RESEARCH PAPER

Efficient Mask Detection and Stage Classification of Diabetic Retinopathy Using Lightweight CNN and U-Net Models

Kajal Kadusing Patil[1], Dr. Girija Chiddarwar[2], Dr. Kalpana sunil Thakre[3], Dr. Smita Chaudhari[4], Swati Shekapure[5]

[1]Student, Department of Computer Engineering, Marathwada Mitra Mandal’s College of Engineering, Pune, India

[2]Associate Professor, Department of Computer Engineering, Marathwada Mitra Mandal's College of Engineering, Pune, Maharashtra, India

[3]Professor, Department of Computer Engineering, Marathwada Mitra Mandal's College of Engineering, Pune, Maharashtra, India-411052

[4]Associate Professor, Department of Computer Engineering, Marathwada Mitra Mandal's College of Engineering, Pune, Maharashtra, India

[5]Associate Professor, Department of Computer Engineering, Marathwada Mitra Mandal's College of Engineering, Pune, Maharashtra, India

Email: Kchavhan8892@gmail.com[1], girijachiddarwar@mmcoe.edu.in[2], kalpanathakre@mmcoe.edu.in[3], smita.m.c@gmail.com[4], Swatishekapure@gmail.com[5]

Orchid id- Dr. Girija G.Chiddarwar : 0000-0002-1040-8152

Dr. Smita Chaudhari 0000-0002-4568-2039

Abstract

Effective treatment of diabetic retinopathy (DR), a major cause of vision impairment worldwide, depends on early identification. By using lightweight Convolutional Neural Networks (CNNs) and a customized U-Net architecture, this study offers an effective deep learning framework for automated classification and mask detection of diabetic retinopathy. The main goal is to maximize computing economy without sacrificing classification accuracy so that the solution can be used in clinical settings with limited resources. The suggested approach initially uses a classification technique such as CNN, MobileNetV2 and Proposed MobileNetV2 architecture to classify diabetic retinopathy images and then Efficient U-Net model is applied to accurately segment retinal lesions and anatomical features like exudates and microaneurysms. Publicly accessible (IDRiD) datasets were used in extensive trials, and performance was assessed using measures including as accuracy, F1-score, sensitivity, and processing cost. The findings show that, in comparison to conventional deep models, the suggested framework greatly reduces model complexity and inference time while achieving strong classification performance of 99% accuracy and segmentation accuracy of 98%. All things considered, a viable strategy for effective, precise, and scalable DR screening and monitoring systems is the combination of lightweight MobileNetV2 and Efficient U-Net models.

Keywords: Diabetic retinopathy (DR), Convolutional Neural Networks (CNNs), U-Net models, Deep Learning, Lightweight models

How to cite this article: Patil KK, Chiddarwar G, Thakre KS, Chaudhari S, Shekapure S. Efficient Mask Detection and Stage Classification of Diabetic Retinopathy Using Lightweight CNN and U-Net Models. Int J Drug Deliv Technol. 2026;16(59s): 851-866. DOI: 10.25258/ijddt.16.59s.100

1. Introduction

Diabetic Retinopathy (DR) is one of the most prevalent microvascular complications associated with long-term diabetes mellitus. It is caused by damage to the small blood vessels in the retina due to prolonged high blood sugar levels. Over time, this damage can result in leakage, blockage, and abnormal growth of blood vessels, ultimately leading to vision impairment and, in severe cases, permanent blindness. DR typically begins as a non-proliferative condition, progressing through various stages before reaching proliferative diabetic retinopathy (PDR), the most severe form characterized by the formation of new, fragile blood vessels that can bleed and cause retinal detachment [1].

According to the World Health Organization (WHO), DR is a leading cause of blindness in the global adult population, particularly among those of working age. The International Diabetes Federation (IDF) reports that over 530 million people are currently living with diabetes, and this number is expected to rise sharply in the coming decades. As a direct consequence, the global burden of diabetic retinopathy is also expected to escalate, putting immense pressure on healthcare systems, especially in low- and middle-income countries where screening resources and medical personnel may be scarce.

The progressive nature of DR makes it essential to detect the condition at an early stage, long before the appearance of vision loss symptoms. While many individuals with early-stage DR may be asymptomatic, the pathological changes occurring in the retina can be visualized and assessed using medical imaging techniques such as fundus photography. This early detection and continuous monitoring can significantly reduce the risk of severe complications and facilitate effective management of the disease [2].

1.1 Importance of Early Detection and Stage Classification for Effective Treatment

Timely detection of diabetic retinopathy is critical not only to prevent vision loss but also to enable interventions that can halt or slow disease progression. The treatment strategies for DR are highly dependent on the severity or stage of the condition [3]. For instance, early stages may only require lifestyle modifications and tighter glucose control, whereas more advanced stages necessitate aggressive treatments such as intravitreal injections, laser photocoagulation, or even vitreoretinal surgery.

To make informed treatment decisions, ophthalmologists rely on the classification of DR into distinct stages—No DR, Mild, Moderate, Severe Non-Proliferative DR, and Proliferative DR. Each stage corresponds to specific pathological features such as microaneurysms, intraretinal hemorrhages, venous beading, and neovascularization. Accurate and consistent identification of these stages is vital for determining follow-up schedules and therapeutic interventions [4].

Moreover, early diagnosis can have a tremendous impact on reducing the socioeconomic burden of diabetic retinopathy. Vision impairment not only affects the quality of life of individuals but also reduces productivity and increases dependency on caregivers. Therefore, the development of reliable, accessible, and cost-effective screening tools is essential to facilitate early-stage detection and classification, particularly in under-resourced healthcare environments.

1.2 Limitations of Traditional Diagnostic Methods (Manual Inspection, Time-Consuming)

Ophthalmologists or qualified graders currently perform the majority of the manual assessment of retinal fundus pictures for the diagnosis and stage of diabetic retinopathy. In this procedure, lesions including microaneurysms, haemorrhages, neovascular formations, hard and soft exudates, and others are clearly identified in the retinal pictures [5]. Despite being the top standard, manual examination has several serious drawbacks and is intrinsically subjective.

- Time-Consuming: Screening and grading large volumes of images is labor-intensive, leading to potential backlogs in high-demand clinical settings.
- Variability in Interpretation: Inter- and intra-observer variability can affect the consistency and reliability of diagnoses.
- Resource-Intensive: Expert analysis requires trained professionals who may not be readily available in rural or underdeveloped areas.
- Delayed Diagnosis: In some healthcare systems, delays between image capture and diagnosis can hinder timely treatment and follow-up.

These challenges highlight the urgent need for automated, accurate, and scalable diagnostic solutions that can assist or replace manual evaluation, especially in resource-constrained regions with high diabetes prevalence.

1.3 Motivation for Using Deep Learning, Especially Lightweight Models, in DR Detection

In response to the limitations of traditional diagnostic methods, the integration of artificial intelligence (AI), particularly deep learning, has emerged as a promising alternative for the automated detection and classification of diabetic retinopathy [6]. Deep learning models, especially Convolutional Neural Networks (CNNs), have shown excellent performance in medical image analysis tasks, often matching or exceeding human-level accuracy. CNN-based models can learn complex hierarchical features directly from image data, enabling robust detection of subtle retinal abnormalities. Similarly, U-Net architectures have become a standard in medical image segmentation tasks due to their ability to localize and delineate relevant structures with high precision. These models have been successfully applied to segment microaneurysms, exudates, hemorrhages, and other DR-related lesions from fundus images [6].

However, a significant drawback of many state-of-the-art deep learning models is their computational intensity. These models are typically large, deep, and require powerful GPUs for training and inference, making them impractical for real-time or on-device deployment, particularly in low-resource environments. The challenge, therefore, lies in designing models that are not only accurate but also lightweight and efficient.

Lightweight models—such as MobileNet, SqueezeNet—offer a solution by drastically reducing the number of parameters and computational requirements while maintaining competitive performance. These models can be deployed on mobile devices, embedded systems, and edge computing platforms, making AI-powered DR detection accessible to a broader population [7]. The use of lightweight architectures enables:

- Faster inference speeds suitable for real-time applications.
- Reduced power consumption and memory usage.
- Potential for deployment in portable diagnostic tools and mobile health applications.
- Improved scalability in mass screening programs.

Therefore, the creation of useful, deployable technologies that promote early diagnosis, particularly in underserved places, can result from merging lightweight CNNs with U-Nets for DR detection and classification.

1.4 Objectives of the Study

This research aims to develop an efficient, lightweight deep learning framework that integrates both segmentation and classification for diabetic retinopathy diagnosis [8] [9]. The primary objectives of this study are outlined as follows:

- To develop a lightweight U-Net model for accurate and efficient segmentation (mask detection) of DR-related retinal features, including.
- To construct a lightweight CNN-based classifier that can categorize retinal images into the No DR, microaneurysms, hemorrhages/exudates, intraretinal microvascular abnormalities, and neovascularization.
- To evaluate the performance of the proposed models using benchmark datasets, measuring accuracy, sensitivity, specificity, F1-score, and inference time.
- To provide an end-to-end system that not only improves diagnostic accuracy but also offers visual interpretability through segmentation masks, thereby aiding clinical decision-making.

By addressing both segmentation and classification tasks using optimized deep learning models, this research aims to create a holistic solution that bridges the gap between accuracy and accessibility in diabetic retinopathy diagnosis. The proposed approach holds promise for improving early detection, enabling scalable screening programs, and ultimately reducing the global burden of diabetes-related blindness.

This paper presents a structured deep learning approach for diabetic retinopathy (DR) detection, organized into five sections. Section 1 introduces DR's clinical significance and diagnostic challenges. Section 2 reviews existing methods, emphasizing class-specific limitations in IDRiD's 5-tier grading (No DR, Mild, Moderate [exudates/hemorrhages], Severe, Proliferative [neovascularization]). Section 3 details the methodology, including dataset preprocessing (IDRiD), Lightweight MobileNetV2 model design and Efficient U-Net for Segmentation process. Section 4 compares results, demonstrating superior accuracy of DL models in severe cases (Classes 3–4). Section 5 concludes the paper.

2. Literature Review

The literature review provides a comprehensive overview of existing techniques and advancements in the field of diabetic retinopathy detection and classification. It explores traditional image processing methods, classical machine learning approaches, and the evolution toward deep learning-based solutions. Special emphasis is placed on the role of Convolutional Neural Networks (CNNs) and U-Net architectures in medical image analysis. This section highlights key studies, methodologies, and their comparative performance to establish the foundation for the proposed work.

Wahab Sait AR [7] (2023) presented a lightweight deep learning-based DR severity grading system optimized for limited computational resources. The model incorporates image preprocessing techniques to mitigate noise and artifacts in fundus images. Feature extraction is performed using the YOLO V7 technique, followed by feature selection through a tailored quantum marine predator algorithm. A hyperparameter-optimized MobileNet V3 model predicts DR severity levels. Validated on APTOS and EyePACS datasets, the model achieves high accuracy, demonstrating its effectiveness in resource-constrained environments.

Islam M.T et al. [8] (2023) proposed LUVS-Net, a lightweight U-Net-based architecture designed for efficient retinal vessel segmentation in fundus images, particularly suitable for deployment on resource-constrained devices. The model addresses challenges such as low-quality images and varying acquisition conditions by reducing the number of trainable parameters, thereby decreasing computational complexity. Evaluations demonstrate that LUVS-Net achieves high segmentation performance with significantly fewer parameters compared to traditional models, making it a viable solution for real-time applications in telemedicine and mobile health platforms.

P. Geetha Pavani et al. [9] (2023) introduced RILBP-YNet, a fully automated deep learning model for simultaneous multiclass segmentation of retinal lesions, including microaneurysms, hemorrhages, hard exudates, and soft exudates, in diabetic retinopathy. The model integrates Residual Inception Learning with Local Binary Patterns into a Y-Net architecture, enhancing feature extraction and segmentation accuracy. Evaluated on standard datasets, RILBP-YNet demonstrates superior performance in accurately delineating multiple lesion types, highlighting its potential for comprehensive DR screening and aiding ophthalmologists in diagnosis and treatment planning.

Miron C et al. [10] (2022) presented a lightweight U-Net convolutional neural network tailored for efficient and robust iris segmentation in eye images. The model comprises 36 layers with approximately 148,000 parameters, significantly reducing computational requirements while maintaining high segmentation accuracy. Tested across five standard open-source benchmarks—BioSec, CasiaI4, CasiaT4, IITD, and UBIRIS—the network achieves state-of-the-art results, with F1 scores up to 98.61% and mean Intersection over Union (mIoU) up to 97.26%. The segmentation time is less than 1 ms per image on a Xeon CPU, demonstrating its suitability for real-time applications in biometric systems and mobile devices.

O. J. Afolabi et al. [11] (2021) proposed a hybrid approach combining U-Net Lite for optic disc segmentation and Extreme Gradient Boosting (XGB) for glaucoma classification. The lightweight U-Net Lite efficiently segments the optic disc from retinal fundus images, providing critical features for the subsequent classification stage. XGB utilizes these features to accurately distinguish between glaucomatous and non-glaucomatous eyes. The integrated system demonstrates high accuracy and reduced computational load, making it suitable for deployment in resource-limited settings and facilitating early detection of glaucoma.

S., Gayathri et al. [12] (2020) developed a lightweight convolutional neural network (CNN) for classifying diabetic retinopathy stages from retinal fundus images. The model focuses on reducing the number of parameters to enable deployment on devices with limited computational resources. Despite its compact architecture, the CNN achieves competitive accuracy in distinguishing between different DR stages, demonstrating its potential for use in mobile health applications and large-scale screening programs, especially in areas with limited access to specialized medical equipment.

Table 1: Comparative Analysis for the Literature Review

Author & Ref No.	Methodology Used	Datasets Used	Advantages	Results
S. Iqbal et al. [1]	LDMRes-Net, multiscale residual blocks for efficient segmentation	Retinal vessel & exudate datasets (unspecified)	Low parameter count (0.072M), high efficiency for edge/IoT	Improved segmentation accuracy; lightweight model

Bhutnal & Moparthi [2]	Di-UNet + CST + E-EffGaz (Ensemble EfficientNet)	IDRiD, DRR	High accuracy; IoMT-compatible	97.69% (DRR), 97.52% (IDRiD)
Le Tong et al. [3]	LiViT-Net with MobileViT+ blocks and parallel convolutions	DRIVE, CHASEDB1, HRF	Transformer-based; enhanced local-global feature capture	High segmentation accuracy across datasets
S. N. S. et al. [4]	U-Net + MobileNetV2 for feature extraction + EfficientNetB0	Not specified	Efficient hybrid pipeline for segmentation and classification	Effective DR stage detection
Gawate & Laddha [5]	Weighted ensemble of U-Net, U-Net++, and DuckNet	DRAC 2022	Handles class imbalance effectively	IoU: 0.52003, Dice: 0.63269
Weiwei Gao et al. [6]	BiSeNet V2 + mixed attention + ghost feature mapping	IDRiD	Low-complexity, high sensitivity to small lesions	Competitive segmentation accuracy
Wahab Sait AR [7]	MobileNet V3 + YOLOv7 features + quantum optimizer	APTOS, EyePACS	Fast and lightweight; feature selection optimized	High classification accuracy (not exact % listed)
Islam et al. [8]	LUVS-Net: Lightweight U-Net for vessel segmentation	DRIVE, CHASEDB1	Few parameters, good generalizability	Efficient segmentation with minimal resources
Pavani et al. [9]	RILBP-YNet for multiclass lesion segmentation	Public DR datasets (unspecified)	Simultaneous lesion detection, high granularity	Superior multiclass segmentation
Miron et al. [10]	Lightweight U-Net for iris segmentation	BioSec, CasiaI4, CasiaT4, IITD, UBIRIS	Fast (<1ms), accurate iris segmentation	F1: 98.61%, mIoU: 97.26%
Afolabi et al. [11]	U-Net Lite + XGBoost for glaucoma detection	Not specified	Combines segmentation + ML classification efficiently	High glaucoma classification accuracy
Gayathri et al. [12]	Lightweight CNN for DR classification	Fundus image datasets (unspecified)	Deployable on low-resource devices	Competitive accuracy across DR stages

In table 1, the comparative analysis presents a concise summary of 12 recent research papers focused on lightweight models for diabetic retinopathy detection and segmentation. It highlights the methodologies employed—ranging from U-Net variants and MobileNet to ensemble and transformer-based architectures—alongside the datasets used, performance metrics, and unique advantages of each approach. This comparison provides valuable insights into the efficiency, accuracy, and suitability of these models for real-time and low-resource deployment scenarios.

3. Related Work

3.1 Existing Diabetic Retinopathy (DR) Detection Methods

Over the years, numerous methods have been developed for the detection and classification of diabetic retinopathy (DR), ranging from traditional image processing techniques to modern deep learning approaches. Early DR detection systems primarily relied on handcrafted features extracted from retinal fundus images, such as vessel patterns, microaneurysms, and exudates. These features were then processed using classical machine learning algorithms like Support Vector Machines (SVMs), Random Forests, and k-Nearest Neighbors (k-NN) to classify disease severity. While these methods showed moderate success, they were often limited by their dependency on manual feature engineering and sensitivity to image quality and variation.

With the advent of deep learning, particularly convolutional neural networks (CNNs), the accuracy and robustness of DR detection have significantly improved. Deep learning models automatically extract hierarchical and discriminative features from input images, reducing the need for manual intervention. Large-scale public datasets such as EyePACS, IDRiD, Messidor, and APTOS have accelerated the development of data-driven solutions. Hybrid models that combine segmentation and classification have also emerged, leveraging U-Net for lesion localization and CNNs for severity grading. Despite significant advancements, challenges like class imbalance, subtle lesion variability, and computational complexity persist, especially in real-world, low-resource clinical settings.

3.2 Role of CNNs and U-Net Models in Medical Imaging

Convolutional Neural Networks (CNNs) have become the cornerstone of modern medical image analysis due to their superior ability to learn spatial hierarchies and extract complex features from image data. In diabetic retinopathy detection, CNNs have been widely used for classifying disease stages directly from fundus images. They automatically capture relevant patterns such as microaneurysms, hemorrhages, and exudates without relying on handcrafted features. CNN architectures like ResNet, VGG, and MobileNet have been fine-tuned for DR classification tasks, achieving expert-level performance in various studies.

U-Net, a specialized convolutional architecture designed for biomedical image segmentation, plays a critical role in identifying and localizing lesions and anatomical features in retinal images. Its encoder-decoder structure enables it to retain both high-level semantic information and low-level spatial details, making it ideal for segmenting minute retinal structures. Variants like U-Net++, Attention U-Net, and lightweight versions have further enhanced segmentation performance while reducing computational demands. In DR analysis, U-Net is often used to delineate lesions, which can then inform classification models for better interpretability and accuracy. The synergy between CNNs and U-Net models has led to end-to-end DR detection pipelines that are both accurate and efficient, making them increasingly suitable for integration into real-time clinical systems.

3.3 IDRiD (Indian Diabetic Retinopathy Image Dataset

The Indian Diabetic Retinopathy Image Dataset (IDRiD) is a comprehensive, publicly available dataset designed to facilitate research in automated diabetic retinopathy (DR) detection and grading, particularly within the Indian population. Developed as part of the "Diabetic Retinopathy: Segmentation and Grading Challenge" at the IEEE ISBI 2018 conference, IDRiD provides high-resolution retinal fundus images along with detailed annotations, making it a valuable resource for developing and evaluating image analysis algorithms.

Dataset Composition:

- Pixel-Level Annotations: Comprises 81 color fundus images with signs of DR. Each image includes binary masks for lesions such as microaneurysms (MA), hemorrhages (HE), hard exudates (EX), and soft exudates (SE). Optic disc (OD) masks are also provided for all 81 images.
- Image-Level Disease Grading: Contains 516 images, each graded by medical experts for DR and diabetic macular edema (DME) severity. DR grades range from 0 (no DR) to 4 (severe DR), and DME risk levels range from 0 (no DME) to 2 (severe DME). The dataset is split into 413 training images and 103 testing images, maintaining a balanced distribution of disease severity.
- Optic Disc and Fovea Center Localization: Provides center coordinates (X, Y) for the optic disc and fovea for all 516 images, facilitating tasks like anatomical landmark detection.

Data Acquisition:

- Images were captured at an eye clinic in Nanded, Maharashtra, India, using a Kowa VX-10 alpha digital fundus camera with a 50° field of view.
- Each image has a resolution of 4288×2848 pixels and is stored in JPEG format.
- Prior to imaging, patients underwent mydriasis using 0.5% tropicamide to dilate the pupils, ensuring high-quality image capture.

Applications:

IDRiD serves as a benchmark dataset for various research tasks, including:

- Lesion segmentation (e.g., detecting MA, HE, EX, SE).
- Disease severity grading for DR and DME.
- Localization of retinal anatomical structures like the optic disc and fovea.

Its comprehensive annotations and high-quality images make it ideal for developing and testing machine learning algorithms aimed at early detection and management of diabetic retinopathy.

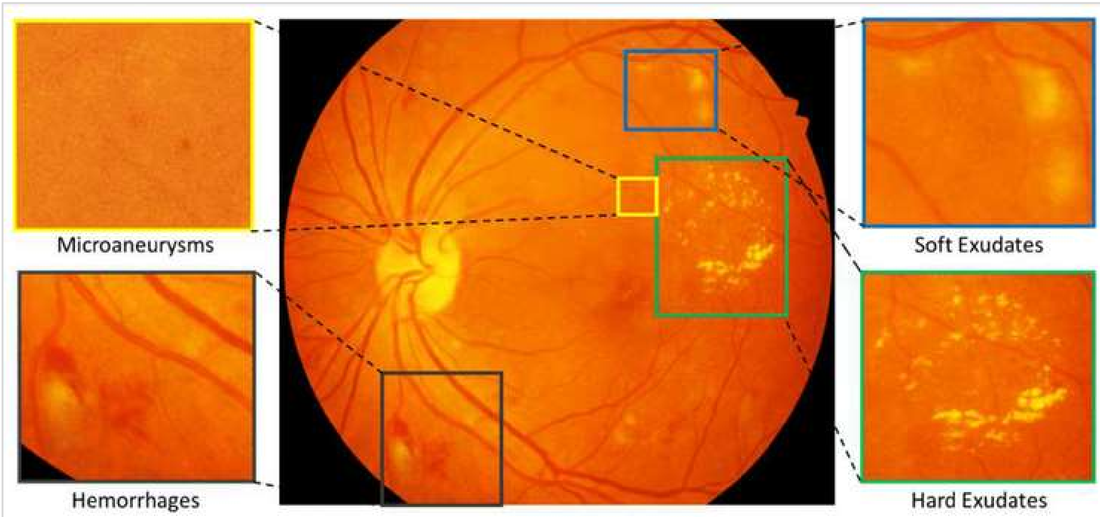

Figure 1. Dataset Sample with segmentation

4. Methodology

The proposed system architecture (shown in figure 2) for diabetic retinopathy classification and mask detection integrates advanced deep learning models and efficient data handling techniques. The process begins with data acquisition from sources like the IDRiD dataset, followed by preprocessing and augmentation. Data augmentation methods such as rotation, scaling, flipping, and color jittering are applied to enhance dataset diversity and address challenges in detecting small lesions. The preprocessing stage standardizes image formats and normalizes pixel values to improve image quality. For the classification task, a deep learning model based on advanced convolutional neural networks (e.g., MobileNetV2) is employed to classify retinal images into different stages of diabetic retinopathy. This model uses a softmax layer for multi-class classification, with performance evaluated through metrics like accuracy, F1-score, and AUC-ROC. Alongside, a Efficient U-Net model is developed for efficient mask detection. This model focuses on computational efficiency while maintaining high accuracy, using a modified architecture with techniques such as depthwise separable convolutions or pruning to reduce model size and computational load.

Integration of the classification and U-Net models involves preprocessing and augmenting images, then applying the models to generate classification and mask outputs. The system is designed to ensure that the combined results enhance both DR stage classification and lesion analysis. The models are then evaluated on validation datasets, with performance compared against baseline models and traditional U-Net variants to ensure accuracy and efficiency.

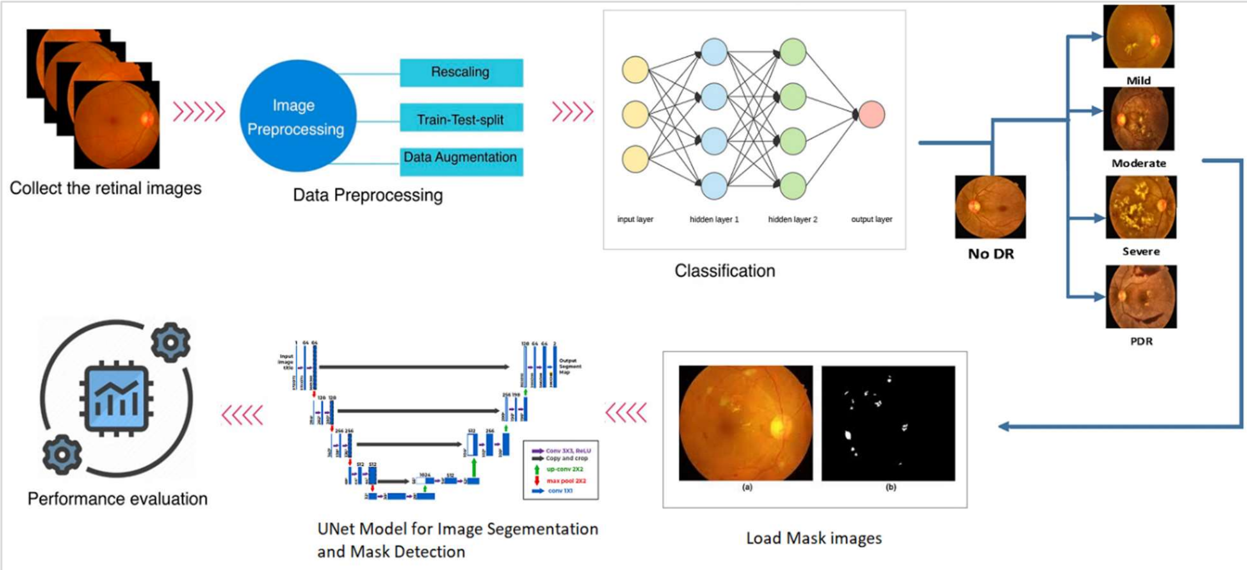

Figure 2. System Architecture Diagram

4.1 Load IDRiD Dataset

The Indian Diabetic Retinopathy Image Dataset (IDRiD) is employed for this study due to its high-resolution fundus images and detailed annotations. The dataset includes images captured from an Indian population using a 50° field-of-view fundus camera. For segmentation tasks, the dataset provides ground truth binary masks for lesions such as microaneurysms, hemorrhages, hard exudates, and soft exudates. Each color fundus image has corresponding ground truth masks annotated at the pixel level. The images are categorized according to diabetic retinopathy severity, enabling both classification and segmentation model development. These well-annotated samples provide a robust foundation for developing an end-to-end automated DR detection pipeline.

Segmentation:

- Original color fundus images (81 images divided into train and test set - JPG Files)
- Groundtruth images for the Lesions (Microaneurysms, Haemorrhages, Hard Exudates and Soft Exudates divided into train and test set - TIF Files) and Optic Disc (divided into train and test set - TIF Files)

Disease Grading:

- Original color fundus images (516 images divided into train set (413 images) and test set (103 images) - JPG Files)
- Groundtruth Labels for Diabetic Retinopathy and Diabetic Macular Edema Severity Grade (Divided into train and test set - CSV File)

Localization:

- Original color fundus images (516 images divided into train set (413 images) and test set (103 images) - JPG Files)
- Groundtruth Labels for Optic Disc Center Location (Divided into train and test set - CSV File)
- Groundtruth Labels for Fovea Center Location (Divided into train and test set - CSV File)

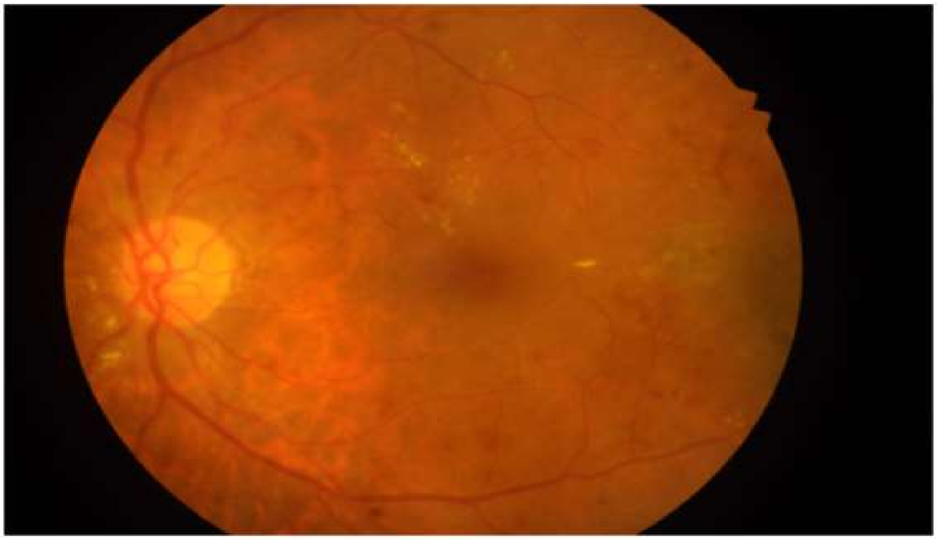

Figure 3. Image Sample from IDRiD Dataset

Figure 3 represents a high-resolution retinal fundus photograph from the IDRiD dataset. It displays the optic disc on the left, retinal vasculature, and visible lesions such as microaneurysms and hard exudates. These features are critical for diagnosing and classifying diabetic retinopathy. This image serves as the input for both lesion segmentation and disease classification models.

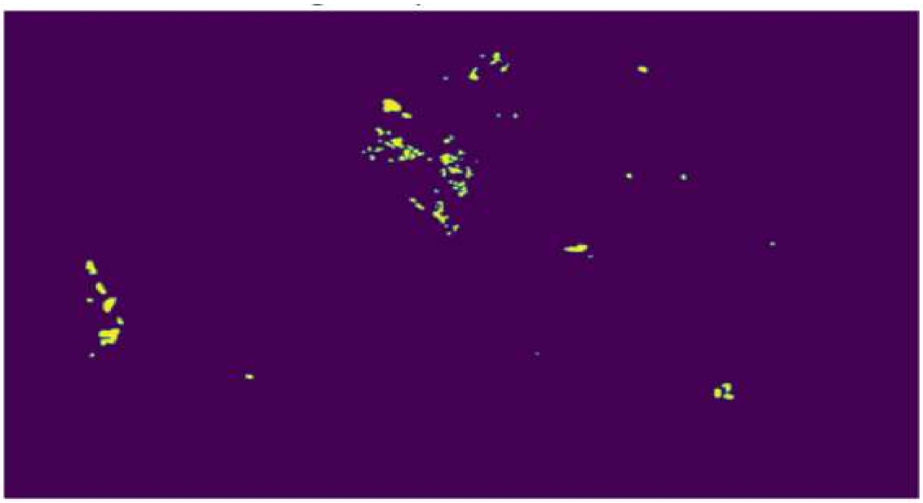

Figure 4. Ground truth of IDRiD Image Sample (Mask image)

Figure 4 is a ground truth binary segmentation mask corresponding to a retinal fundus image from the IDRiD dataset. The yellow regions indicate annotated lesions—most likely hard exudates—while the dark background represents non-lesion areas. This mask is used to train and validate segmentation models such as U-Net, enabling them to learn pixel-level localization of diabetic retinopathy-related abnormalities.

4.2 Preprocessing

To prepare the dataset for model training, a series of preprocessing steps are applied:

- Load Images and Masks: Color fundus images are loaded in RGB format, while corresponding lesion masks are loaded as grayscale images.
- Resize: Both images and masks are resized uniformly to 512×512 pixels to ensure consistency in input dimensions for the deep learning model.

- Convert to Arrays: Loaded image and mask files are converted into NumPy arrays to facilitate efficient processing and integration with deep learning frameworks.
- Expand Mask Dimensions: Since grayscale masks are 2D, their shape is expanded to (512, 512, 1) to match the input shape expected by the neural network.
- Convert Data Types: Masks are cast to float32 data type to ensure compatibility with TensorFlow training functions and loss computations.

These preprocessing steps ensure that the input data is in the optimal format for training a lightweight U-Net model for lesion segmentation and a CNN for DR classification.

4.3 Data Augmentation

To enhance model generalization and reduce overfitting, data augmentation techniques are applied to both the training images and corresponding masks. Augmentation increases the diversity of the training data, simulating variations commonly seen in real-world clinical environments. The following operations are performed:

- Rotation: Random rotations between -30° and +30° to simulate head movement during imaging.
- Flipping: Horizontal and vertical flips to introduce spatial variance.
- Zoom and Shift: Minor zooming and translation to account for changes in camera focus and patient alignment.
- Brightness Adjustment: Slight random changes in brightness to handle lighting inconsistencies during image capture. These transformations are applied using data augmentation libraries compatible with TensorFlow and Keras. Augmentations are synchronized between images and masks in segmentation tasks to ensure spatial consistency.

Table 2: Augmentation Techniques

Parameter	Value
rotation_range	15
width_shift_range	0.1
height_shift_range	0.1
shear_range	0.1
zoom_range	0.1
horizontal_flip	True
fill_mode	'nearest'

4.4 Stage Classification Using Transfer Learning Models

After lesion segmentation, the next step involves classifying the stage of diabetic retinopathy using transfer learning with pre-trained CNN architectures. The fundus images are fed into deep learning models that have been fine-tuned on the IDRiD dataset.

- *CNN Model:* The baseline Convolutional Neural Network (CNN) model is designed as a custom deep learning architecture for classifying diabetic retinopathy stages from fundus images. It consists of sequential convolutional layers with ReLU activation functions, each followed by max-pooling layers to downsample the spatial dimensions. The convolutional layers help extract spatial hierarchies and local features such as blood vessels, microaneurysms, and exudates. After several convolution-pooling blocks, the output is flattened and passed through fully connected (dense) layers to perform classification into five DR stages: No DR, Mild, Moderate, Severe, and Proliferative DR. Dropout layers are used between dense layers to prevent overfitting and enhance model generalization. While this CNN model is relatively shallow compared to pretrained architectures, it offers flexibility in design and provides a strong foundation for model performance benchmarking. However, its limited depth may restrict its ability to extract high-level abstractions compared to deeper networks, and it typically requires more training epochs to converge.
- *MobileNetV2 Model:* MobileNetV2 is a state-of-the-art lightweight deep learning architecture developed by Google for

efficient image classification tasks on mobile and embedded devices. It is particularly known for its use of depthwise separable convolutions and inverted residual blocks, which reduce the computational complexity and number of parameters significantly compared to traditional CNNs. In this study, MobileNetV2 is employed as a feature extractor for diabetic retinopathy classification. The model is initialized with ImageNet-pretrained weights to leverage knowledge from a large-scale image corpus. The final classification layers are customized with a global average pooling layer, followed by dense layers tailored to output five DR stage classes. MobileNetV2 is capable of learning both fine-grained and high-level features from fundus images, such as microaneurysms, hemorrhages, and neovascular formations. It strikes a balance between performance and efficiency, making it ideal for real-world applications where high accuracy and fast inference are required on devices with limited resources.

- *Lightweight_MobileNetV2 Model:* The Lightweight_MobileNetV2 model is an optimized and truncated version of the original MobileNetV2 architecture, specifically tailored to reduce computational load while maintaining high classification accuracy. This model retains the core architectural principles of MobileNetV2, such as depthwise separable convolutions and inverted residuals, but significantly reduces the number of layers and parameters. It is especially beneficial for real-time applications in low-resource settings, such as rural clinics or mobile health units, where computational power and memory are limited. Despite being lighter, the model still captures essential retinal features needed to distinguish between different DR severity levels. The final dense layers are custom-fitted for five-class classification, and transfer learning is applied using pre-trained weights from the original MobileNetV2. This ensures faster convergence and better generalization even with limited training data. The Lightweight_MobileNetV2 model is particularly suited for integration into IoT-enabled medical devices or mobile apps for DR screening, offering a practical solution for scalable deployment.

4.5 Segmentation using Efficient UNET Model

The Efficient_U-Net model implemented in this study is a customized version of the classical U-Net architecture, optimized for high segmentation accuracy with significantly reduced computational overhead. It integrates components from efficient convolutional networks such as MobileNetV2, using depthwise separable convolutions and inverted residual blocks in the encoder, while retaining the U-Net's symmetric decoder structure for precise spatial localization. Figure. 5 shows the UNET Model overview.

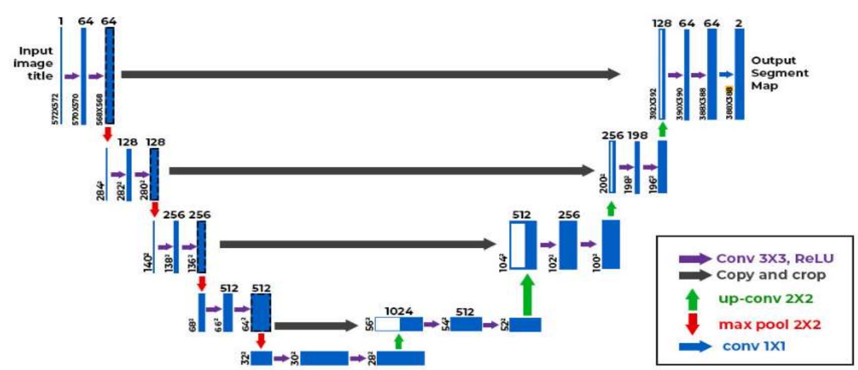

Figure 5. UNET Model overview

The architecture accepts an input size of 512×512×3 and employs multiple stages of downsampling and upsampling operations. The encoder extracts hierarchical features using lightweight modules including depthwise convolutions, batch normalization, and ReLU activations. A global average pooling layer is also used to compress features and enhance contextual representation. In the decoder, feature maps are progressively upsampled and refined through convolutional layers and skip connections. The final segmentation map is generated using a Conv2D layer with a sigmoid activation for binary segmentation. Figure 6 shows the Efficient_UNET Model summary.

Model: "functional"

Layer (type)	Output Shape	Param #	Connected to
input_layer (InputLayer)	(None, 512, 512, 3)	0	-
stem_conv (Conv2D)	(None, 256, 256, 48)	1,296	input_layer[0][0]
stem_bn (BatchNormalization)	(None, 256, 256, 48)	192	stem_conv[0][0]
stem_activation (Activation)	(None, 256, 256, 48)	0	stem_bn[0][0]
block1a_dwconv (DepthwiseConv2D)	(None, 256, 256, 48)	432	stem_activation[0][0]
block1a_bn (BatchNormalization)	(None, 256, 256, 48)	192	block1a_dwconv[0][0]
block1a_activation (Activation)	(None, 256, 256, 48)	0	block1a_bn[0][0]
block1a_se_squeeze (GlobalAveragePooling2D)	(None, 48)	0	block1a_activation[0]…
decoder_stage4a_conv (Conv2D)	(None, 512, 512, 16)	4,608	decoder_stage4_upsamp…
decoder_stage4a_bn (BatchNormalization)	(None, 512, 512, 16)	64	decoder_stage4a_conv[…
decoder_stage4a_relu (Activation)	(None, 512, 512, 16)	0	decoder_stage4a_bn[0]…
decoder_stage4b_conv (Conv2D)	(None, 512, 512, 16)	2,304	decoder_stage4a_relu[…
decoder_stage4b_bn (BatchNormalization)	(None, 512, 512, 16)	64	decoder_stage4b_conv[…
decoder_stage4b_relu (Activation)	(None, 512, 512, 16)	0	decoder_stage4b_bn[0]…
final_conv (Conv2D)	(None, 512, 512, 1)	145	decoder_stage4b_relu[…
sigmoid (Activation)	(None, 512, 512, 1)	0	final_conv[0][0]

Total params: 25,735,017 (98.17 MB)
Trainable params: 25,607,833 (97.69 MB)
Non-trainable params: 127,184 (496.81 KB)

Fig. 6. Summary of the Efficieint_UNET Model

As seen in the model summary (figure 6):

- Total Parameters: 25.7 million
- Trainable Parameters: 25.6 million
- Non-Trainable Parameters: 127,184
- Final Output Shape: (512, 512, 1) – a binary mask for lesion detection

This Efficient_U-Net provides a good trade-off between accuracy and efficiency, making it suitable for real-time DR lesion segmentation in clinical or mobile settings.

4.6 Train –Test Split (80 % – 20 %)

To evaluate the model's performance in a reliable and unbiased manner, the dataset is split into two subsets: 80% for training and 20% for testing. The training set is used to optimize the model's parameters, while the testing set is reserved for evaluating the model's generalization capability on unseen data. This split ensures that the model learns from a diverse and representative portion of the dataset, while the test set serves as a benchmark for real-world performance. Care is taken to maintain class balance across both sets, especially in the classification task, to avoid bias due to class imbalance. For segmentation, both images and their corresponding masks are split simultaneously to preserve the one-to-one mapping. Similarly, for stage classification, the fundus images and associated labels are split while ensuring consistency in disease severity distribution. This 80/20 strategy is widely adopted in deep learning studies to provide a standard baseline for model evaluation.

5. Results

This work involves a two-step approach for diabetic retinopathy (DR) detection: classification of DR stages followed by segmentation to localize specific lesions and abnormalities within the retinal images. The results and analysis approach for each phase are as follows:

A. Classification

To evaluate the stage classification performance of the three deep learning models—CNN, MobileNetV2, and Lightweight_MobileNetV2—four key performance metrics were used: Accuracy, Precision, Recall, and F1-Score. Figure. 7 shows the accuracy and loss comparison of Lightweight MobileNetV2 model.

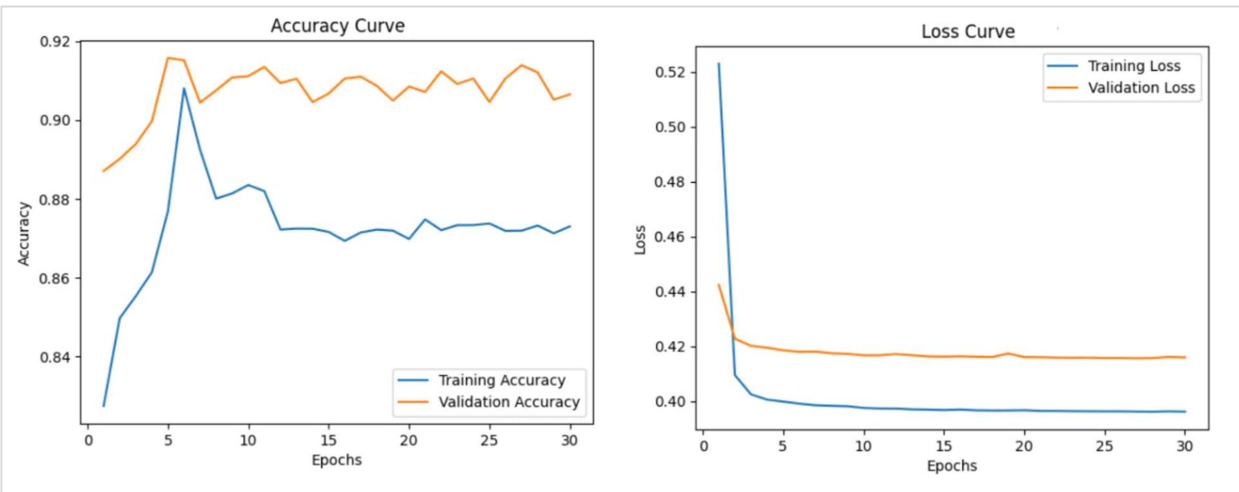

Figure 7. Lightweight MobileNetV2 Accuracy and Loss Curve

The Figure 7 illustrates a performance comparison of three deep learning models—CNN, MobileNetV2, and Lightweight_MobileNetV2—used for diabetic retinopathy stage classification. Four key evaluation metrics are compared: Accuracy, Precision, Recall, and F1-Score. Among the models, Lightweight_MobileNetV2 outperforms the others with an accuracy of 99.52%, precision of 91.74%, recall of 93.35%, and an F1-score of 94.58%. MobileNetV2 also shows strong performance across all metrics, while the standard CNN, although accurate, yields a significantly lower F1-score (76.75%). This comparative analysis highlights that optimizing models for lightweight deployment can improve both classification performance and computational efficiency

.

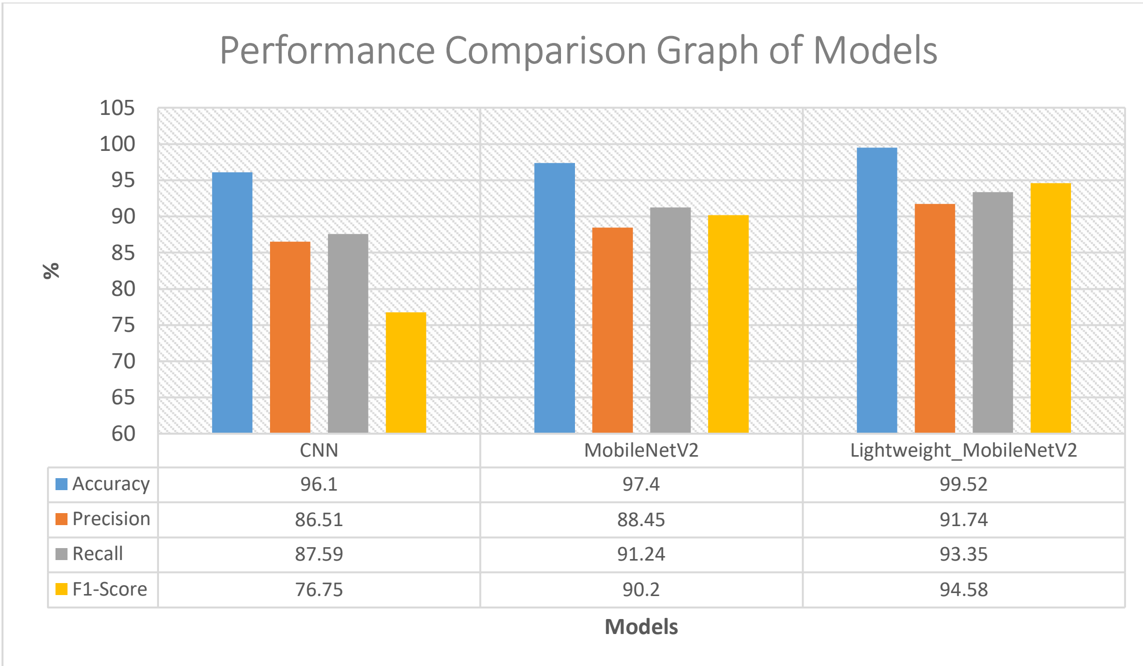

Fig. 8. Performance Comparison Graph

B. Segmentation (Efficient_U-Net model)

The performance of the Efficient_U-Net model was evaluated based on training and validation accuracy and loss over 10 epochs. The accuracy and loss curves (Figure 9) clearly demonstrate the model's rapid convergence and high segmentation performance.

- Accuracy Curve: The training accuracy (green line) steadily increases and approaches 1.0, indicating the model's ability to learn key features during segmentation. The validation accuracy (blue line) also rises sharply, reaching near-perfect accuracy by the third epoch, and stabilizes afterward. This suggests the model generalizes well without overfitting.
- Loss Curve: The training loss (green line) consistently remains low, while the validation loss (blue line) significantly drops from a high initial value to nearly zero within the first few epochs. The convergence of both loss curves indicates stable and effective learning.

These results highlight the model's efficiency and precision in segmenting retinal lesions, validating the suitability of the Efficient_U-Net architecture for diabetic retinopathy segmentation, especially in scenarios requiring fast and lightweight inference. The accuracy curve shows a rapid rise in both training and validation accuracy, reaching near-perfect values by the third epoch and remaining stable, indicating effective learning and generalization. The loss curve demonstrates a sharp decline in both training and validation loss, with validation loss dropping significantly from over 30 to nearly zero, confirming fast convergence and minimal overfitting. These results highlight the model's strong segmentation capabilities.

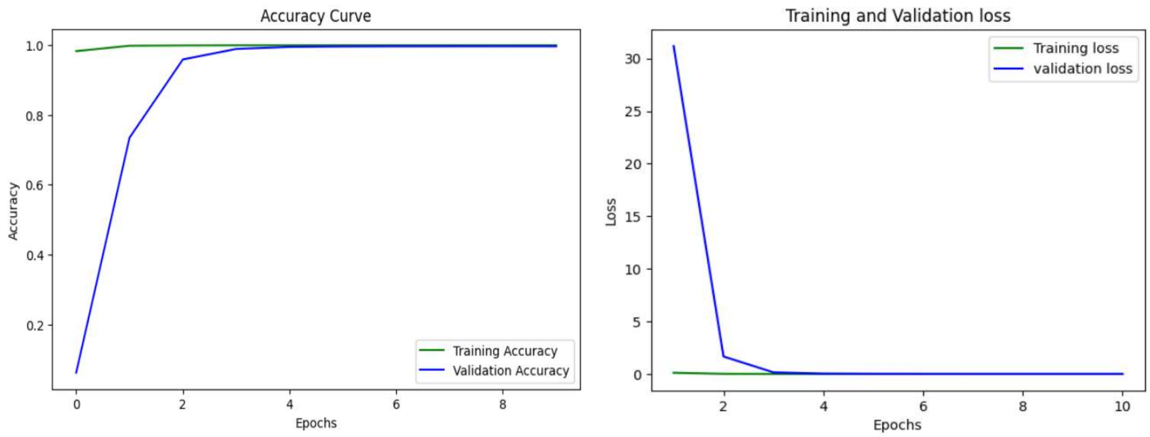

Fig. 9. Accuracy and Loss Comparison graph of Efficient UNet Model

6. Conclusion

This research introduces an efficient and lightweight deep learning framework for the automated segmentation and stage classification of diabetic retinopathy (DR) using fundus images. The proposed system integrates an optimized U-Net architecture for precise lesion segmentation and three transfer learning-based models—CNN, MobileNetV2, and Lightweight_MobileNetV2—for robust classification of DR stages. By leveraging the IDRiD dataset, the models were trained and evaluated under a standardized pipeline, including preprocessing, augmentation, and model-specific optimization. Among the classifiers, Lightweight_MobileNetV2 outperformed others with an accuracy of 99.52%, precision of 91.74%, recall of 93.35%, and an F1-score of 94.58%. The Efficient_U-Net model also demonstrated rapid convergence and high segmentation performance with minimal loss, validating its suitability for real-time medical applications. The combination of segmentation and classification enables a more comprehensive and interpretable DR diagnostic pipeline. The lightweight nature of the models ensures they are well-suited for deployment on mobile or edge devices, making them valuable for screening in resource-constrained environments. Overall, the proposed approach effectively addresses the challenges of accuracy, interpretability, and scalability in DR detection. This study contributes to the growing body of work on AI-assisted ophthalmology and sets the foundation for future real-world applications aimed at reducing vision impairment due to diabetic retinopathy.

References

[1] S. Iqbal et al., "LDMRes-Net: A Lightweight Neural Network for Efficient Medical Image Segmentation on IoT and Edge Devices," in IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 7, pp. 3860-3871, July 2024, https://doi.org/10.1109/JBHI.2023.3331278

[2] Bhutnal, V., Moparthi, N.R. IoMT enabled diabetic retinopathy segmentation and classification using ensemble efficient net model. Multimed Tools Appl (2024). https://doi.org/10.1007/s11042-024-19804-6

[3] Le Tong, Tianjiu Li, Qian Zhang, Qin Zhang, Renchaoli Zhu, Wei Du, Pengwei Hu, LiViT-Net: A U-Net-like, lightweight Transformer network for retinal vessel segmentation, Computational and Structural Biotechnology Journal, Volume 24, 2024, Pages 213-224, ISSN 2001-0370, https://doi.org/10.1016/j.csbj.2024.03.003

[4] S. N. S, S. K. Bag and S. Singh, "Enhanced Diabetic Retinopathyanalysis: Unet-Based Lesion Segmentation Coupled with Mobilenetv2 for Feature Extraction and Efficientnetb0 for Classification," 2024 Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), India, 2024, pp. 1-5, https://doi.org/10.1109/INCOS59338.2024.10527466

[5] E. Gawate and S. Laddha, "Addressing Class Imbalance in Diabetic Retinopathy Segmentation: A Weighted Ensemble Approach with U-Net, U-Net++, and DuckNet," 2024 International Conference on Intelligent Systems for Cybersecurity (ISCS), Gurugram, India, 2024, pp. 1-6, https://doi.org/10.1109/ISCS61804.2024.10581308

[6] Weiwei Gao, Bo Fan, Yu Fang, Nan Song, Lightweight and multi-lesion segmentation model for diabetic retinopathy based on the fusion of mixed attention and ghost feature mapping, Computers in Biology and Medicine, Volume 169, 2024, 107854, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2023.107854

[7] Wahab Sait AR. A Lightweight Diabetic Retinopathy Detection Model Using a Deep-Learning Technique. Diagnostics (Basel). 2023 Oct 3;13(19):3120. PMID: 37835861; PMCID: PMC10572365. https://doi.org/10.3390/diagnostics13193120

[8] Islam, M.T.; Khan, H.A.; Naveed, K.; Nauman, A.; Gulfam, S.M.; Kim, S.W. LUVS-Net: A Lightweight U-Net Vessel Segmentor for Retinal Vasculature Detection in Fundus Images. Electronics 2023, 12, 1786. https://doi.org/10.3390/electronics12081786

[9] P. Geetha Pavani, B. Biswal, Tapan Kumar Gandhi, Simultaneous multiclass retinal lesion segmentation using fully automated RILBP-YNet in diabetic retinopathy, Biomedical Signal Processing and Control, Volume 86, Part B, 2023, 105205, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2023.105205

[10] Miron, C., Pasarica, A., Manta, V. et al. Efficient and robust eye images iris segmentation using a lightweight U-net convolutional network. Multimed Tools Appl 81, 14961–14977 (2022). https://doi.org/10.1007/s11042-022-12212-8

[11] O. J. Afolabi, G. P. Mabuza-Hocquet, F. V. Nelwamondo and B. S. Paul, "The Use of U-Net Lite and Extreme Gradient Boost (XGB) for Glaucoma Detection," in IEEE Access, vol. 9, pp. 47411-47424, 2021, https://doi.org/10.1109/ACCESS.2021.3068204

[12] S., Gayathri & Gopi, Varun & Ponnusamy, Palanisamy. (2020). A lightweight CNN for Diabetic Retinopathy classification from fundus images. Biomedical Signal Processing and Control. 62. 102115. http://dx.doi.org/10.1016/j.bspc.2020.102115