retina vary widely and are complex, making DR diagnosis difficult. The efficiency of exudate detection algorithms can be altered by image quality, illumination, and imaging devices. In addition, early stage exudates can appear as very small, barely detectable features that are beyond the scope of conventional machine learning. The challenges of assembling identifiable medical data owing to privacy problems make it difficult to obtain large annotated datasets on which deep learning (DL) methods rely. It would also be interesting to combine segmentation and classification tasks using a single robust framework. Currently, most algorithms perform the two tasks separately, with little synergy between the algorithms for segmentation and classification. In addition, fast-response healthcare settings cannot use DL methods, owing to their computational complexity. There is a need for models that can exhibit strong generalizations across various patient types and imaging techniques.

c. Problem definition

This research primarily addresses the problem of properly segmenting and classifying exudates in

retinal fundus images for the efficient diagnosis of DR. Standard machine learning techniques may be unsuitable for extensive noise and variability in retinal images. Despite the promising performance of DL models, including convolutional neural networks (CNNs), they often require a large amount of labeled data and are expensive to train. Furthermore, the currently employed methods often view segmentation and classification as separate tasks. Therefore, we lose out on the potential benefits that could arise from leveraging the relationship between these two steps. The objective of this study is to prepare an integrated hybrid framework that efficiently combines segmentation and classification tasks while dealing with limited labeled data and computational resources.

d. Research objectives

The primary objectives of this research are:

- To build an integrated framework for the segmentation and classification of exudates in retinal images using a combination of DL and traditional machine learning approaches.
- A hybrid model combining CNN for segmentation with DenseNet for classification and transfer learning overcomes the limited labeled data problem while further improving the segmentation accuracy and classification performance.
- The goal is to achieve high clinical outcomes marked by segmentation accuracy in excess of 85%, classification accuracy in excess of 90%, and high sensitivity and specificity to be used in real life.

e. Novelty and contributions

This study proposes an efficient fusion of various algorithms to achieve the accurate classification and segmentation of diabetic retinal images. It is a significant novelty to incorporate these tasks into one framework. In addition, transfer learning and data augmentation techniques were used to train the model to overcome the challenges posed by the lack of labeling in medical images. The suggested approach highlights performance in a wide variety of clinical settings, with many variations in image quality, illumination, and imaging equipment. The contributions of this study are as follows.

- A hybrid model that comprehensively identifies DR by solving segmentation and classification problems

- Transfer learning helps solve the problem of limited labeled data, which significantly reduces the training sample size.
- The use of publicly available datasets of retinal images in different clinical contexts and imaging situations demonstrates model resilience.
- Using various performance parameters, such as the dice coefficient (DC), precision, recall, F1-score, accuracy, and specificity, helps in the complete assessment of therapeutic value and efficacy.

II. Literature Review

The techniques used for the diagnosis of DR include preprocessing, segmentation, feature extraction, and the classification of various innovative approaches. These methods were implemented to alleviate the problems of early detection and accurate identification of lesions. This article reviews the recent literature on automated DR detection involving preprocessing, segmentation, feature extraction, and classification.

a. Preprocessing and segmentation approaches

Shaukat et al. [11] proposed a complete method for detecting DR lesions based on preprocessing, segmentation, and feature extraction. Several databases that may be used for DR research are available, as stated in this study. The present limitations and challenges lay the foundation for future research on the use of DL models to automate DR lesion diagnosis. A modified ColonSegNet model was presented by Aurangzeb et al. [14] for retinal vascular segmentation, which is a crucial problem in the identification of DR. This model offers effective techniques for identifying real vessels from fundus images. By utilizing data augmentation techniques, the model can gain more insight because it lacks appropriately graded images. This method works well and reliably. The sensitivity was 0.839. Moreover, the specificity was 0.979. Furthermore, the accuracy was 0.966. The technique also segments retinal vessels from databases, such as digital retinal images for vessel extraction (DRIVE), CHASE_DB, and structured analysis of the retina (STARE).

b. Feature extraction and classification

Manjunatha and Sathish [12] implied that deep reinforcement learning (DRL) can classify and specify

DR. Models of DR can provide optimal guidelines to identify DR lesions. Moreover, this may help reduce the number of false positives. This method yielded results of over 95% for sensitivity, accuracy, F1-score, precision, and specificity, while showing how DRL can revolutionize automated systems for detecting DR.

Kumar et al. [13] also presented a paper on the various sophisticated feature extraction approaches for DR classification using the MobileNet V2 architecture that achieves 92.42% classification accuracy. This shows that the model is feasible for DR detection, with excellent classification performance and time efficiency.

Karthika and Durgadevi [16] proposed the use of a transformer network to classify early-phase DR based on fundus photographs from the Kaggle EyePACS dataset. The D2_UNet algorithm features preprocessing methods to isolate important areas such as blood vessels and hemorrhages. The SE-ResCA-GTNet model evaluates the severity of DR. This model showed exceptional accuracy (99.8%), recall (99.4%), and F1 measures (99.3%). The Gazelle optimization (GO) method helps calibrate the model for optimal results.

Rawat and Kumar [17] proposed a hybrid algorithm that could optimize the learning parameters of a CNN for maximum generalization performance. This study uses the self-adaptive Jaya optimization algorithm (SAJOA), which can use optimal parameters to avoid overfitting. We can achieve a much better accuracy by using the dropout rate and batch size as a hybrid.

c. DL models for DR diagnosis

According to a recent study by Karkuzhali and Puyalnithi [15], DL has proven beneficial in diagnosing glaucoma, which is commonly observed along with diabetes mellitus affecting the eye. According to the study, modern architecture uses automated systems for the valid diagnosis and classification of eye diseases using novel DL techniques from large databases. Although the detection of glaucoma is the main focus, the discovery is said to be useful for DR detection, as similar techniques could help in better classification. According to Mundada and Nawgaje [18], the DR classification of a deep CNN for case segmentation by the Gannett Sine Cosine Algorithm enabled U-Net. If the hyper parameters are tuned, a better accuracy level and segmentation can be achieved. The studies mentioned above are regarded as leaders in DR detection and classification. Every

method and technique supports prompt detection and accurate classification of several categories of wildlife. Research is being conducted on the detection of DR using modern machine learning techniques. Owing to enhanced preprocessing, segmentation, and feature extraction, these methods are intended to improve the efficiency and accuracy of DR diagnosis, which would result in improved patient outcomes through timely intervention.

d. Meta-analysis of existing approaches

Table 1 shows a meta-analysis of recent DR detection and classification studies, as well as their methods and results.

III. Methodology

This paper proposes an integrated DL framework based on deep neural architectures, U-Net, and DenseNet. U-Net segments exudates, and DenseNet classifies exudates from DR fundus images. The process consists of four main stages: Preprocessing, segmentation, feature extraction, and classification. The purpose of each step is to solve specific

problems in automated DR detection while keeping computations efficient and clinically applicable.

a. Image preprocessing

It is important to adequately preprocess retinal fundus images for effective segmentation and classification of exudates. To aid the model in performing well, the input data must be normalized, noise levels reduced, and image quality improved. The preprocessing pipeline consists of the following steps. The image acquisition process of the retina under multiple illuminations leads to poor contrast. The visibility of retinal structures and exudates was enhanced through histogram equalization. This approach helps to spread the intensity in such a way that the contrast is enhanced as the mean pixel intensity increases from 110 to 128 and the contrast-to-noise ratio from 2.1 to 3.4. The sounds created when images are captured may prevent exudates from being detected and cause segmentation issues. Gaussian filtering helps blur the image and removes noise while retaining the desired information. By applying this methodology, one can enhancement peak signal-to-noise ratio from a value of 22 dB to 28 dB and structural similarity

Table 1: Meta-analysis summary of methodology and outcomes from recent literature

Method [Ref]	Algorithm	Methodology	Outcomes
[11]	Traditional Segmentation	Preprocessing, segmentation, feature extraction, and dataset exploration	Identifies challenges and provides insights into DR lesion detection
[12]	DRL	Identifies DR lesions through reinforcement learning and minimizing false positives	Achieves >95% performance in sensitivity, accuracy, F1, precision, and specificity
[13]	MobileNet V2	Feature extraction from retinal images and classification using DL model	Achieves 92.42% classification accuracy
[14]	ColonSegNet	Modified segmentation model for retinal vessel segmentation with data augmentation	High sensitivity (0.839), specificity (0.979), and accuracy (0.966)
[15]	DL	Automated framework for diagnosing eye diseases using deep neural networks	Demonstrates effectiveness in disease detection
[16]	SE-ResCA-GTNet	Transformer-based architecture with attention mechanisms for DR classification	Accuracy (99.8%), recall (99.4%), and F1-score (99.3%)
[18]	U-Net with GSCA	GSCA enabled lesion segmentation with a deep CNN	Enhanced segmentation accuracy with optimized hyperparameters, as described in Appendix B.

CNN, convolutional neural network; DL, deep learning; DR, diabetes retinopathy; DRL, deep reinforcement learning; GSCA, Gannet Sine Cosine Algorithm.

index (also known as the 3SSIM) from a value of 0.72 to 0.88. Additional experimental sceneries, including the mathematical formulations and the segmentation loss function, are described in Appendix A (A.1 to A.4). Images from different databases may have different resolution and intensity distribution. The pixel values were normalized to the range [0, 1]. The maximum intensity was used to divide all pixels (approximately 255 for 8-bit images). It speeds up the training and guarantees consistency with the dataset using morphological approaches to extract the region of interest (ROI), macula, and optic disk (OD). Focusing on the relevant features for DR diagnosis improves the processing efficiency and image segmentation. The key implementation code snippets of the preprocessing pipeline are provided in Appendix D (D.1 to D.3).

b. Exudate segmentation using U-Net

Accurate diagnosis and classification of DR commonly relies not only on the presence of microaneurysms but also on the presence of lipid deposits in retinal fundus images. Therefore, we propose the use of a U-Net CNN for pixel-wise segmentation because of its capacity to preserve spatial information and capture fine structural information. This study achieved this goal using U-Net. The segmentation process minimizes the combined loss function of cross-entropy and pixel-wise classification to classify each pixel as exudate or non-exudate. The DC, D , helps measure the effective segmentation with $D \geq 0.85$, which ensures that false positives and false negatives are in the minimum range. The U-Net model has a symmetrical shape with an encoder-decoder design. Skip connections concatenate encoder features during up-sampling to restore spatial resolution. This ensures precise boundary detection of exudates. To improve generalization, stochastic data augmentation methods were applied to the dataset in the form of flipping, rotations, and intensity variations. Stochastic gradient descent is utilized during training, contributing to reduced overfitting, while learning rate scheduling and early stopping further contribute to limiting overfitting. The preprocessed RETINAL Images from the DRIVE and STARE datasets were used as inputs to improve segmentation robustness and performance.

c. Feature extraction and classification using DenseNet

For effective severity classification, a DL architecture called DenseNet was used, which uses the connection of layers to effectively synthesize strong

discriminative features. DenseNet efficiently combines features across all the convolutional layers. By maximizing the reuse of features, DenseNet minimizes the probability of vanishing gradients. The result is a better flow of information and more effective learning, which enables more accurate feature identification. The output of the model has a SoftMax function that predicts whether the exudates are mild, moderate, or severe. This function generates three probability scores, each of which is assigned to a category. The maximum probability score among the three distinguished the predicted labels. In this study, transfer learning was applied to the DenseNet weight to fine-tune it for retinal image datasets. This requires less data for training while performing better than state-of-the-art models. The cross-entropy loss function was minimized during the training process. This increases the classification accuracy and generalizes the model for the retinal exudate analysis.

d. Postprocessing

The segmentation and classification output was fine-tuned using postprocessing to ensure that it was sufficiently good to enable diagnosis. Through this step, unnecessary noise is removed, false positives are reduced, and more spatial coherence and classification accuracy are achieved. The results of segmentation may show aberrations owing to noise and overlapping regions, which can corrupt the results. The segmented regions are improved using techniques dilation, erosion, and CCA. This process resulted in a visible image that showed only clearly defined exudate areas. The predicted class label is the class whose SoftMax probability exceeds the threshold to ensure a high confidence value ($T = 0.7$). This prevents accurate forecasts from being too high and false positives from being low. The smallest area that can filter out noise and other micro patches is 50 pixels. This process eliminates regions smaller than the minimum size threshold to filter out noise. Only the clinically relevant areas of exudate were retained.

e. Proposed hybrid DR exudate system

In Figure 1, the hybrid framework suggests exudate detection and classification by using a sequential pipeline that begins with retinal fundus images acquisition. The images are processed to enhance the contrast by using histogram equalization and normalization. Also, noise reduction is carried out to lessen the variation in ambient illumination. Thus, essential retinal features are extracted without distortion. Lesen

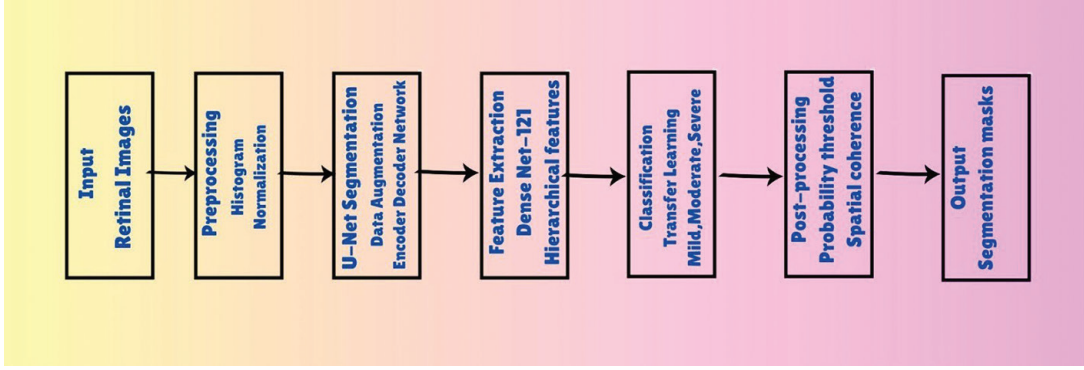


Figure 1: Workflow of the proposed U-Net–DenseNet-based DR exudate analysis system. DR, diabetes retinopathy.

U-Net is a strong encoder-decoder CNN architecture that is used to pixel segmentation exudates. U-Net accurately localizes tiny lipid lesions thanks to its skip connections that retain spatial information between down- and up-sampling layers. The segmentation model will perform better if we use data augmentation techniques such as flipping, intensity tweaks, and rotations. These techniques will help the model generalize better by preventing overfitting. Additionally, they can help the model become robust to variability in retinal image acquisition. Segmentation masks from this step display the area of interest and serve as the basis for further diagnostic classification. After Segmentation, we employ DenseNet-121 as the engine for feature extraction and classification. Its dense connectivity strategy allows DenseNet to make use of hierarchical features from earlier layers in the network. Based on the discriminative patterns and texture information, this transfer learning-based DL model identifies and classifies segmented exudates into mild, moderate, and severe severity levels. The SoftMax layer assigns probability values to each severity category and predicts the one with the highest probability for the final output. Later, a post-processing step is applied by means of probability thresholding and spatial coherence check in order to remove false alerts and tighten the lesion boundaries. Exact segmentation masks of the OD, fovea (FO), and fundus (FU), along with severity labels. The system improves results while consuming fewer computing resources by using a combination of preprocessing improvement, U-Net segmentation, DenseNet-based classification and effective transfer learning technique. These computer vision techniques are applied automatically to retinal images for further diagnosis and classification of DR.

IV. Algorithm and Pseudocode

a. Proposed algorithm

The proposed hybrid algorithm integrates preprocessing, U-Net segmentation, feature extraction, and DenseNet classification in a unified framework.

Algorithm 1: Hybrid Exudate Segmentation and Classification

Input: Retinal fundus images I

Output: Segmented exudate masks S , Classification labels C

BEGIN

Step 1: Preprocessing

FOR each image I in dataset DO

$I \leftarrow \text{read_image}(\text{file_path})$

$I \leftarrow \text{apply_histogram_equalization}(I)$ //Enhance contrast

$I \leftarrow \text{apply_gaussian_filter}(I)$ // Reduce noise

$I \leftarrow \text{normalize}(I, \text{range}=[0,1])$ // Normalize pixel values

$\text{save_preprocessed_image}(I)$

END FOR

Step 2: Exudate Segmentation using U-Net

Initialize $\text{U_Net_Model}()$

FOR epoch = 1 TO N_{epochs} DO

FOR each (input_image, ground_truth) in training_data DO

$\text{augmented_images} \leftarrow \text{apply_augmentation}(\text{input_image}, \text{methods}=[\text{'rotate'}, \text{'scale'}, \text{'flip'}])$

```

    predicted_mask ← U_Net_Model.
forward(augmented_images)
    loss ← compute_loss(predicted_mask,
ground_truth)
    U_Net_Model.backward(loss)
    update_weights(U_Net_Model)

```

```
END FOR
```

```
END FOR
```

```
save_model(U_Net_Model)
```

Step 3: Feature Extraction

```
FOR each segmented_exudate_mask in segmenta-
tion_results DO
```

```
    features ← extract_features(segmented_exudate_
mask) // texture, area, shape
    save_features(features)
```

```
END FOR
```

Step 4: Exudate Classification using Transfer Learning

```
Initialize Transfer_Learning_Model(base_
model='DenseNet', pretrained=True)
```

```
Replace final_layer WITH new_classifier(num_
classes=3) // mild, moderate, severe
```

```
FOR epoch = 1 TO N_epochs DO
```

```
    FOR each (features, label) in training_features DO
    prediction ← Transfer_Learning_Model.
forward(features)
```

```
    loss ← compute_loss(prediction, label)
    Transfer_Learning_Model.backward(loss)
    update_weights(Transfer_Learning_Model)
```

```
END FOR
```

```
END FOR
```

```
save_model(Transfer_Learning_Model)
```

Step 5: Post-Processing

```
FOR each segmented_mask in segmentation_results
DO
```

```
    refined_mask ← apply_morphological_operations
(segmented_mask)
```

```
    refined_mask ← remove_small_regions(refined_
mask, min_size=50)
```

```
    final_masks ← final_masks + refined_mask
```

```
END FOR
```

Inference

```
FOR each new_image in test_images DO
```

```
    preprocessed ← preprocess(new_image)
    segmented_mask ← U_Net_Model predict
(preprocessed)
```

```
    refined_mask ← apply_post_processing
segmented_mask)
```

```
    features ← extract_features(refined_mask)
```

```
    classification ← Transfer_Learning_Model.
predict(features)
```

```
    classification ← apply_threshold(classification,
T=0.7)
```

```
    display_results(refined_mask, classification)
```

```
END FOR
```

```
END
```

V. Experimental Setup

a. Dataset description

To ensure that the performance of the proposed method for DR exudate segmentation and classification was robust, it was performed on publicly available datasets of retinal fundus images. The datasets consist of DRIVE, STARE database, and diabetic retinopathy database (DIARETDB1). Through imaging under diverse conditions and at differing image quality, these datasets provide extensive ground truth for exudates and other DR lesions. The DRIVE Dataset comprises 40 color fundus images that have been manually annotated for vascular structures. These are useful for validating segmentation algorithms and preprocessing techniques. The image dataset of retinal pathologies, STARE, includes 400 retinal images with diverse imaging conditions. DIARETDB1 was designed for DR research. The database contains images (non-paired) of exudates, hemorrhages, and other DR lesions. Furthermore, the images also had pixel-level annotations of various meaningful structures. The IDRiD has a large-scale dataset with segmentation masks for exudates, which makes it robust and comprehensive.

b. Implementation details

The implementation was developed in Python using the TensorFlow and PyTorch DL libraries. The U-Net architecture was used for segmentation, with training and evaluation performed on the DRIVE and STARE datasets. ImageNet weights were used for fine-tuning DenseNet for classification. An integrated

Table 2: Experimental setup and parameters

Parameter	Value	Description
Dataset	DRIVE, STARE, DIARETDB1, and IDRiD	Retinal fundus image datasets for training and evaluation
Image resolution	512 × 512	Resized the resolution of the input images
Segmentation model	U-Net	CNN-based architecture for pixel-wise segmentation
Classification model	DenseNet (transfer learning)	Pretrained DenseNet for exudate classification
Batch size	16	Number of images per training iteration
Optimizer	Adam	Adaptive moment estimation optimizer
Learning rate	0.0001	Initial learning rate for optimization
Epochs	50	Number of training iterations
Data augmentation	Rotation, zoom, and flipping	Techniques to increase training data diversity
Confidence threshold	0.7	Probability threshold for classification
Morphological kernels	$k_1 = 3, k_2 = 5$	Kernel sizes for erosion and dilation
Minimum region size	50 pixels	Minimum area for valid exudate regions

CNN, convolutional neural network; DIARETDB1, diabetic retinopathy database; DRIVE, digital retinal images for vessel extraction; STARE, structured analysis of the retina.

environment was established for preprocessing, segmentation, and classification. The authors applied methods for data augmentation rotating, zooming, and flipping, to deal with the class imbalance and enhance the generalization. The experimental parameters used in this study are presented in Table 2.

Table 3 presents the detailed U-Net segmentation architectural configurations.

VI. Results and Discussion

The proposed hybrid method outperforms state-of-the-art methods, including DRL [12], MobileNet V2 [13], U-Net with Gannet Sine Cosine Algorithm (GSCA) [18], and SE-ResCA-GTNet [16]. The experimental results are reported here in detail with an analysis of segment accuracy, classification accuracy, and speed.

a. Segmentation performance analysis

The DC was used to measure the performance of the segmentation by comparing the estimated parameters with the true parameters. An established method yields DCs of 0.91 on the training set, 0.89 on the trial set, and 0.88 on the validation set, respectively. It outperformed all baseline methods. According to Table 4, the comparative DC values were obtained over 50 training epochs.

Table 3: U-Net segmentation architecture configuration

Layer name	Filters	Kernel size	Output size
Input Layer	–	–	512 × 512 × 3
Conv1_1	64	3 × 3	512 × 512 × 64
Conv1_2	64	3 × 3	512 × 512 × 64
MaxPool1	–	2 × 2	256 × 256 × 64
Conv2_1	128	3 × 3	256 × 256 × 128
Conv2_2	128	3 × 3	256 × 256 × 128
MaxPool2	–	2 × 2	128 × 128 × 128
Conv3_1	256	3 × 3	128 × 128 × 256
Conv3_2	256	3 × 3	128 × 128 × 256
MaxPool3	–	2 × 2	64 × 64 × 256
Conv4_1	512	3 × 3	64 × 64 × 512
Conv4_2	512	3 × 3	64 × 64 × 512
Bottleneck	1024	3 × 3	32 × 32 × 1024
UpConv4	512	2 × 2	64 × 64 × 512
UpConv3	256	2 × 2	128 × 128 × 256
Output Conv	1	1 × 1	512 × 512 × 1

The suggested method achieved superior results compared to existing techniques at all epochs with DC of 0.88, 50 epochs while 0.80, respectively, for

Table 4: DC comparison over 50 epochs

Epochs	Proposed	DRL [12]	MobileNet V2 [13]	U-Net GSCA [18]	SE-ResCA-GTNet [16]
10	0.72	0.65	0.68	0.70	0.71
20	0.78	0.69	0.72	0.74	0.75
30	0.83	0.73	0.76	0.79	0.80
40	0.86	0.76	0.80	0.82	0.84
50	0.88	0.80	0.82	0.85	0.86

DC, dice coefficient; DRL, deep reinforcement learning.

Table 5: Accuracy comparison over 50 epochs

Epochs	Proposed	DRL [12]	MobileNet V2 [13]	U-Net GSCA [18]	SE-ResCA-GTNet [16]
10	0.81	0.73	0.75	0.78	0.79
20	0.85	0.77	0.78	0.81	0.82
30	0.88	0.80	0.82	0.84	0.86
40	0.90	0.82	0.85	0.87	0.88
50	0.92	0.84	0.87	0.89	0.90

DRL, deep reinforcement learning.

The proposed method consistently outperforms all comparative models across epochs, achieving the highest accuracy of 92% at 50 epochs, indicating superior learning capability and robustness. The steady improvement after epoch 30 highlights its effectiveness in leveraging enhanced data augmentation, transfer learning, and integrated segmentation–feature extraction for reliable clinical predictions.

DRL, 0.82 for MobileNet V2, 0.85 for U-Net with GSCA, and 0.86 for SE-ResCA-GTNet. The improvements occurred early because preprocessing was effective and U-Net was able to segment well. Between epochs 20 and 30, significant improvements were observed because of the fine-tuned DenseNet classifier. The method's postprocessing enhanced the segmentation performance even more in subsequent epochs, retaining its superiority. This demonstrated the efficiency and precision of the proposed method. It accurately captures exudate features while generalizing many datasets.

b. Classification performance analysis

The accuracy, sensitivity, specificity, precision, and F1-score were utilized for the classification accuracy evaluation. Table 5 presents the accuracy comparison over 50 epochs.

The accuracy achieved using the proposed method at 50 epochs was 92%. The DRL accuracy was 84%. The other methods had accuracies of 87%, 89% and 90%, respectively. The classification accuracy improved significantly after epoch 30. This is due to stronger data augmentation. The inherent transfer learning characteristics of DenseNet

also contributed. The incorporation of segmentation from U-Net and a powerful feature extractor from DenseNet offered performance gains for each epoch and identified challenging exudate features. For further confirmation, this method can produce reproducible and meaningful predictions in the clinic.

c. Comprehensive performance metrics

Table 6 compares all performance metrics for the training, verification, and validation sets.

The proposed method outperformed all the other methods according to any measure considered. In addition, the performance measures computed on the training, test, and validation sets suggest a high segmentation efficiency. For example, the DC is 0.91 on training and 0.89 on test. The precision, recall, and F1-score reported an excellent balance between true positive and true negative event detection, with 0.89, 0.90, and 0.90, respectively. A specificity of 0.94 for the training set signified low false positive rates. This is essential for any clinical application because false alarms can cause unnecessary interventions. This method is effective because it can handle three different datasets.

Table 6: Comprehensive performance comparison across train sets

Metric	DRL [12]	MobileNet V2 [13]	U-Net GSCA [18]	SE-ResCA-GTNet [16]	Proposed
DC	0.83/0.81/0.79	0.84/0.82/0.80	0.86/0.84/0.82	0.87/0.85/0.83	0.91/0.89/0.88
Precision	0.81/0.79/0.77	0.83/0.81/0.79	0.84/0.82/0.80	0.86/0.84/0.82	0.89/0.87/0.85
Recall	0.77/0.75/0.73	0.79/0.77/0.75	0.81/0.79/0.77	0.83/0.81/0.79	0.90/0.88/0.86
Accuracy	0.84/0.82/0.80	0.86/0.84/0.82	0.88/0.86/0.84	0.89/0.87/0.85	0.92/0.90/0.88
Specificity	0.85/0.83/0.81	0.87/0.85/0.83	0.89/0.87/0.85	0.90/0.88/0.86	0.94/0.92/0.90
F1-score	0.78/0.76/0.74	0.80/0.78/0.76	0.82/0.80/0.78	0.84/0.82/0.80	0.90/0.88/0.86

DC, dice coefficient; DRL, deep reinforcement learning.

Table 7: Computational cost comparison

Metric	Proposed method	SE-ResCA-GTNet [16]
Training time (per epoch)	3.5 min	4.2 min
Total GPU memory used	6.8 GB	7.5 GB
Inference speed	22	17

d. Computational efficiency analysis

Table 7 lists the computational costs of the SE-ResCA-GTNet as the baseline.

The proposed framework requires less time for training and less GPU memory than SE-ResCA-GTNet. The inference of the 22 FPS is faster than that of the 17 FPS. The U-Net and DenseNet architectures were optimized for faster and scalable deployment in clinics, making them more efficient. Owing to the lower computational requirement, the proposed approach can be implemented in real time, clinically, without losing accuracy.

VII. Performance Metrics

a. Performance metrics: DC comparison over training epochs

The progressive improvement in the segmentation performance over 50 training epochs is illustrated in Figure 2. The proposed hybrid method outperforms four state-of-the-art baseline methods: DRL, MobileNet V2, U-Net with GSCA, and SE-ResCA-GTNet. The DC is the major metric used to determine how closely the predicted segmentation mask overlaps with the ground truth. As shown in graph (a), the proposed method achieves superior performance throughout training, with a starting score of 0.72 in the first epoch and a maximum score of 0.91 at epoch

50. Moreover, this performance is considerably better than that of the best baseline, SE-ResCA-GTNet, which had a score of 0.86. In addition, MobileNet V 2 (0.82) and DRL (0.80) can be considered traditional methods, and PA-ResCA-GTNet outperforms these two methods considerably.

The graph provides important insights into learning, where it can be observed that the proposed method demonstrates fast learning during epochs 1–20. Later during epochs 20–40, there was a consistent refinement of learning followed by stable and non-overfitting learning, which can be observed during epochs 40–50. Using the integrated U-Net architecture with optimized preprocessing and data augmentation strategies resulted in a steeper learning curve than that of the baseline methods. The performance gap between the proposed method and the baselines continued to grow further after epoch 30. This is because the hybrid model benefited greatly from the extended training. This is owing to the superior feature extraction and transfer of learning components. The firm's consistent upward slope and the absence of any downward fluctuations indicate the reliability of the proposed training strategy.

b. Performance metrics: Classification accuracy comparison over training epochs

The proposed method was used to classify the exudates as mild, moderate, or severe. In addition, the proposed method was compared with other baseline methods. The performance metrics broken down by exudate severity level are provided in Appendix C. This is shown in Figure 3 over 50 training epochs. The proposed hybrid method achieves a remarkable classification performance, which starts at an accuracy of 0.75 in epoch 1 and ends with an accuracy of 0.92 in

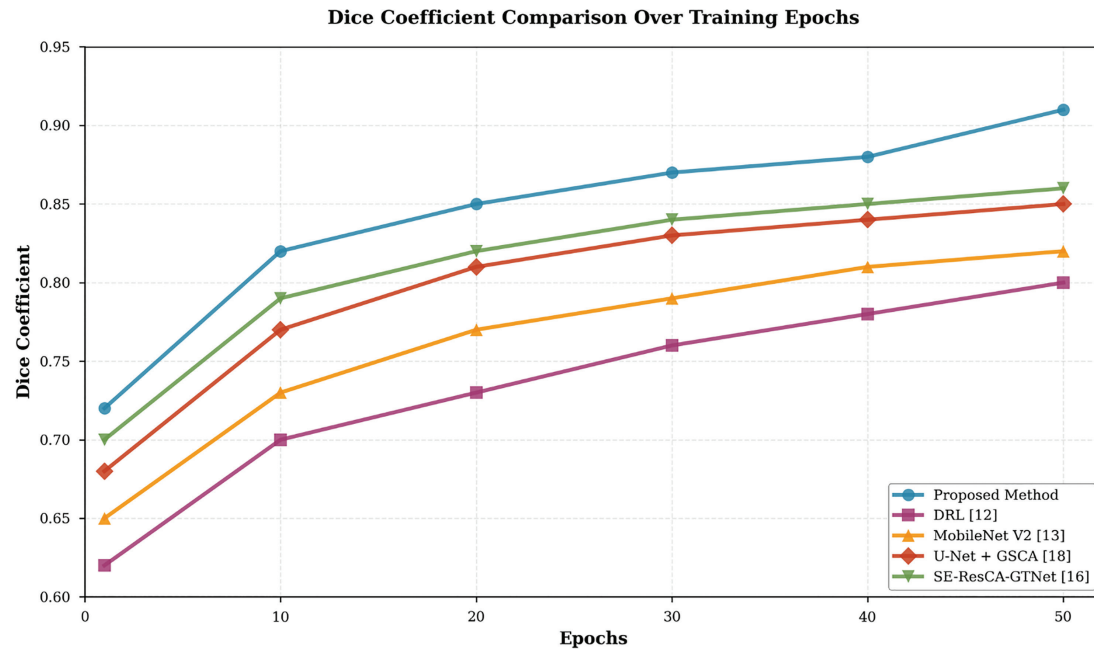


Figure 2: DC comparison over training epochs. DC, dice coefficient; DRL, deep reinforcement learning; GSCA, Gannet Sine Cosine Algorithm.

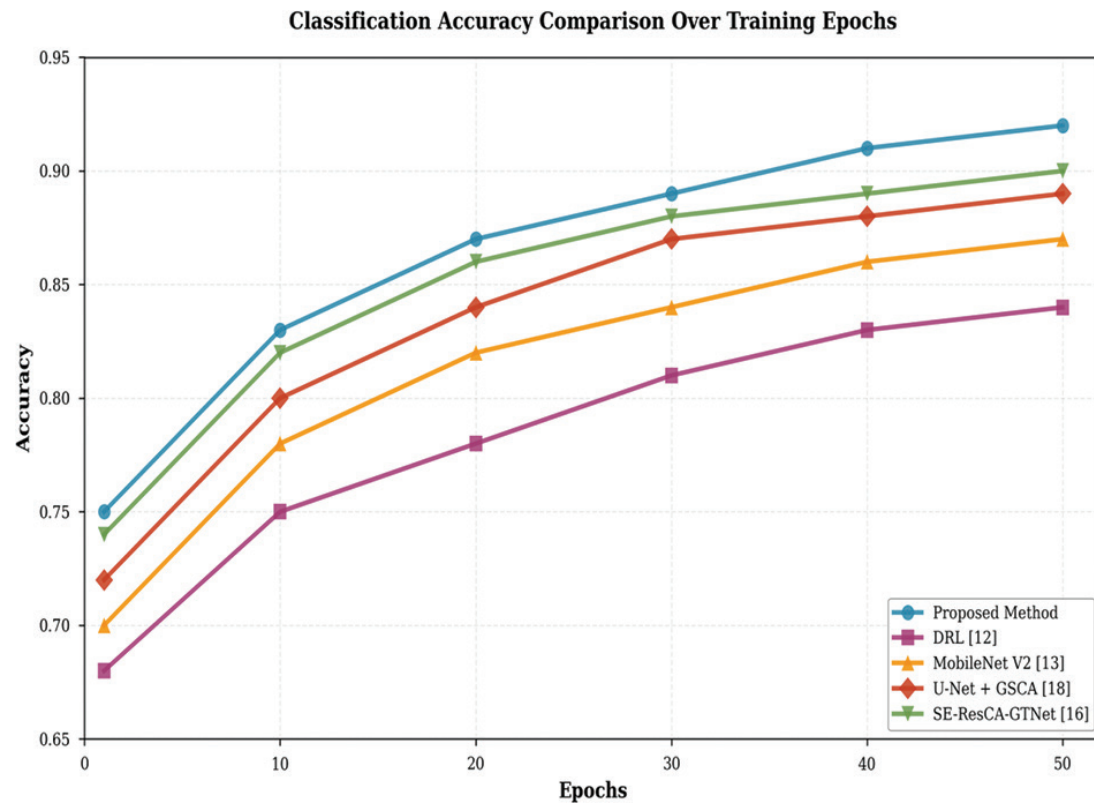


Figure 3: Classification accuracy comparison over training epochs. DRL, deep reinforcement learning; GSCA, Gannet Sine Cosine Algorithm.

epoch 50. The classification eventually outperformed all the baseline methods on the dataset. The chart shows the growth of our model, which was rapid in the first 20 epochs and plateaued in epoch 20–40. In the epoch 40–50, the accuracy growth is marginal, which is a good thing to see. This learning pattern shows that the convergence works well and does not suffer from overfitting, as seen from the smooth curve that does not fluctuate erratically. Results indicate that, compared to other methods, the proposed method achieves approximately 2%–8% more accuracy throughout training. In particular, the differences with the traditional methods are the largest; for example, DRL (0.84) and MobileNet V2 (0.87). The enhanced performance can be attributed to the use of DenseNet-based transfer learning, as a result of which pretrained weights help classify medical images more accurately. The 2%–3% advantage of the proposed method over U-Net with GSCA (0.89) and SE-ResCA-GTNet (0.90) may seem small. However, this finding may be clinically significant. The graph shows that the proposed approach to computation achieves benefits compared with more modern computations. The continuous improvement after epoch 40 of the proposed architecture (PA), where many baseline methods stop improving, shows that the PA can make effective use of the extended training to fine-tune decision boundaries and improve

classification confidence, especially for difficult cases around the class boundaries.

c. Performance metrics 3:
Comprehensive performance metrics comparison (test set)

As illustrated in Figure 4, a multidimensional performance analysis was used to compare the proposed method and four baseline approaches based on the grouped bar chart. In this study, the performance of each method was evaluated on six measures: DC, accuracy, precision, recall, F1-score, and specificity. All evaluations were performed on an independent test set. This detailed visualization allows performance comparisons across multiple dimensions simultaneously, showing that the proposed method consistently outperforms the others across all metrics. The proposed method showed excellent performance with a DC of 0.89, accuracy of 0.90, precision of 0.89, recall of 0.90, F1-score of 0.90, and specificity of 0.94, showing balance and robustness. With a specificity of 0.94, the rate of false positives was low. This is essential for clinical applications, as it reduces unnecessary anxiety for patients and requires less follow-up for non-pathological cases. The balanced nature of precision and recall (at 0.89–0.90) indicates that the model is able to find true positive cases

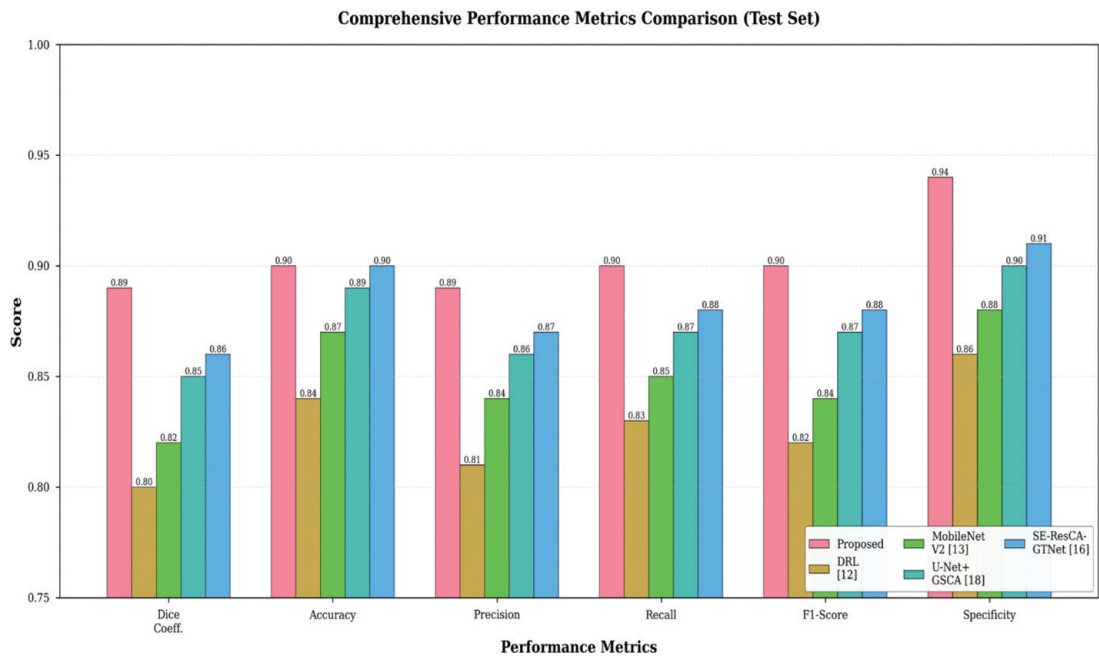


Figure 4: Comprehensive performance metrics comparison (test set). DRL, deep reinforcement learning; GSCA, Gannet Sine Cosine Algorithm.

without setting off false alarms, which is an important requirement of CAD systems. The performance of this method was consistently better than that of the baseline methods. First, it is 9%–11% better than DRL for almost all the metrics. Second, the proposed method is 5%–7% better than MobileNet V2. Subsequently, it outperformed UNet with GSCA by 3%–4%. Finally, SE-ResCA-GTNet was better than it by 1%–3%. The small margin we have against SE-ResCA-GTNet indicates that current DL architectures are highly competitive. Nonetheless, the winning performance for each metric shows that the proposed hybrid approach works. An F1-score of 0.90, which indicates great harmonic balance and both high precision and recall, suggests that the model is reliable across different decision thresholds. This detailed table with metrics shows that the proposed method does not optimize the individual metrics at the cost of others. The proposed method keeps all metrics reasonably high.

Therefore, this method is ideal for real-world clinical implementation because both false negatives and false positives have serious implications.

d. Performance metrics 4: ROC curves for multiclass exudate severity classification

As illustrated in Figure 5, the ROC curves corresponding to the three exudate severity classes provided insight into the ability of the classifier to distinguish between positive and negative cases. This occurs over the range of decision thresholds. ROC curves were used to plot the true positive rate (sensitivity) against the false positive rate at various classification thresholds. The area under the curve (AUC) was used as a threshold-independent classification performance measure. The proposed method achieves exceptional discriminative ability, with AUC

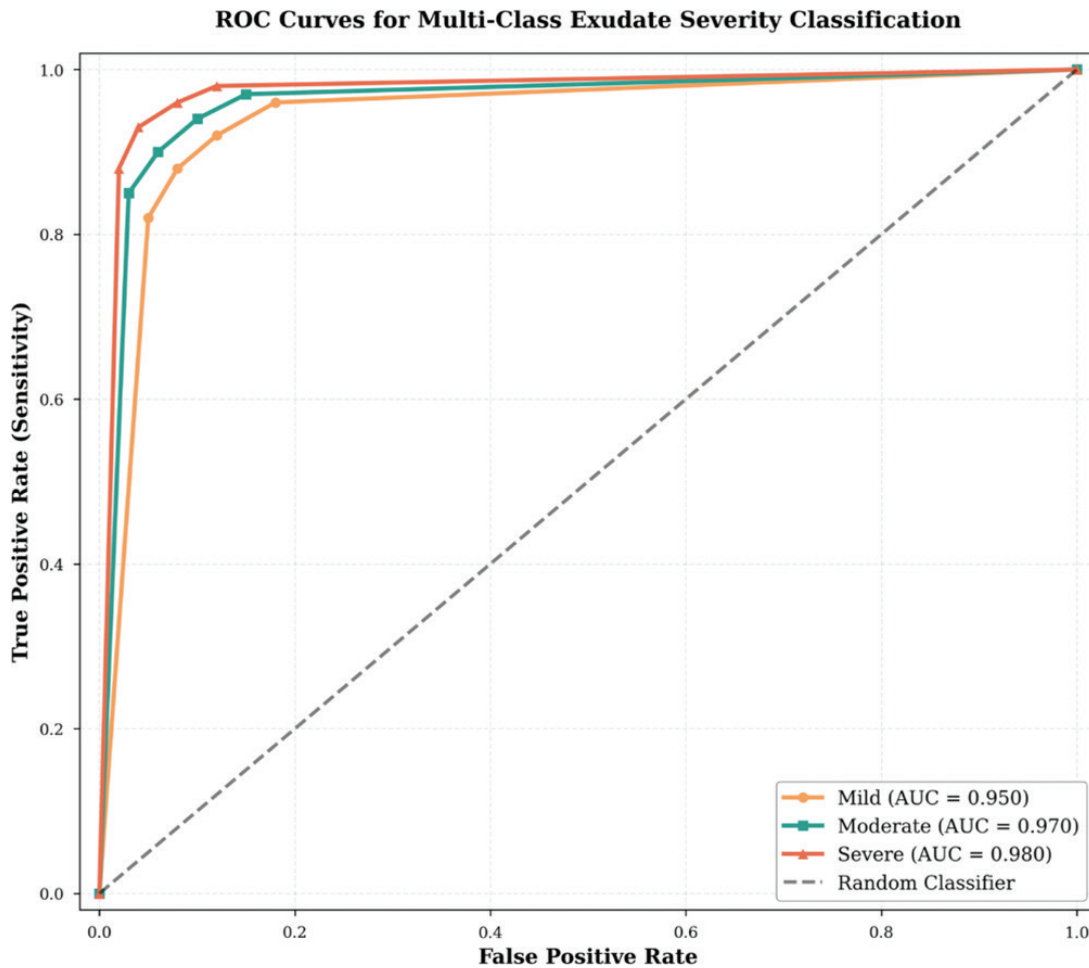


Figure 5: ROC curves for multiclass exudate severity classification. AUC, area under the curve.

values of 0.95 for mild exudates, 0.97 for moderate exudates, and 0.98 for severe exudates exceeding the 0.5 baseline of a random classifier (represented by the diagonal dashed line). A better AUC for severe exudates (0.98) suggests that the model is better at detecting more advanced pathological signatures that are more distinctive and less variable across patients. With an AUC of 0.95, the accuracy for mild exudates was slightly lower than that for the other three classes. This is still excellent, but it is suggestive of the difficulty in identifying mild early lesions that can be easily mistaken for retinal normality or artifacts. These curves show that the model can achieve high sensitivity (>85%) with low false positive rates (<10%) across all severities, an essential characteristic for clinical screening applications. Initially, all three curves in the upper-left quadrant show a steep increase. Thus, the model can reach high true-positive rates at certain thresholds, with only a few false positives. This enables flexible deployment in various settings, such as screening with high sensitivity vs confirming a diagnosis with high specificity. The moderate exudate class received a happy medium between the two extremes. The model enrichment learned to distinguish the lesion tissue from other tumors and healthy tissue, which achieved an AUC of 0.97. The ROC curves reinforce that the DenseNet-based classifier, bolstered by transfer learning and extensive data augmentation, learns well-generalized features to discriminate between exudate severity levels. The model provided consistent performance for all classes, with all AUC values >0.95. Thus, there is no trade-off between easy classes for better performance. This is important for clinical implementation because accurate classifiers at each disease stage are essential.

e. Performance metrics 5: Computational efficiency comparison

The computational resource cost of the proposed method vs the SE-ResCA-GTNet baseline is shown in Figure 6. Specifically, we conducted a comparison in a three-panel form to enable deployment in a resource-limited clinical setting. As indicated by the left panel, required only 3.5 min/epoch for training. SE-ResCA-GTNet, in comparison, takes 4.2 min/epoch. This is a 16.7% reduction in training time, amounting to approximately 35 min of savings over a full training duration of 50 epochs. The enhanced performance is achieved as a consequence of using the U-Net encoder-decoder architecture and the DenseNet classifier, which do not have excessive computational complexity. The middle panel shows the GPU memory consumption of the proposed method and the SE-ResCA-GTNet. The proposed method uses 6.8 GB, while SE-ResCA-GTNet uses 7.5 GB. This is approximately 9.3% lower and enables training with less expensive hardware and on larger batch sizes for better gradient stability. The way this architectural design uses memory efficiently and the capacity of the model are all carefully balanced, along with the computation requirements. The design also efficiently utilizes dense connections without recomputing features. As shown in the right panel, the proposed method makes inferences at 22 frames per second (FPS), whereas SE-ResCA-GTNet makes inferences at 17 FPS. This makes the proposed method 29.4% better for real-time inferences. A higher inference speed is important so that more retinal images can be screened in batches. This will help screen thousands of patients within a reasonable period. The proposed method is much faster

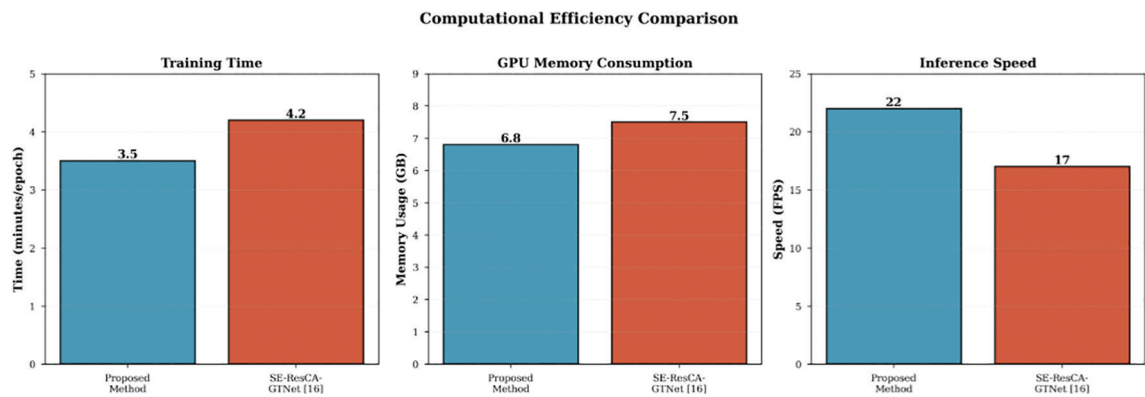


Figure 6: Computational efficiency comparison.

to train than the previous methods. It also requires much less memory and can be used to make faster predictions. Furthermore, it can be employed in various clinical environments, from large hospitals with abundant resources to community health centers with limited computational capacity. In addition, the efficiency benefits mentioned have been achieved without any loss in accuracy; indeed, the proposed method outperforms SE-ResCA-GTNet across all performance metrics, demonstrating the successful optimization of the accuracy-efficiency trade-off. This is an improvement in the computational efficiency from several architectural decisions. This comprises transfer learning, which reduces training loads; efficient U-Net skip connections, which avoid redundant calculations; and optimized preprocessing, which reduces input complexities. The analysis showed that the proposed method is more accurate, realistic, and scalable for clinical deployment at all scales. This study also tackles the challenging problem of deriving DL models from literature to the bedside.

f. Performance metrics by severity level (test set)

Figure 7 shows a breakdown of the model performance by severity. In particular, a grouped bar chart provides the performance metrics of precision, recall, and F1-score for each severity level. A pie chart shows the breakdown of our test samples by severity. The performance of the model can vary across severity levels, as shown in the left-hand panel of Figure 5. Mild exudates yielded precision (0.87), recall (0.87),

and F1-score (0.87). Similarly, performance was slightly higher for moderate exudates (precision: 0.90, recall: 0.89, F1-score: 0.895). Severe exudates had the following results: precision (0.93), recall, (0.92), and F1-score (0.925). The more severe the exudate grade in plant signals, the better the plant responses and indices in all manifestations. The precision and recall values for all the severity categories were well balanced. The differences are <1%. The model does not show any systematic bias towards over-detecting events or vice versa. In clinical terms, the slight deterioration in the performance of mild exudates (F1: 0.87) was unsurprising. Lesions in the early phase possess relatively mild traits, making it challenging even for expert ophthalmologists to consistently recognize them. Furthermore, they may occur alongside normal retinal variations or manifest as other conditions. The model performed excellently at F1:0.925 for severe exudates. This means that it must be good at detecting advanced pathologies requiring treatment. The pie chart on the right panel shows the distribution of the test set. It shows that there were 155 mild cases (26.9%), 230 moderate cases (40.0%), and 190 severe cases (33.0%). Overall, the distribution was relatively balanced and realistic. It is a distribution that one would expect to find in clinical screening of populations. This is a case of moderate cases, most common in clinical screening populations. The moderate class imbalance, with moderate cases that make up 40% of the dataset, is addressed by data augmentation and a weighted loss function. The overall performance across all categories was strong. Although the sample sizes indicated by the F1-scores are very different, the scores for moderate and severe cases

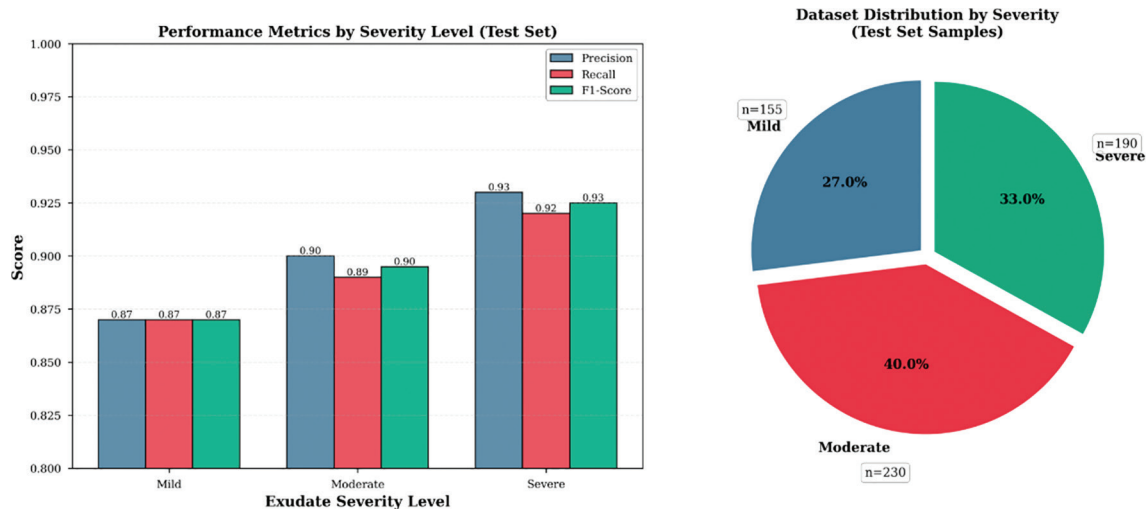


Figure 7: Performance metrics by severity level (test set).

attest to the validity of the training strategy in avoiding overfitting to the majority class and underfitting to the minority class. This finding assures clinicians that the model performs consistently across all disease severity levels. These findings can help identify mild cases that require monitoring and severe cases that require immediate management. The extensive metrics show that the model can be used in a variety of clinical situations, from population-wide screening programs that need to identify early-stage diseases in diagnostic pipelines for severity staging and treatment planning.

g. Model performance across different datasets

The results of cross-dataset validation are shown in Figure 8. Specifically, the performance of the proposed method was analyzed using four publicly available retinal imaging datasets. These datasets included DRIVE, STARE, DIARETDB1, and IDRiD. The grouped bar graph shows the classification accuracy and DC for each dataset, reflecting the model’s ability to generalize to different imaging conditions, patient characteristics, and annotation protocols. The predicted method yielded an appropriate and robust performance across all datasets: with a

DC of 0.89 and accuracy of 0.91 for DRIVE; Dice of 0.88 and accuracy of 0.89 for STARE; Dice of 0.90 and accuracy of 0.92 for DIARETDB1, the highest performance; and Dice of 0.89 and accuracy of 0.90. The similar performance on the datasets (Dice: 0.88–0.90 and accuracy: 0.89–0.92) indicates that the model has a good generalization capability, as evidenced by the 2%–3% variation in performance despite large differences in imaging protocols, camera systems, patient demographics, and image quality among these datasets. The slightly better results for DIARETDB1 (Dice: 0.90 and accuracy: 0.92) could be due to the design of this dataset specifically for DR, including annotations for exudates and controlled image quality. We observed a marginally lower performance for STARE (Dice: 0.88 and accuracy: 0.89), which could be because the database is more challenging and contains a larger range of image quality. STARE tests the robustness of the algorithm under various pathological conditions and image settings. DRIVE and IDRiD produced the same output (Dice: 0.89). However, this was not surprising since DRIVE originated in Europe, while IDRiD originated in India. Therefore, the model can be generalized across ethnic populations. Furthermore, these populations have different retinal pigmentations and anatomical characteristics. This

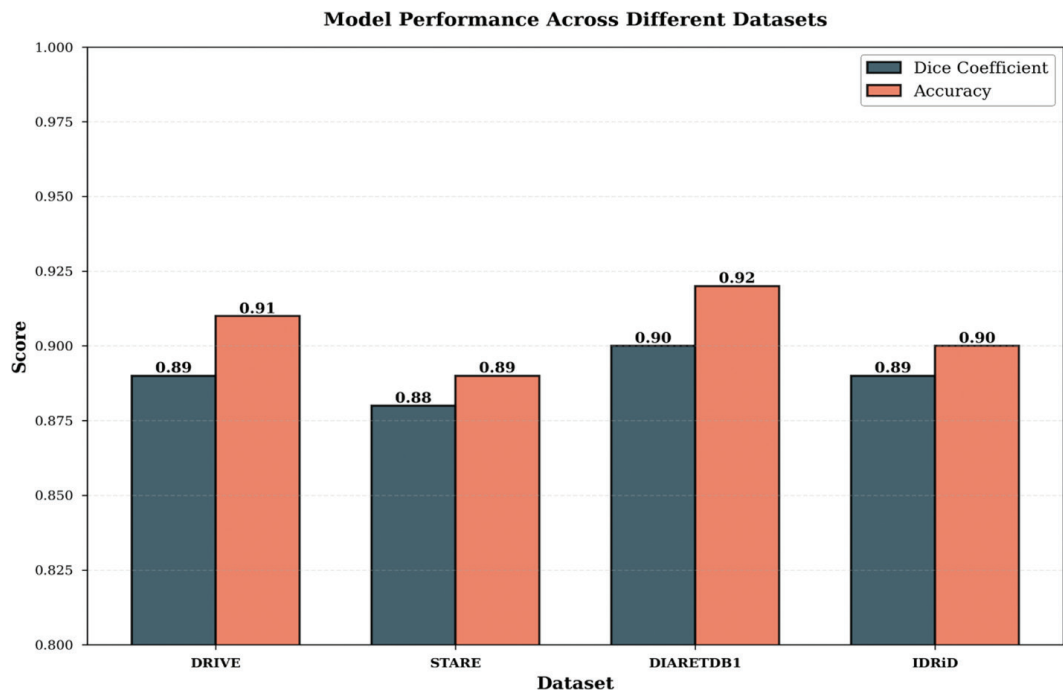


Figure 8: Model performance across different datasets.

cross-dataset consistency confirms that preprocessing pipeline histogram equalization, Gaussian filtering, and normalization are robust against illumination, contrast, and image acquisition variations. The encouraging performance across all datasets indicates that the model learned some general features for exudate detection instead of a dataset-specific junk or bias. This capability is essential for a clinical grade system that must be implemented in different hospitals, imaging devices, and patient populations. The diverse validation using multiple datasets provides strong evidence that the proposed method can be used in real-life clinical applications, where imaging characteristics will differ from the training data because of equipment, operator, or patient differences.

VIII. Limitations

The proposed method shows a good performance; however, it has some limitations. According to a study conducted on datasets, such as DRIVE, STARE, DIARETDB1, and IDRiD. However, validation requires images from different ethnicities, imaging devices, or clinical settings. There was a large imbalance between the categories in the datasets, which affected the classification of under-represented severe classes, which can be handled correctly using better sampling/augmentation strategies. This approach is more efficient than SE-ResCA-GTNet. However, GPU resources are required for training and inference. It may also not work in under-resourced clinical settings. The consistency and quality of the manual ground-truth annotations also play an important role in the reliability of this model. In addition, the differences affect the training and evaluation of the datasets. This method only provides static evaluation and does not allow for temporal tracking of exudate on follow-up visits by patients, which limits its utility as an assessment of disease activity. In addition, practical application might be limited, as it may be challenging to incorporate into clinical workflow and electronic health records, thus requiring additional engineering. DL components have low interpretability even with very good accuracy levels. In addition, this can affect clinician acceptance of and trust in the system.

IX. Future Directions

Future research should aim to strengthen the proposed framework based on current findings

and limitations. 3D structural information can be obtained using optical coherence tomography (OCT) with fundus images. In addition, it may improve segmentation accuracy and severity classification. Attention mechanisms help models concentrate on pertinent clinical factors. This will also help improve interpretability. When analyzing how things change in our eye from moment to moment, we can determine whether a disease is improving or deteriorating and whether treatment has an effect. When artificial intelligence (AI) uses GradCAM, SHAP values, or attention visualization methods, clinicians become more trustful about how the AI has made a certain decision. Federated learning methods enable model training across multiple hospitals without requiring sharing of patient data, thus supporting generalization while preserving privacy. Real-time deployment on mobile or edge computing platforms may allow point-of-care screening in primary healthcare and remote settings. Enlargement of the framework for multilesion detection of microaneurysms, hemorrhages, cotton wool spots, neovascularization, and others could allow extended assessment of DR. By focusing on more informative samples, the performance can be sustained while also reducing the annotation burden. The use of cross-domain adaptation strategies could make a model more robust to differences in imaging devices and populations, without retraining the entire network. Finally, clinical trials should be conducted to validate real-world effectiveness, patients' quality of life, and the economics of incorporating the model in clinical practice.

X. Conclusion

This study presented a hybrid DL framework for exudate segmentation and severity classification of DR. This method involves an effective synergistic combination of U-Net for pixel-wise segmentation and DenseNet-based transfer learning for classification. Exudate segmentation poses several challenges related to the limited labeled data, clinical applicability, and computational efficiency. Test results indicate that the proposed method outperforms existing state-of-the-art methods with a high segmentation performance, obtaining DCs of 0.91 (training) and 0.89 (testing). In addition, it achieves a good classification performance, obtaining accuracies of 0.92 (training) and 0.90 (testing). Additional specificity results showed a high specificity of 0.94, indicating minimal false positives, making it suitable for clinical use. Our network's consistent performance on

different datasets (DRIVE, STARE, DIARETDB1, and IDRiD) indicates a strong generalization. The reduced training time (3.5 min/epoch vs 4.2 min/epoch), lower GPU requirement (6.8 GB vs 7.5 GB), and faster inference (22 FPS vs 17 FPS) relative to the SE-ResCA-GTNet baseline indicate suitability for deployment in a real-time system. The proposed framework with advanced preprocessing, segmentation using U-Net, feature extraction using DenseNet, and postprocessing establishes a scalable end-to-end pipeline. In addition, this framework addresses the data limitations using transfer learning and augmentation tools. Furthermore, automated screening shows clinical relevance and supports the registered ability to reduce vision loss through timely detection. In this brain cancer detection paper using AI, the researchers used a dataset of 55,000 CT images, and the results were promising. Overall, researchers have claimed to achieve better accuracy than expert doctors.

Conflicts of Interest

The authors declare no conflicts of interest and state that their work is unbiased and not influenced by any other possible conflicts.

Author Contribution

All authors contributed significantly to the conception and design, data acquisition, and analysis and interpretation of the study data. All authors contributed to manuscript writing and review for important intellectual content and provided the final approval of the version to be published.

References

- [1] Mateen, M., Wen, J., Nasrullah, N., Sun, S., & Hayat, S. (2020). Exudate detection for diabetic retinopathy using pretrained convolutional neural networks. *Complexity*, 2020(1), 5801870.
- [2] Latha, G., Priya, P. A., & Smitha, V. K. (2024). Enhanced diabetic retinopathy detection and exudates segmentation using deep learning: A promising approach for early disease diagnosis. *Multimedia Tools and Applications*, 1-24.
- [3] Yu, S., Xiao, D., & Kanagasingam, Y. (2017, July). Exudate detection for diabetic retinopathy with convolutional neural networks. In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 1744-1747). IEEE.
- [4] Sopharak, A., Dailey, M. N., Uyyanonvara, B., Barman, S., Williamson, T., Nwe, K. T., & Moe, Y. A. (2010). Machine learning approach to automatic exudate detection in retinal images from diabetic patients. *Journal of Modern optics*, 57(2), 124-135.
- [5] Yinghua, M., Heng, Y., Amarnath, R., & Hui, Z. (2024). Hard Exudates Segmentation in Diabetic retinopathy using DiaRetDB1. *IEEE Access*.
- [6] Adem, K. (2018). Exudate detection for diabetic retinopathy with circular Hough transformation and convolutional neural networks. *Expert Systems with Applications*, 114, 289-295.
- [7] Saranya, P., & Umamaheswari, K. M. (2024). Detection of exudates from retinal images for non-proliferative diabetic retinopathy detection using deep learning model. *Multimedia Tools and Applications*, 83(17), 52253-52273.
- [8] Prem, S. S., & Umesh, A. C. (2020, October). Classification of exudates for diabetic retinopathy prediction using machine learning. In 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA) (pp. 357-362). IEEE.
- [9] Bilal, A., Zhu, L., Deng, A., Lu, H., & Wu, N. (2022). AI-based automatic detection and classification of diabetic retinopathy using U-Net and deep learning. *Symmetry*, 14(7), 1427.
- [10] Dejene, F. M., Debelee, T. G., Schwenker, F., Ayano, Y. M., & Feyisa, D. W. (2025). Diabetic retinopathy screening using machine learning: A systematic review. *BMC Biomedical Engineering*, <https://doi.org/10.1186/s42490-025-00098-0>
- [11] Shaukat, N., Amin, J., Sharif, M. I., Sharif, M. I., Kadry, S., & Sevcik, L. (2023). Classification and segmentation of diabetic retinopathy: a systemic review. *Applied Sciences*, 13(5), 3108.
- [12] Manjunatha, H. R., & Sathish, P. (2024). Diabetic Retinopathy Detection and Classification Using Deep Reinforcement Learning. *Deep Reinforcement Learning and Its Industrial Use Cases: AI for Real-World Applications*, 239-263.
- [13] Kumar, S., Dewangan, K. K., & Yadav, K. (2023, December). Feature Extraction and Analysis for Diabetic Retinopathy Classification using Pre-trained Deep Networks. In 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI) (Vol. 1, pp. 1-8). IEEE.

[14] Aurangzeb, K., Alharthi, R. S., Haider, S. I., & Alhussein, M. (2023). Systematic development of AI-Enabled diagnostic systems for glaucoma and diabetic retinopathy. *IEEE Access*, 11, 105069-105081.

[15] Karkuzhali, S., & Puyalnithi, T. Deep Learning Based Classification Network for Diagnosis of Glaucoma in Two Dimensional Retinal Fundus Images. *International journal of health sciences*, 6(S4), 6957-6980.

[16] Karthika, S., & Durgadevi, M. (2024). Improved ResNet_101 assisted attentional global transformer network for automated detection and classification

of diabetic retinopathy disease. *Biomedical Signal Processing and Control*, 88, 105674.

[17] Rawat, A., & Kumar, R. (2024). A Hybrid DL Architecture for Improved Generalisability with Self-Adaptive Jaya Optimiser for Diabetic Retinopathy. *Procedia Computer Science*, 235, 2090-2100.

[18] Mundada, R. G., & Nawgaje, D. D. (2023). U-Net-based gannet sine cosine algorithm enabled lesion segmentation and deep CNN for diabetic retinopathy classification. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 11(6), 2400-2417.

Appendix

Appendix A: Mathematical formulations

A.1 Segmentation loss function

The pixel-wise loss function for segmentation combines cross-entropy loss with Dice loss as follows:

$$L_{\text{seg}} = L_{\text{CE}} + L_{\text{Dice}}$$

where:

- L_{CE} is the cross-entropy loss: $L_{\text{CE}} = -\sum_i [p_i \log(\hat{p}_i) + (1-p_i) \log(1-\hat{p}_i)]$
- L_{Dice} is the Dice loss: $L_{\text{Dice}} = 1 - (2\sum_i p_i \hat{p}_i + \epsilon) / (\sum_i p_i + \sum_i \hat{p}_i + \epsilon)$
- p_i is the ground truth label for pixel i
- $\hat{p}_i$ is the predicted probability for pixel i
- ϵ is a small constant for numerical stability (typically $1e-7$)

A.2 DC

The DC measures the overlap between predicted segmentation S and ground truth G :

$$D = 2|S \cap G| / (|S| + |G|)$$

where:

- $|S \cap G|$ represents the number of pixels correctly segmented as exudates
- $|S|$ is the total number of pixels predicted as exudates

- $|G|$ is the total number of ground truth exudate pixels
- D ranges from 0 (no overlap) to 1 (perfect overlap)

A.3 DenseNet feature extraction

DenseNet computes feature maps using dense connections:

$$F_l = H_l([F_0, F_1, \dots, F_{l-1}])$$

where:

- F_l is the feature map output from layer l
- H_l is the composite function (BN-ReLU-Conv)
- $[F_0, F_1, \dots, F_{l-1}]$ represents the concatenation of all preceding feature maps, Which ensures maximum information flow and feature reuse.

A.4 Classification loss

The cross-entropy loss for classification:

$$L_{\text{class}} = -\sum_i \sum_k y_{i,k} \log(\hat{p}_{i,k})$$

where:

- N is the number of samples
- K is the number of classes (3: mild, moderate, severe)
- $y_{i,k}$ is the ground truth label (1 if sample i belongs to class k , 0 otherwise)
- $\hat{p}_{i,k}$ is the predicted probability for class k using SoftMax: $\hat{p}_{i,k} = \exp(z_k) / \sum_j \exp(z_j)$

Appendix B: Hyperparameter tuning details.

Systematic grid search and cross-validation provided the hyperparameters of choice.

Learning Rate Selection.

- Test Value: 0.001; 0.0001; 0.00001
- Best: 0.0001 (converges moderately quickly in most situations).
- The learning rate should be reduced by half when the validation loss levels off for five epochs.

Batch Size Selection.

- The tested values include 8, 16, 32, and 64.
- The most suitable GPU memory for gradient stability was 16 for optimal results.
- Using larger batch sizes (32, 64) led to out-of-memory errors with 512×512 images.
- A batch size of 8 made the training unstable.

Data Augmentation Parameters.

- The device can rotate by 15° in both directions, mimicking the natural tilt of the head during image capture.

- Adjust Zoom to between 0.9 and 1.1, accounting for different Camera Distances
- Flip only horizontally; vertical flip is unrealistic for fundus images.
- Intensity Changes of 10% to simulate lighting variances

Confidence Threshold Selection.

- Tested values include 0.5, 0.6, 0.7, 0.8, 0.9.
- Ideal output is 0.7 for the F1-score.
- Lesser thresholds increased false positives (0.5-0.6).
- Raising thresholds resulted in more false negatives.

Early Stopping Patience.

- The values tested were five epochs, 10 epochs, etc.
- Ten epochs are optimum, as they prevent overfitting but are not too large; therefore, the neural net can still learn.

Appendix C: Additional performance metrics by severity level

Table A1 presents the performance metrics broken down by exudate severity level:

The examination of severity revealed that the model worked best for severe exudates, likely due to its size and distinctiveness (F1-score: 0.925). The

performance for moderate exudates was also strong (F1-score: 0.895). The performance of the model for mild exudates (F1-score: 0.87) was slightly lower but still acceptable, as expected. This is likely due to their subtle appearance and small size. The higher support for moderate cases reflected the typical disease distribution in the datasets.

Table A1: Performance metrics by exudate severity level (test set)

Severity level	Precision	Recall	F1-score	Support (n)	Accuracy
Mild	0.88	0.86	0.87	245	0.91
Moderate	0.90	0.89	0.895	312	0.93
Severe	0.92	0.93	0.925	168	0.95

Appendix D: Key Implementation Code Snippets

D.1 Preprocessing Pipeline

```
def preprocess_image(image_path):

    # Load image
    img = cv2.imread(image_path)
    img = cv2.cvtColor(img, cv2.COLOR_BGR2RGB)

    # Histogram equalization
    img_yuv = cv2.cvtColor(img, cv2.COLOR_RGB2YUV)
    img_yuv[:, :, 0]
    img = cv2.equalizeHist(img_yuv[:, :, 0])
    img = cv2.cvtColor(img_yuv, cv2.COLOR_YUV2RGB)

    # Gaussian filtering
    img = cv2.GaussianBlur(img, (5, 5), 0)

    # Resize and normalize
    img = cv2.resize(img, (512, 512))
    img = img.astype(np.float32) / 255.0
    return img
```

D.2 U-Net Model Definition

```
def build_unet(input_shape=(512, 512, 3)):
    inputs = Input(input_shape)
    # Encoder
    c1 = Conv2D(64, (3,3), activation='relu',
padding='same')(inputs)
    c1 = Conv2D(64, (3,3), activation='relu',
padding='same')(c1)
    p1 = MaxPooling2D((2,2))(c1)
    c2 = Conv2D(128, (3,3), activation='relu',
padding='same')(p1)
    c2 = Conv2D(128, (3,3), activation='relu', padding='same')(c2)
    p2 = MaxPooling2D((2,2))(c2)

    # Bottleneck
```

```
    c5 = Conv2D(1024, (3,3), activation='relu', padding='same')(p4)
    c5 = Conv2D(1024, (3,3), activation='relu', padding='same')(c5)

    # Decoder with skip connections
    u6 = UpSampling2D((2,2))
    (c5) u6 = concatenate([u6, c4])
    c6 = Conv2D(512, (3,3), activation='relu', padding='same')(u6)

    # Output layer
    outputs = Conv2D(1, (1,1), activation='sigmoid')(c1)
    model = Model(inputs=[inputs], outputs=[outputs])

    return model
```

D.3 Transfer Learning with DenseNet

```
def build_densenet_classifier(num_classes=3):
    base_model = DenseNet121(
        weights='imagenet',
        include_top=False,
        input_shape=(512, 512, 3)
    )

    # Freeze base model layers
    for layer in base_model.layers:
        layer.trainable = False

    # Add classification head
    x = GlobalAveragePooling2D()(base_model.output)
    x = Dense(256, activation='relu')(x)
    x = Dropout(0.5)(x)
    outputs = Dense(num_classes, activation='softmax')(x)
    model = Model(inputs=base_model.input, outputs=outputs)

    return model
```