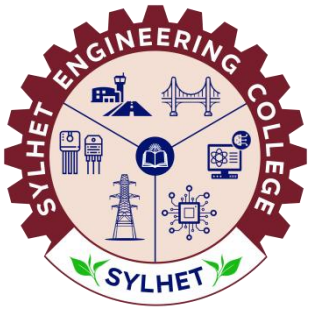
A Thesis Submitted to the Sylhet Engineering College for the Degree of
Bachelor of Science in Electrical and Electronic Engineering

Multi-Task Deep Learning-Based Approaches for Comprehensive Diabetic Retinopathy Assessment: Severity Classification, Lesion Localization, and Retinal Vessel Segmentation

By
Bishal Roy &
Syed Abu Sufwan

Supervised by ,
Salman Fazle Rabby
Assistant Professor
Department of Electrical and Electronic Engineering
Sylhet Engineering College



June, 2025
Sylhet Engineering College, Sylhet

The thesis titled **"Multi-Task Deep Learning-Based Approaches for Comprehensive Diabetic Retinopathy Assessment: Severity Classification, Lesion Localization, and Retinal Vessel Segmentation"** submitted by **Bishal Roy** & **Syed Abu Safwan**; Student ID: **2018338515** and **2018338540**; Session **2019-2020**, to the Department of Electrical and Electronic Engineering, Sylhet Engineering College, has been accepted as satisfactory in partial fulfillment of the requirements for the Degree of Bachelor of Science in Electrical and Electronic Engineering and approved as to its style and contents.

BOARD OF EXAMINERS

_____
Md. Shahid Iqbal — Chairman
Assistant Professor and Head
Department of Electrical and Electronic Engineering
Sylhet Engineering College, Sylhet.


_____
Salman Fazle Rabby — Member & Supervisor
Assistant Professor
Department of Electrical and Electronic Engineering
Sylhet Engineering College, Sylhet.

_____
Apurba Biswas — Member
Assistant Professor
Department of Electrical and Electronic Engineering
Sylhet Engineering College, Sylhet.

_____
Md. Ashraful Alam — Member
Lecturer
Department of Electrical and Electronic Engineering
Sylhet Engineering College, Sylhet.

_____
Mahedi Kamal Ahmed — Member
Lecturer
Department of Electrical and Electronic Engineering
Sylhet Engineering College, Sylhet.


_____
Arif Ahammad — Member (External)
Assistant Professor
Department of Electrical and Electronic Engineering
Shahjalal University of Science & Technology, Sylhet

Acknowledgements

We begin by extending our gratitude to the Almighty, whose guidance and sustenance have enabled us to embark on this journey and complete the requirements for the Bachelor of Science in Electrical and Electronics Engineering (EEE), including our thesis work. It is through His grace that we have attained the strength and determination to achieve this milestone.

We would like to express our sincere appreciation to our esteemed supervisor, **Salman Fazle Rabby**, Assistant Professor of the Department of Electrical and Electronics Engineering at Sylhet Engineering College. His invaluable academic mentorship, exemplary leadership, positive encouragement, insightful advice, and unwavering support have been instrumental throughout our academic pursuit. We are truly grateful for his guidance. Furthermore, we extend our thanks to our other faculty members, **Md. Shahid Iqbal** and **Md. Ashraful Alam**, for their kindness, support, and cooperation during our academic journey. Their assistance has been greatly appreciated.

Our heartfelt gratitude also goes to our parents, siblings, classmates, and friends in the EEE department at SEC. We acknowledge the sacrifices, prayers, encouragement, and unwavering support they have provided, which have contributed significantly to our success. We are indebted to them for their role in shaping our academic and personal development.

Abstract

Diabetic Retinopathy (DR) is a leading cause of vision loss worldwide, particularly among diabetic patients. Early detection and detailed assessment are essential to prevent irreversible blindness. However, manual grading of DR severity, lesion identification, and vessel analysis are time-consuming, subjective, and dependent on expert availability. This thesis proposes a multi-task deep learning-based framework for comprehensive DR assessment, focusing on three critical tasks: DR severity classification, lesion localization, and retinal vessel segmentation. Each task was addressed independently using optimized deep learning models tailored for medical imaging. For severity classification, a DenseNet121-based CNN was trained on the APTOS 2019 dataset, incorporating several preprocessing steps to address class imbalance . The model achieved a validation accuracy of 93.1%, F1-score of 0.80, and a Quadratic Weighted Kappa (QWK) score of 0.91, demonstrating strong capability in distinguishing DR severity levels. For lesion localization, a pretrained YOLOv8n model that was initially trained on the COCO dataset was fine-tuned using the IDRiD dataset to detect five types of retinal lesions within fundus images. For vessel segmentation, a deep learning-based approach was implemented using a U-Net++ architecture for automated vessel segmentation on fundus images from the DRIVE dataset. A custom U-Net++ model was then trained with Binary Cross-Entropy loss and evaluated using Dice Coefficient, Intersection over Union (IoU), and pixel-wise accuracy. The model achieved a mean Dice score of 0.7966, IoU of 0.6633, and pixel accuracy of 96.25% on the test set. By addressing all three tasks through modular and task-specific deep learning approaches, this work presents a robust pipeline for automated DR assessment. The proposed system demonstrates competitive performance with existing methods and holds strong potential for real-world deployment in clinical ophthalmology to facilitate early diagnosis and intervention.

Keywords: *Diabetic Retinopathy(DR); Deep Learning; Severity Classification; Lesion Localization; Retinal Vessel Segmentation; Fundus Image Analysis.*

Table of Contents

Acknowledgements ..... **iii**

Abstract ..... **iv**

Table of Contents ..... **v**

List of Figures ..... **vii**

List of Tables ..... **viii**

Chapter 1: Introduction **1**

1.1 Overview 1
1.2 Diabetic Retinopathy: Background and Clinical Relevance 1
1.3 Fundus Imaging in Retinal Analysis 2
1.4 Lesions in Diabetic Retinopathy 2
1.4.1 Microaneurysms 2
1.4.2 Optic Disc 3
1.4.3 Hard Exudates 3
1.4.4 Soft Exudates 3
1.4.5 Hemorrhages 4
1.5 Importance of Retinal Vessel Segmentation 4
1.6 Objectives of the Study 5
1.7 Significance of the Study 5
1.8 Scope and Limitations 6
1.9 Summary and Structure 6

Chapter 2: Literature Review **8**

Chapter 3: Materials and Methodology **11**

3.1 Dataset Overview 11
3.1.1 APTOS Dataset 11
3.1.2 IDRiD Dataset 12
3.1.3 DRIVE Dataset 13

3.2 Image Preprocessing 14
3.2.1 Preprocessing for Classification 14
3.2.2 Preprocessing for Localization 16
3.2.3 Preprocessing for Segmentation 16

3.3 Diabetic Retinopathy Classification 17
3.3.1 Model Architecture for Classification (DenseNet121) 17
3.3.2 Severity Grading Criteria 18

3.4 Localization Model Architecture (YOLOv8n) 19
3.5 Retinal Vessel Segmentation Model Architecture (UNet++) 19

3.6 Evaluation Metrics 20

Chapter 4: Results and Discussion 23

4.1 Overview 23
4.2 Diabetic Retinopathy Classification Results Analysis 23
4.3 Fundus Image Lesion Localization Results Analysis..... 26
4.4 Retinal Vessel Segmentation Results Analysis 27

Chapter 5: Conclusion and Future Works 31

5.1 Conclusion 31
5.2 Future Works 31

References 33

List of Figures

Fig. 1.1: A typical fundus image labeled with different anatomical components 3

Fig. 3.1: Different Stages of Diabetic Retinopathy (DR).....12

Fig. 3.2: Annotated Fundus Image from IDRiD Dataset.....13

Fig. 3.3: Retinal Fundus Image and Corresponding Vessel Segmentation Mask.....14

Fig. 3.4: Workflow Diagram of the DR Classification Pipeline.....15

Fig. 3.5: Image Segmentation Data Preparation Workflow.....17

Fig. 3.6: DenseNet121-Based Convolutional Neural Network Architecture for Diabetic Retinopathy Classification.....18

Fig. 3.7: UNet++ Architecture for retinal vessel segmentation from fundus Images.....20

Fig. 4.1: Model Accuracy and Loss during Training an Validation.....24

Fig. 4.2 : Diabetic Retinopathy Classification Results – Sample Predictions vs. Ground Truth.....25

Fig. 4.3: Confusion Matrix of our proposed model tested on 550 images.....26

Fig. 4.4: Lesion Localization Using YOLOv8 on Fundus Image.....27

Fig. 4.5: Training and Validation Dice Coefficient Over Epochs.....28

Fig. 4.6: Training and Validation Accuracy Over Epochs.....28

Fig. 4.7 : Extraction of blood vessels using Proposed algorithm.....29

List of Tables

Table 3.1: Class Distribution of the APTOS 2019 Diabetic Retinopathy Dataset.....11

Table 4.1: Classification Report (Precision, Recall, F1-score) by Class.....24

Table 4.2: Final Segmentation Performance Metrics and Clinical Interpretation.....28

Chapter 1: Introduction

1.1 Overview

One of the most prevalent side effects of diabetes mellitus is diabetic retinopathy (DR), which continues to be a major global cause of blindness and vision impairment. If not identified and treated promptly, the progressive retinal degeneration brought on by DR might result in irreversible visual loss. A prompt and precise diagnosis is essential to avoiding the disease's vision-threatening stages. Ophthalmologists' manual inspection is a major component of the traditional diagnosis of DR, but it is subjective, time-consuming, and vulnerable to inter-observer variability. New possibilities for the automated evaluation of diabetic retinopathy utilizing retinal fundus pictures have been made possible by recent developments in deep learning and medical image processing. This thesis focuses on developing a multi-task deep learning-based framework that performs three critical tasks: severity classification, lesion localization, and retinal vessel segmentation to comprehensively assess diabetic retinopathy.

1.2 Diabetic Retinopathy: Background and Clinical Relevance

Diabetic Retinopathy (DR) is a progressive microvascular complication of diabetes caused by prolonged hyperglycemia, which gradually damages the blood vessels in the retina. It is one of the leading causes of vision impairment and blindness among the working-age population worldwide. Clinically, DR is categorized into the following stages:

- **Mild DR:** Characterized by the presence of microaneurysms, which are small bulges in the retinal blood vessels and represent the earliest visible signs of the disease.
- **Moderate DR:** Involves partial blockage of retinal vessels, leading to impaired blood flow.
- **Severe DR:** Marked by widespread vessel blockage, which deprives areas of the retina of oxygen, signaling the retina to grow new abnormal vessels.
- **Proliferative DR (PDR):** The most advanced stage, where abnormal new blood vessels proliferate on the retina and vitreous. These vessels are fragile and prone to bleeding, increasing the risk of retinal detachment and permanent vision loss.

.

Early detection and timely intervention are critical to halting or slowing disease progression. Retinal screening through fundus imaging, combined with automated image analysis tools, enhances the speed and accuracy of diagnosis [1]. These tools assist ophthalmologists by mapping retinal blood vessels, detecting pathological lesions, and grading DR severity, thereby supporting effective clinical decision-making and reducing the burden of manual assessments.

1.3 Fundus Imaging in Retinal Analysis

The most popular non-invasive imaging method for looking at the retina is fundus photography. It offers color, high-resolution images of the retinal surface that show vital features such the macula, optic disc, blood vessels, and pathological lesions. Diagnosing and treating diabetic retinopathy depends on these pictures [13]. By emphasizing characteristics like microaneurysms, hemorrhages, exudates, and vascular alterations, fundus pictures allow for a thorough examination of retinal disorders. These photos are used by contemporary deep learning algorithms to accurately and automatically identify, categorize, and segment pertinent retinal components. The creation of reliable, data-driven diagnostic tools has been made easier by the availability of enormous databases of fundus images with annotations. The multi-task learning system created in this study uses fundus imaging as its main input. In order to evaluate the severity of diabetic retinopathy, locate lesions, and segment retinal vessels all of which are crucial for a thorough evaluation of the condition it is imperative to be able to discern significant patterns from fundus images.

1.4 Lesions in Diabetic Retinopathy

Lesions are critical indicators of diabetic retinopathy severity. Automated detection and localization of these lesions can significantly aid early diagnosis and disease monitoring. **Fig. 1.1** highlights key lesions retinal fundus image. The key lesions typically observed in DR include:

1.4.1 Microaneurysms

The first obvious symptoms of DR are microaneurysms, which show up in the retina as tiny, spherical red dots. They are generally the initial sign of vascular injury brought on by persistent hyperglycemia and are represented by localized capillary dilation [2]. For the diagnosis of DR in its early stages, microaneurysms must be accurately detected.

1.4.2 The Optic Disc

The natural anatomical structure where the optic nerve leaves the eye is called the optic disc [7]. In retinal imaging, it is a crucial landmark that needs to be accurately recognized to prevent misunderstandings while localizing and segmenting lesions. Reducing false positives in automated systems requires the ability to distinguish between lesions and the optic disc.

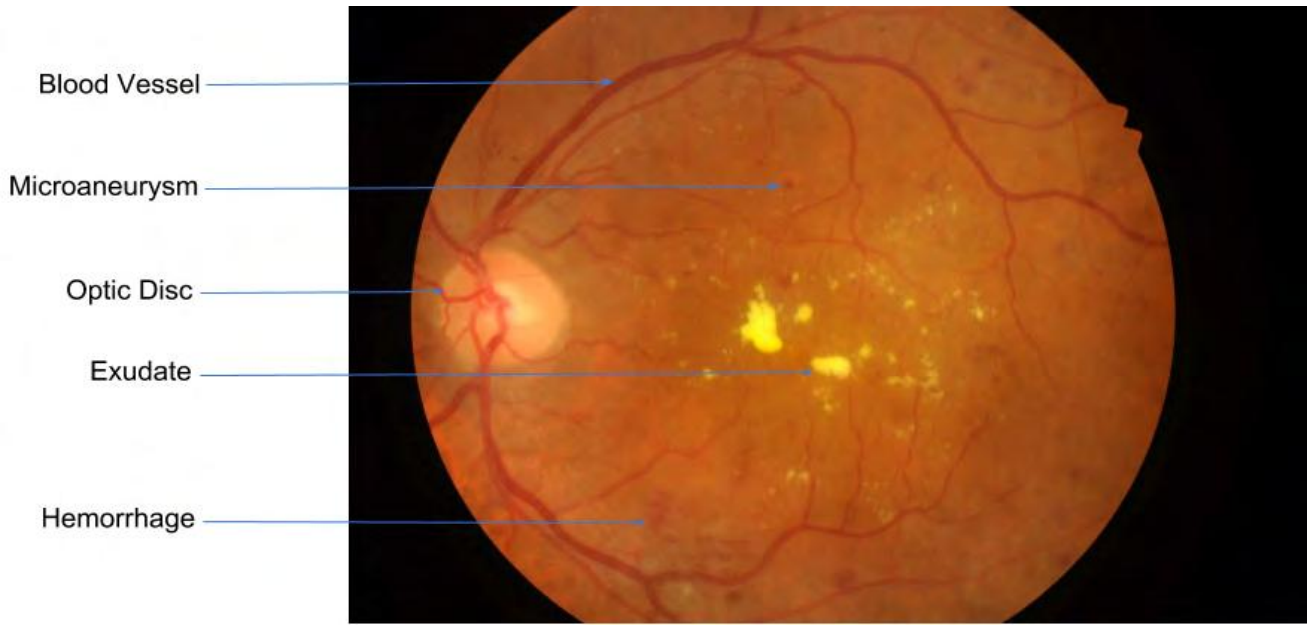


Fig. 1.1: A typical fundus image labeled with different anatomical components.

1.4.3 Hard Exudates

Hard exudates are deposits of lipids and proteins that show up in fundus imaging as yellowish patches with distinct borders. Lipids from damaged retinal blood vessels seep out, causing them to occur. They are frequently linked to macular edema, a dangerous condition that can cause blindness. Assessing the degree of DR and the chance of vision impairment requires the detection of hard exudates.

1.4.4 Soft Exudates

Soft exudates, also referred to as cotton wool spots, are white, fluffy patches that show up on the retina. Localized infarctions of the retinal nerve fiber layer brought on by capillary blockage are the

source of these lesions. More advanced stages of diabetic retinopathy and a reduced retinal blood supply are suggested by the development of soft exudates.

1.4.5 Hemorrhages

Retinal hemorrhages occur when weakened blood vessels rupture, leading to blood leakage into the retinal layers. Hemorrhages can appear in various shapes and sizes and may indicate a progression toward severe or proliferative stages of DR. Accurate detection and localization of hemorrhages are essential for timely treatment to prevent further complications.

1.5 Importance of Retinal Vessel Segmentation

Retinal vessel segmentation plays a vital role in the assessment of diabetic retinopathy (DR) and other ocular pathologies. The morphology of retinal blood vessels such as their width, branching patterns, and tortuosity serves as a key indicator of vascular health and the progression of retinal disease. In DR, pathological changes like neovascularization, vessel narrowing, and abnormal vessel growth are hallmarks of disease advancement. Effective segmentation of retinal vessels enables:

- Precise quantification of vessel width, tortuosity, and branching structure, aiding in the analysis of microvascular changes.

- Detection of proliferative DR, which is marked by neovascularization and abnormal vascular proliferation.

- Improved lesion detection, by isolating vessels from lesions such as microaneurysms, hemorrhages, and exudates.

- Enhanced classification accuracy, as vessel morphology provides contextual information that supports DR grading.

Beyond diabetic retinopathy, retinal vessel analysis is valuable in diagnosing other ocular and systemic conditions such as glaucoma, hypertension, and cardiovascular diseases. The integration of deep learning in vessel segmentation has significantly advanced the automation, speed, and accuracy of this process. Automated segmentation models reduce clinicians' workload while maintaining high diagnostic reliability, making them indispensable in modern ophthalmic screening and diagnostic workflows.

1.6 Objectives of the Study

This study aims to develop a unified multi-task deep learning framework for comprehensive diabetic retinopathy (DR) assessment from retinal fundus images. The specific objectives are outlined as follows:

1. Develop a multi-class DR classification model:

- Accurately categorize fundus images into five DR severity levels: normal, mild, moderate, severe, and proliferative.
- Support early diagnosis and facilitate effective disease monitoring by learning visual patterns specific to each stage.

2. Design an automated lesion localization model:

- Detect and highlight retinal regions affected by DR.
- Specifically identify key lesions such as microaneurysms, hemorrhages, hard exudates, and soft exudates.
- Generate bounding boxes or heatmaps to assist clinicians in locating diseased regions efficiently.

3. Implement a retinal vessel segmentation technique:

- Precisely extract retinal blood vessels from the background and other anatomical structures.
- Support vascular anomaly detection, improve DR severity classification, and enhance lesion localization by providing vascular context.

1.7 Significance of the Study

This research significantly advances automated diabetic retinopathy (DR) evaluation and deep learning-based medical image analysis through a multifaceted approach. Unlike previous studies that often focus on singular aspects of DR detection, this work integrates classification, localization, and segmentation into a unified framework, enhancing diagnostic comprehensiveness. By combining lesion localization, vascular segmentation, and severity grading, the proposed system delivers interpretable, clinically relevant outputs that aid ophthalmologists in diagnosis and treatment planning, thereby improving clinical utility. The adoption of multi-task learning leverages shared features across tasks to boost diagnostic accuracy and model robustness while optimizing resource efficiency by reducing the need for separate models, thus lowering computational costs and training

time. The system's scalability makes it ideal for large-scale DR screening programs, particularly in underserved regions with limited access to specialized care. Furthermore, this study lays the groundwork for future multi-task learning applications in medical imaging, fostering innovation in integrated diagnostic methodologies and aligning technical advancements with practical clinical needs.

1.8 Scope and Limitations

This research focuses on developing deep learning models using publicly accessible diabetic retinopathy datasets of retinal fundus images, targeting multi-class classification of DR severity levels, localization of critical pathological lesions (e.g., microaneurysms, hemorrhages, hard and soft exudates, segmentation of retinal vasculature for detailed vascular analysis, and exploration of multi-task learning to integrate these components, with model performance evaluated via metrics like accuracy, precision, recall, Intersection over Union (IoU), Dice coefficient, and localization accuracy, using Python-based libraries such as TensorFlow or PyTorch; however, limitations include reliance on dataset quality and diversity, which may introduce biases or limit generalizability, challenges with low-resolution or low-contrast images reducing lesion detection and segmentation precision, lack of real-world clinical validation across diverse patient groups, high computational demands restricting deployment on resource-constrained devices, and limited model interpretability in clinical settings, with future work aiming to address these by incorporating diverse datasets, improving generalizability, and collaborating with healthcare institutions for clinical validation.

1.9 Summary and Structure

Diabetic Retinopathy (DR) is a leading cause of blindness caused by damage to retinal blood vessels due to diabetes. Early detection is crucial but manual diagnosis is time-consuming and prone to errors. This study proposes a multi-task deep learning framework that combines DR severity classification, lesion localization, and retinal vessel segmentation using fundus images. The approach aims to improve diagnostic accuracy, support clinical decision-making, and enable scalable, automated DR screening. Challenges include dataset limitations and computational demands, but the system shows promise for future clinical applications.

Following chapters outlines the methodology, results and finally conclusion for developing a unified multi-task deep learning framework for diabetic retinopathy (DR) assessment using retinal fundus images. Three publicly available datasets were used: APTOS 2019 for DR severity classification,

IDRiD for lesion localization, and DRIVE for retinal vessel segmentation. Preprocessing steps varied by task, including cropping, resizing, contrast enhancement (Ben Graham method), and data augmentation techniques such as Mixup and real-time image transformations. The framework employs DenseNet121 for classification, YOLOv8n for lesion localization, and UNet++ for vessel segmentation. Each model is trained and evaluated using appropriate metrics: Precision, Recall, F1-score, and Confusion Matrix for classification; mAP@0.5 for localization; and Dice Coefficient, IoU, and Pixel Accuracy for segmentation. This integrated approach aims to enhance diagnostic accuracy, provide clinical interpretability, and streamline DR screening through automation.

Chapter 2: Literature Review

Diabetic retinopathy (DR) remains one of the leading causes of vision impairment worldwide, and early detection plays a crucial role in preventing blindness. In recent years, researchers have extensively investigated both traditional machine learning and advanced deep learning approaches to improve DR detection, lesion localization, and severity grading. The following section reviews key contributions in this domain, highlighting different methodologies, datasets, and outcomes.

Ali et al. (2020) proposed a hybrid machine learning framework for DR severity classification, combining clustering-based region growing with hybrid feature extraction techniques, including histograms, co-occurrence matrices, and wavelets. By applying Fisher score and mutual information for dimensionality reduction, and using classifiers like MLP and logistic regression, the model attained higher accuracy on a local dataset of 2500 images, showcasing robust performance across all DR stages.

Santos et al. (2022) introduced a YOLO-based deep learning framework for detecting diabetic retinopathy lesions in fundus images, underscoring the critical role of preprocessing and augmentation in addressing low contrast and small lesion sizes. Their approach incorporated background cropping, image tiling, and contrast enhancement to enhance lesion visibility. Trained on the DDR and IDRiD datasets, the model delivered competitive performance, effectively identifying small lesions despite dataset limitations. This work highlights the efficacy of combining advanced preprocessing with real-time object detection systems like YOLO to advance diabetic retinopathy analysis.

Challa et al. (2019), published in the PReMI 2019 proceedings, proposed a deep All-CNN architecture for multi-class diabetic retinopathy (DR) detection. Their pipeline employs Gaussian filtering and retinal boundary removal to emphasize key retinal features, followed by processing through a 10-layer convolutional network culminating in a softmax classifier. Trained on 30,000 fundus images and evaluated on 3,000, the model achieved an accuracy of 86.6%, a loss of 0.46, and an average F1-score of 0.63, surpassing several existing methods. This study highlights the effectiveness of tailored preprocessing combined with deep learning for improving multi-class DR classification performance.

Sarki et al. (2021) developed a convolutional neural network (CNN)-based framework for multi-class classification of diabetic eye diseases using retinal fundus images. The model achieved an overall accuracy of 81.33%, effectively distinguishing diabetic retinopathy stages and other ocular conditions. By incorporating contrast enhancement and data augmentation to address limited dataset sizes, along with transfer learning from pretrained models for improved feature extraction, the study demonstrated robust performance. This work emphasizes the critical role of image preprocessing and meticulous model tuning in achieving reliable classification across diverse retinal diseases.

Aswathi et al. (2021) developed a deep learning-based system for diabetic retinopathy (DR) severity grading using transfer learning on the Messidor dataset. They evaluated multiple pre-trained CNN architectures, including VGG19, InceptionV3, ResNet50, MobileNet, and NASNet, for binary and multi-class classification tasks. Preprocessing involved CLAHE and power-law transformation to enhance image quality. The InceptionV3 model, combined with CLAHE and optimized hyperparameters, achieved the best performance, with a validation accuracy of 58% for multi-class classification and 78% for binary classification (mild vs. normal DR). The study emphasized the challenges posed by inter-class similarity and the critical role of extensive fine-tuning in improving classification accuracy.

Pratt et al. (2016) introduced a deep learning-based convolutional neural network (CNN) for five-class diabetic retinopathy (DR) severity classification using color fundus images. Their approach automatically detected key retinal features, such as microaneurysms, exudates, and hemorrhages, without manual feature extraction. Trained on a large Kaggle dataset of 80,000 images, the model utilized advanced preprocessing, including color normalization and data augmentation. It achieved a validation accuracy of 75% and high specificity, demonstrating potential for real-time DR screening. This work highlights the capability of CNNs to automate DR grading effectively with large-scale datasets.

Dutta et al. (2018) employed deep learning models, including CNNs and DNNs, for classifying DR stages from fundus images. CNNs excelled in identifying critical retinal features such as hemorrhages, exudates, and micro-aneurysms, outperforming traditional back-propagation networks. The model's high accuracy in severity prediction highlights the advantage of automated feature learning over handcrafted-feature methods in DR classification tasks.

Parmar et al. (cited in Ali et al., 2020) utilized neural networks to classify diabetic retinopathy into five severity levels based on fundus images, achieving approximately 85% accuracy. The study

highlighted the potential for enhancing performance through deep learning and hybrid feature integration. It provided a foundational framework for multi-class DR severity prediction, underscoring the critical role of effective data preparation and feature extraction in achieving reliable classification outcomes.

Aujih et al. (2018) explored the impact of retinal vessel segmentation on DR classification using U-Net and InceptionV1 models. U-Net was employed to segment retinal vessels, and their presence or absence was analyzed for its effect on DR severity classification. The findings showed that vessel information significantly enhances detection in advanced DR stages but offers limited benefits in early stages, emphasizing the role of vessel morphology in DR diagnosis.

Santos et al. (2022) introduced a YOLO-based deep learning framework for detecting diabetic retinopathy lesions in fundus images, underscoring the critical role of preprocessing and augmentation in addressing low contrast and small lesion sizes. Their approach incorporated background cropping, image tiling, and contrast enhancement to enhance lesion visibility. Trained on the DDR and IDRiD datasets, the model delivered competitive performance, effectively identifying small lesions despite dataset limitations. This work highlights the efficacy of combining advanced preprocessing with real-time object detection systems like YOLO to advance diabetic retinopathy analysis. So, the summary at a glance -

- CNN-based and U-Net-inspired architectures have consistently proven effective for DR classification and lesion segmentation, validating their use in this study's proposed U-Net++ model.
- Preprocessing techniques such as CLAHE, normalization, background removal, and contrast enhancement are crucial across studies for improving classification and segmentation outcomes.
- Hybrid frameworks and YOLO-based lesion detection approaches demonstrate the benefits of combining preprocessing, segmentation, and classification for more comprehensive DR analysis.

Research gaps include the tendency of most existing studies to address classification, segmentation, or lesion detection independently rather than through an integrated framework, which limits their capacity for comprehensive and robust severity prediction. Another gap lies in the high inter-class similarity between DR stages, which makes accurate grading particularly challenging. Additionally, limited dataset diversity and imbalance reduce the generalizability of many proposed models, underscoring the need for more fine-tuned architectures, advanced augmentation strategies, and frameworks capable of adapting effectively to real-world clinical scenarios.

Chapter 3: Materials and Methodology

3.1 Dataset Overview

Datasets were selected to align with the requirements of each task. The APTOS 2019 Blindness Detection dataset was used for DR severity classification, providing labeled fundus images across five severity levels. The IDRiD dataset, formatted for YOLO, was employed for lesion localization, with bounding box annotations for microaneurysms, hemorrhages, hard/soft exudates, and optic disc. The DRIVE dataset was chosen for vessel segmentation, offering high-resolution images with expert-annotated binary vessel masks. Datasets were retrieved via the Kaggle API, organized into structured directories, and verified for integrity (image-label pairings). Subsections detail directory structures, data inspection, and exploratory analyses to confirm label distributions and file consistency.

Table 3.1: Class Distribution of the APTOS 2019 Diabetic Retinopathy Dataset

Label	DR-Severity Level	Frequency
0	No DR	1800 images
1	Mild DR	350 images
2	Moderate DR	1000 images
3	Severe DR	250 images
4	Proliferative DR	300 images

3.1.1 APTOS Dataset

The APTOS 2019 Blindness Detection dataset is a publicly available collection of retinal fundus images curated for the task of diabetic retinopathy (DR) severity classification, a critical challenge given DR's role as a leading cause of vision loss globally as showed in **Fig. 3.1** . The dataset contains 3,662 high-resolution images, divided into training and validation sets. Each training image is labeled on a five-point severity scale ranging from 0 (No DR) to 4 (Proliferative DR). Notably, the **Table 3.1** exhibits a significant class imbalance, with 1,800 images labeled as No DR, followed by 1000 Moderate, 350 Mild, 250 Severe, and 300 Proliferative DR cases. Furthermore, the dataset includes considerable variability in image quality, illumination conditions, and the presence of

artifacts, which necessitates robust preprocessing techniques to ensure reliable model performance. These characteristics make the APTOS 2019 dataset a valuable benchmark for developing and evaluating automated DR grading systems based on deep learning.

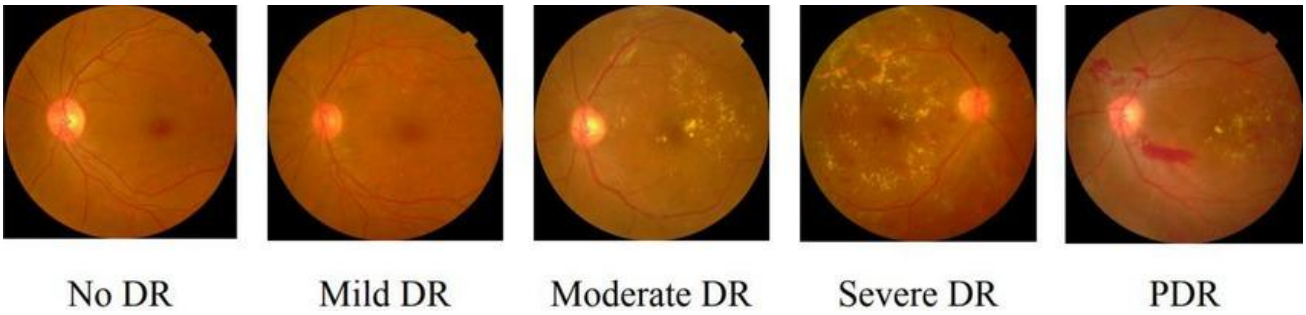


Fig. 3.1: Different Stages of Diabetic Retinopathy (DR)

3.1.2 IDRiD Dataset

The Indian Diabetic Retinopathy Image Dataset (IDRiD) is a specialized and comprehensive dataset curated for diabetic retinopathy (DR) research, particularly suited for the development and evaluation of deep learning models such as YOLO. It supports multiple critical tasks including lesion detection, severity classification, and retinal vessel segmentation. The dataset comprises 81 high-resolution retinal fundus images (4288×2848 pixels, JPG format), captured using a Kowa VX-10a digital fundus camera with a 50° field of view, and highlights various signs of DR. The dataset is divided into 54 training and 27 validation images, each accompanied by TXT annotation files that mark key DR-related lesions—microaneurysms, hemorrhages, hard exudates, and soft exudates—as bounding boxes or pixel-level masks, making it directly compatible with YOLO's object detection framework. Additionally, CSV files provide ground-truth labels for both DR severity and diabetic macular edema based on international clinical grading standards. The dataset also includes optic disc and fovea center coordinates to facilitate anatomical localization. **Fig. 3.2** shows a fundus image with annoted precise position of lesions present .

Hosted on IEEE DataPort and introduced through the ISBI-2018 Diabetic Retinopathy Segmentation and Grading Challenge, IDRiD is the first publicly available dataset representing an Indian population. With its detailed pixel-level annotations and clinical richness, it serves as a valuable benchmark for training and validating advanced computer-aided diagnostic systems aimed at the early detection and analysis of DR.

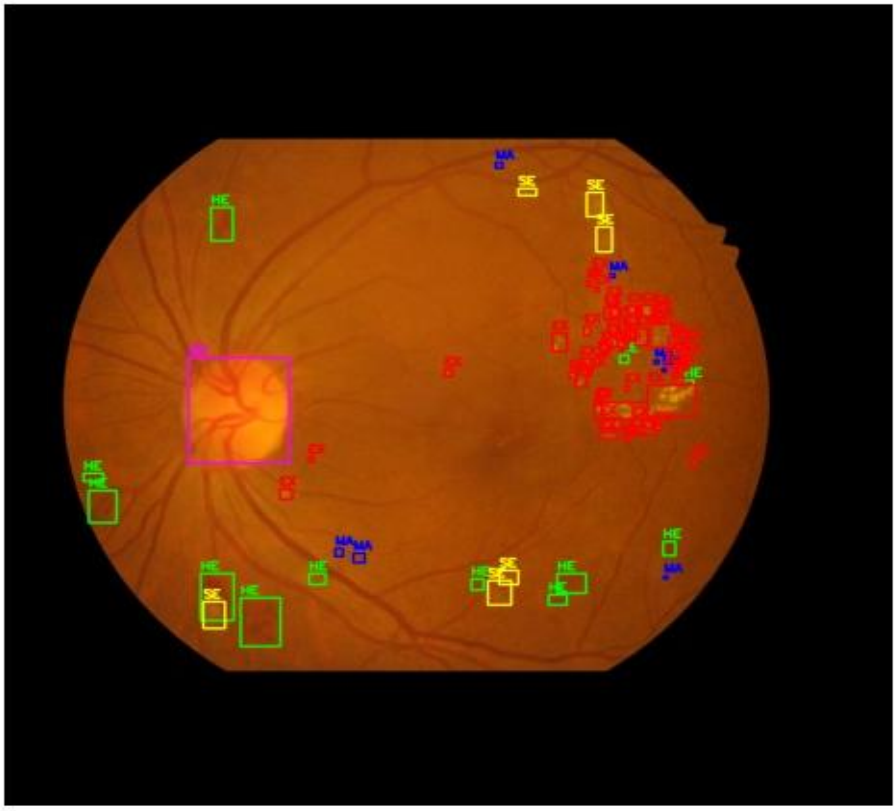


Fig. 3.2: Annotated Fundus Image from IDRiD Dataset

3.1.3 DRIVE Dataset

The DRIVE (Digital Retinal Images for Vessel Extraction) as shown in **Fig. 3.3** dataset is a standard benchmark for retinal vessel segmentation, comprising 40 high-resolution fundus images captured under uniform conditions with a non-mydriatic camera (45° field of view). Each image shows a circular retinal region, typically processed using the green channel for better vessel contrast.

The dataset is evenly divided into 20 training and 20 testing images, each paired with expert-annotated binary vessel masks and field-of-view (FOV) masks that isolate the retinal area. This structured format supports supervised learning and reliable evaluation using metrics such as the Dice coefficient, IoU, and pixel-wise accuracy, all computed within the FOV.

Despite its small size, DRIVE's consistent imaging, clinical relevance, and high-quality annotations make it a trusted resource for developing and benchmarking vessel segmentation models.

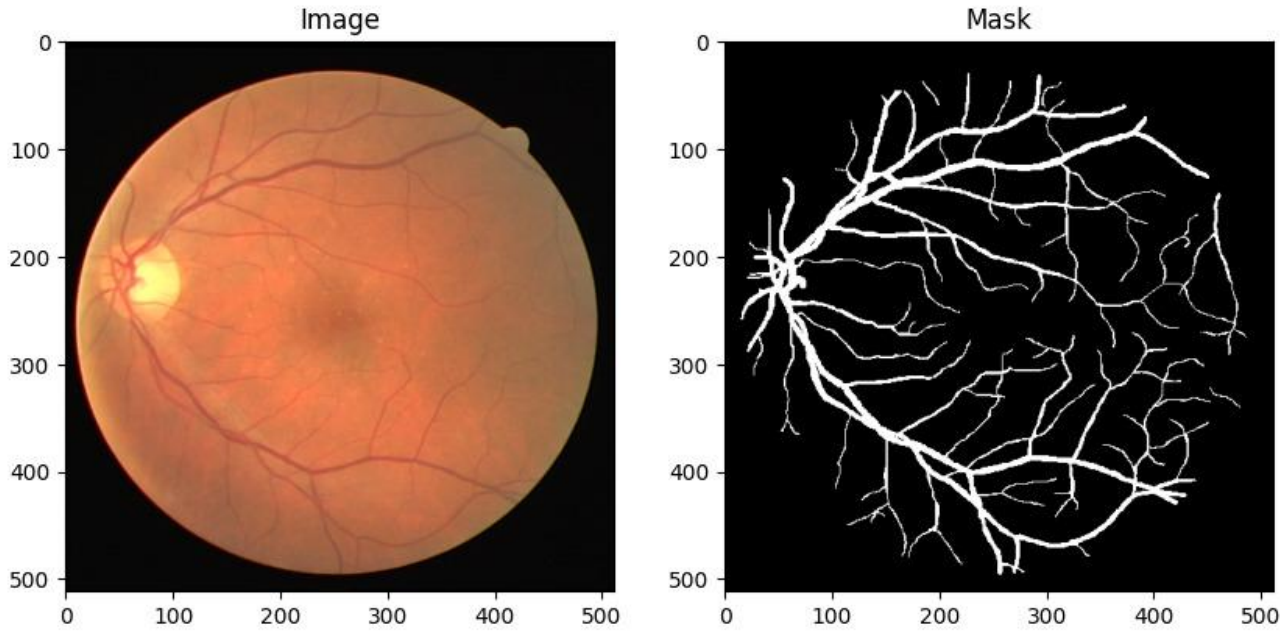


Fig. 3.3: Retinal Fundus Image and Corresponding Vessel Segmentation Mask

3.2 Image Preprocessing

Task-specific preprocessing pipelines were designed to meet the unique input needs of each task, adhering to best practices. Images were read, converted to NumPy arrays, and normalized. Preprocessing included cropping dark borders for classification, resizing to uniform dimensions for classification and segmentation, normalizing pixel intensities to [0,1], enhancing contrast for classification, thresholding masks for segmentation, and verifying annotation formats for localization. Augmentation varied by task: classification used Keras and custom mixup; localization leveraged YOLO's built-in augmentations; segmentation applied synchronized image-mask augmentations via Keras.Subsections detail each pipeline.

3.2.1 Preprocessing for Classification

To ensure optimal model performance and robustness, the preprocessing pipeline shown in **Fig. 3.4** for the APTOS 2019 retinal fundus dataset includes several critical steps designed to handle class imbalance, reduce noise, and enhance the discriminative features of retinal images. First, each fundus image is read using OpenCV and converted from BGR to RGB, ensuring correct color channel interpretation. Next, the images undergo automatic cropping using brightness thresholding, where dark background regions are removed by masking pixels with intensity values below a given threshold (controlled via the crop_image_from_gray function). This isolates the informative retinal area and discards non-informative dark borders.

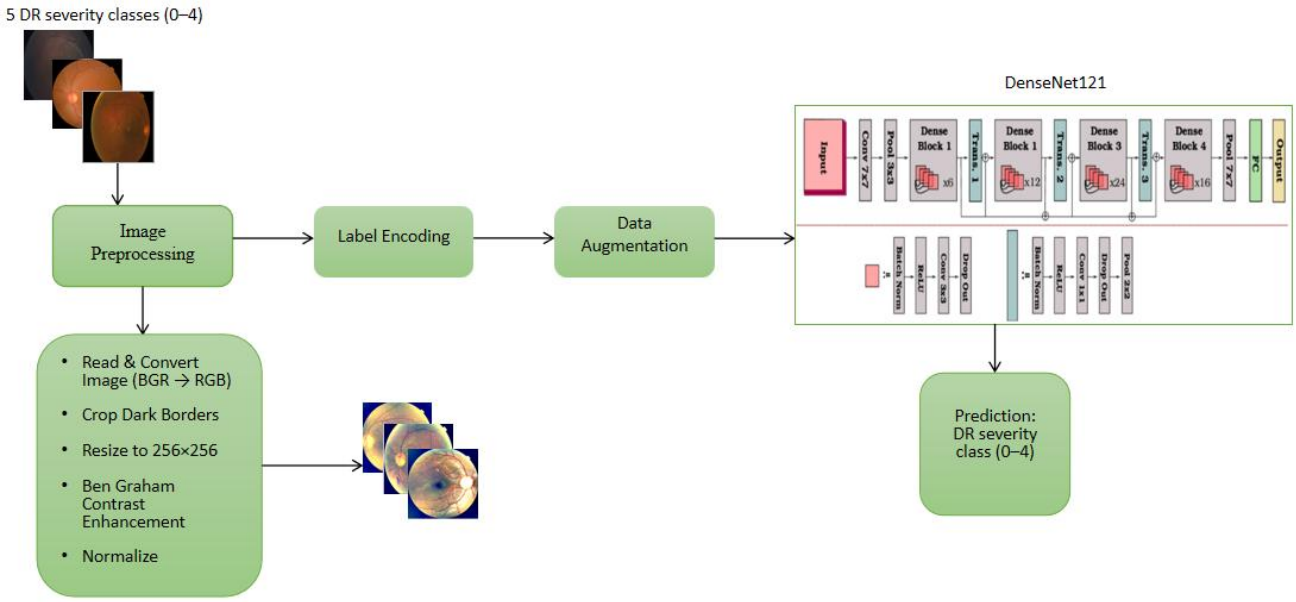


Fig. 3.4: Workflow Diagram of the DR Classification Pipeline

After cropping, images are resized to 256×256 pixels, ensuring uniform input size across the dataset, which is essential for efficient batch training in deep learning models. To enhance lesion visibility, the Ben Graham preprocessing technique is applied: a weighted subtraction of a Gaussian-blurred version of the image from the original is used (cv2.addWeighted). This method sharpens blood vessels and lesion regions while suppressing background variations, improving contrast and feature localization. To address class imbalance, the target labels are converted from single-class to multi-label format by propagating severe labels down the scale (e.g., if a sample is labeled as class 4, it is also considered positive for classes 0–3). This softens the boundary between severity levels and aids learning in skewed datasets.

Furthermore, a Mixup data augmentation strategy is integrated into training. Mixup linearly interpolates pairs of training examples and their labels, generating synthetic samples that regularize the model and improve generalization. This is combined with traditional augmentations like random zooms, horizontal/vertical flips, and constant fill padding, using ImageDataGenerator. The use of the green channel during preprocessing (implicitly done in contrast enhancement) is also a well-established practice in retinal imaging, as vessels and lesions exhibit higher contrast in that channel.These preprocessing techniques collectively contribute to more stable training, increased robustness to imaging artifacts, and enhanced sensitivity to subtle pathological features within retinal fundus images.

3.2.2 Preprocessing for Localization

To prepare the dataset for lesion localization in diabetic retinopathy using the YOLOv8n model, a comprehensive set of preprocessing steps was performed. Initially, the dataset was retrieved from Kaggle and unzipped into structured directories containing training and validation images along with their corresponding YOLO-format annotation files. The images were resized and filtered using Gaussian blurring to reduce noise and enhance lesion visibility, as this step preserves edge information while suppressing background variation. Following this, contrast enhancement techniques such as CLAHE (Contrast Limited Adaptive Histogram Equalization) were implicitly applied during training augmentation to improve the visibility of microaneurysms, hemorrhages, exudates, and optic discs. Label files were formatted according to the YOLO object detection standard, where each annotation consists of the class index and normalized bounding box coordinates. A custom visualization function was created to overlay bounding boxes and labels on images for inspection. During training, Ultralytics' YOLOv8 framework applied additional real-time augmentations such as random flipping, scaling, and color-space transformations to improve model robustness and generalization. Albumentations library was internally used to add controlled variations like grayscale conversion, blur, and adaptive histogram equalization. All images were rescaled to a fixed resolution (1024×1024) to match YOLOv8's input requirement and allow for consistent feature extraction across samples. The resulting image-label pairs were cached for faster I/O during training.

3.2.3 Preprocessing for Segmentation

Fig. 3.5 shows the retinal vessel segmentation pipeline, the preprocessing begins by loading images and corresponding vessel masks from the DRIVE dataset directories. The images are read using PIL.Image, and masks (in .gif format) are carefully aligned with their corresponding images. Each image and mask pair is resized to a standard resolution of 512×512 pixels to ensure uniformity in model input size. The pixel values of the images are then normalized to the [0, 1] range by dividing by 255, converting them to float32 for compatibility with neural networks. For the masks, thresholding is applied (with a cutoff value of 70) to convert grayscale mask images into binary masks, separating vessels from the background clearly.

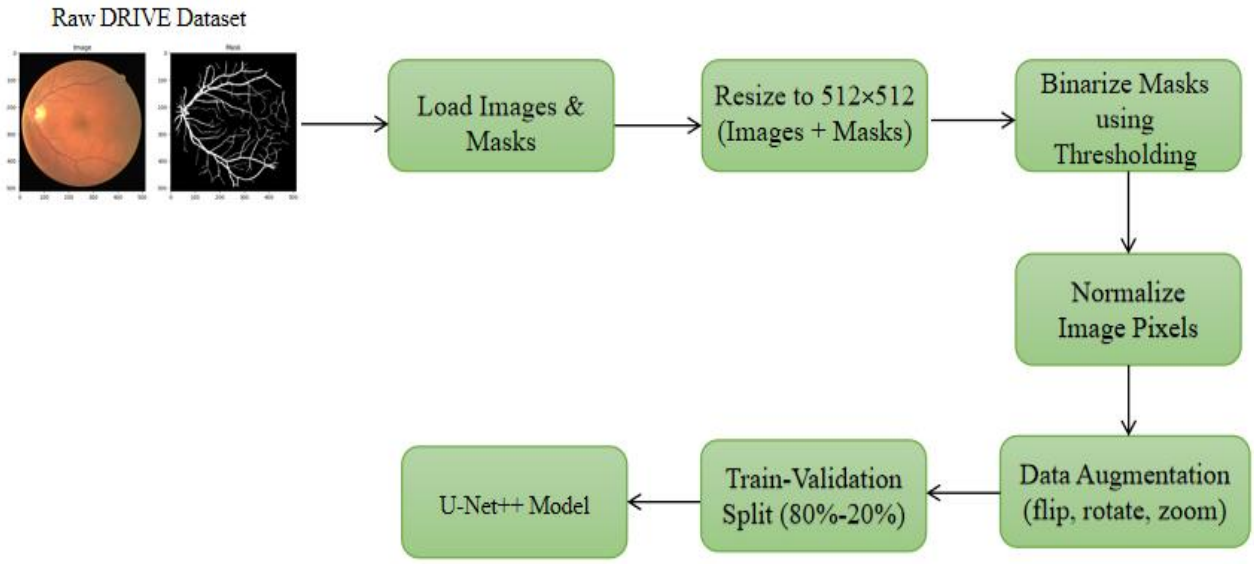


Fig. 3.5: Image Segmentation Data Preparation Workflow

Following resizing and binarization, data augmentation is applied to expand the dataset and improve generalization. Augmentation includes rotation, width/height shift, zooming, and both horizontal and vertical flips, ensuring that the model becomes robust to various orientations and transformations of the retinal structures. Importantly, the same random seed is used for both image and mask generators to maintain synchronization between augmented images and masks. Finally, all processed images and masks are stored in NumPy arrays, reshaped as necessary for use in a deep learning model.

3.3 Diabetic Retinopathy Classification

This section outlines the design, implementation, and training of the DR severity classification model, focusing on architecture, label encoding, and training procedures.

3.3.1 Model Architecture for Classification (DenseNet121)

The DenseNet121-based Convolutional Neural Network (CNN) architecture is a powerful and efficient deep learning model used for Diabetic Retinopathy (DR) classification shown in **Fig. 3.6**. DenseNet121, a variant of the Dense Convolutional Network family, connects each layer to every other layer in a feed-forward fashion, promoting feature reuse and efficient gradient flow, which helps mitigate the vanishing gradient problem in deep networks. In the context of DR classification, the DenseNet121 model is typically initialized with ImageNet pre-trained weights and fine-tuned on

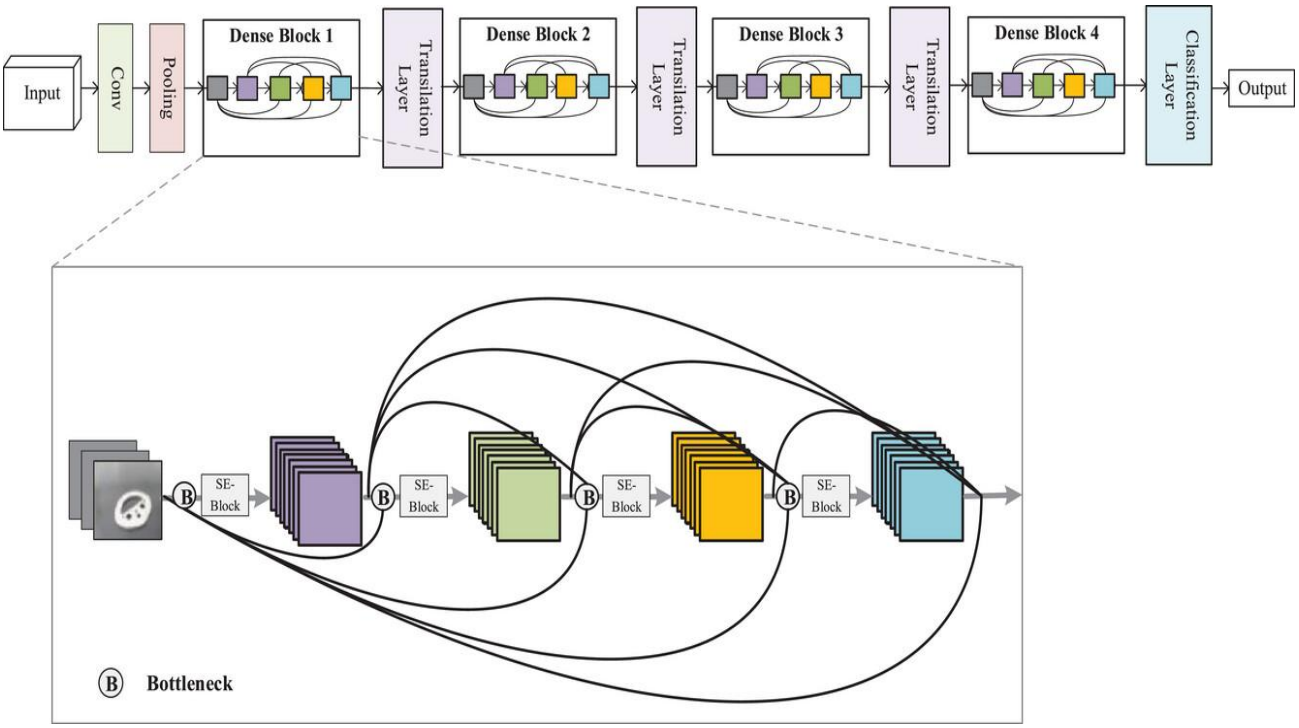


Fig. 3.6: DenseNet121-Based Convolutional Neural Network Architecture for Diabetic Retinopathy Classification (Natesan et al. 2020)

retinal fundus images to learn disease-relevant features. Its deep feature extraction capability enables the detection of subtle pathological signs such as microaneurysms, hemorrhages, and exudates across different severity levels of DR. The model's dense connectivity and compact architecture make it well-suited for medical image analysis, offering high classification accuracy even with limited training data. Additionally, integrating DenseNet121 with techniques like data augmentation and class balancing can further enhance its performance in real-world clinical DR screening applications.

3.3.2 Severity Grading Criteria

DR severity is categorized into five classes: 0 (No DR), 1 (Mild: microaneurysms only), 2 (Moderate: additional lesions but not severe), 3 (Severe: extensive hemorrhages/microaneurysms, venous beading), and 4 (Proliferative: neovascularization, vitreous hemorrhage). Labels are encoded as cumulative multilabel vectors to reflect ordinal progression, where class k sets outputs 0 to k as 1, enhancing the model's ability to learn severity hierarchies.

3.4 Localization Model Architecture (YOLOv8n)

YOLOv8n (You Only Look Once version 8 – Nano) is the lightest variant of the YOLOv8 object detection architecture, designed for real-time inference on resource-constrained devices. It features a fully convolutional architecture with a streamlined backbone and head for efficient feature extraction and prediction. The backbone employs CSP (Cross Stage Partial) connections and convolutional layers to extract multi-scale features while reducing computational cost. The neck uses a PAN (Path Aggregation Network) structure to fuse features from different scales, enhancing spatial information crucial for small object detection. The detection head outputs bounding box coordinates, objectness scores, and class probabilities directly, making it an anchor-free design, which simplifies training and improves adaptability across datasets. YOLOv8n integrates batch normalization and SiLU (Sigmoid Linear Unit) activation for better convergence. Despite its compact size, YOLOv8n maintains competitive accuracy, making it suitable for real-time applications like lesion localization in medical imaging.

3.5 Retinal Vessel Segmentation Model Architecture(UNet++)

UNet++ is an enhanced version of the traditional U-Net architecture (**Fig. 3.7**), tailored to improve segmentation accuracy by addressing the semantic gap between encoder and decoder feature maps. It introduces nested and dense skip connections, where intermediate convolution layers—called dense convolution blocks—are inserted along the skip pathways. These blocks refine the features before passing them to the decoder, resulting in semantically richer and more aligned representations. This is especially useful in retinal vessel segmentation, where capturing fine vascular details is critical. In addition to its redesigned skip pathways, UNet++ employs deep supervision, enabling the network to generate segmentation outputs at multiple semantic levels. This technique improves gradient flow during training by adding auxiliary loss functions at intermediate layers. Two supervision modes are supported: accurate (averaging all outputs) and fast (using a single output for reduced complexity). The training objective typically combines binary cross-entropy with the Dice coefficient, ensuring both pixel-level accuracy and region-level overlap. Together, these enhancements make UNet++ highly effective for segmenting thin and complex vessel structures in retinal images.

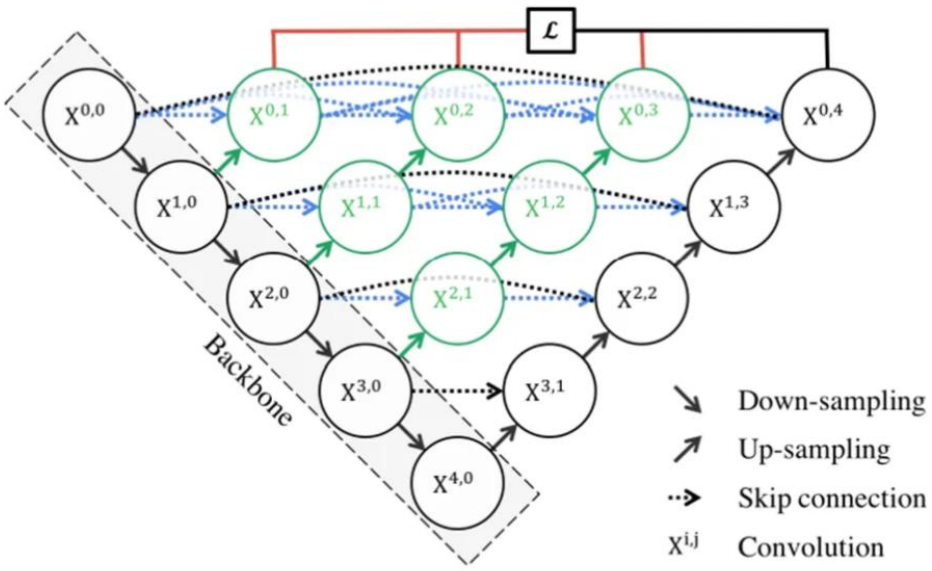


Fig. 3.7: UNet++ Architecture for retinal vessel segmentation from fundus images. (Tsang, 2021)

3.6 Evaluation Metrics

To evaluate the performance of the deep learning models developed for diabetic retinopathy (DR) analysis, task-specific metrics were applied to the classification, localization, and segmentation tasks. For classification, precision, recall, and F1-score were utilized to assess the model's ability to accurately distinguish DR severity levels, effectively reducing false positives and negatives. In lesion localization, mean Average Precision at an IoU threshold of 0.5 (mAP50) was the primary metric, evaluating both detection accuracy and spatial precision of localized lesions. For retinal vessel segmentation, the Dice Coefficient served as the main metric, measuring the overlap between predicted vessel regions and ground truth masks. These tailored metrics collectively provide a comprehensive and quantitative assessment of each model's performance across diverse clinical objectives.

Across all three tasks, the core foundation of evaluation begins with the confusion matrix, which comprises True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Key definitions include:

- **True Positive (TP)**: Correctly identifying a specific DR class (e.g., labeling 'Severe DR' correctly).

- **False Positive (FP)**: Incorrectly predicting a class (e.g., labeling 'Moderate DR' as 'Severe DR').
- **False Negative (FN)**: Failing to identify a class (e.g., missing 'Severe DR' by predicting 'Moderate DR').
- **True Negative (TN)**: Correctly identifying non-class instances.

Precision quantifies the proportion of correct positive predictions, making it a critical metric when false positives carry significant consequences, such as incorrectly diagnosing healthy eyes as having diabetic retinopathy (DR). It indicates the reliability of positive classifications by measuring the fraction of predicted positive cases that are truly positive.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.1}$$

Recall, also known as sensitivity, measures the proportion of actual positive cases correctly identified by the model. In diabetic retinopathy (DR) screening, high recall is vital to minimize false negatives, as missing a true DR case could delay critical treatment, potentially worsening patient outcomes.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3.2}$$

The F1-Score, calculated as the harmonic mean of precision and recall, offers a single metric that balances these two measures. It is particularly valuable in diabetic retinopathy (DR) classification,where imbalanced datasets are common, ensuring a comprehensive evaluation of model performance by accounting for both false positives and false negatives.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.3}$$

For lesion localization, mean Average Precision at IoU threshold 0.5 (mAP@0.5) is the primary metric, averaging precision across lesion classes, with predictions correct if IoU with ground truth is ≥0.5. Defined as

$$\text{mAP}_{50} = \frac{1}{N} \sum_{i=1}^{N} \text{AP}_i \tag{3.4}$$

where N is the number of classes and i is the average precision for class (i) from the precision-recall curve, mAP@0.5 balances detection confidence and localization accuracy, ensuring robust lesion identification and positioning.

In retinal vessel segmentation, the Dice Coefficient quantifies the overlap between predicted and ground truth masks, assessing the accuracy of positive predictions. It is particularly sensitive to class imbalance, making it ideal for evaluating segmentation tasks where vessel pixels are significantly outnumbered by background pixels, ensuring a robust measure of model performance in delineating vascular structures.

$$\text{Dice Coefficient} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{3.5}$$

Ranging from 0 (no overlap) to 1 (perfect overlap), it is sensitive to both false positives and negatives, making it ideal for imbalanced datasets where vessels occupy a small image area.

Intersection over Union (IoU), also called the Jaccard Index, measures the ratio of the overlap area to the combined area of prediction and ground truth, providing a robust indicator of segmentation accuracy—particularly when assessing how well the model captures fine structures like retinal vessels.

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{3.6}$$

Additionally, pixel accuracy calculates the proportion of correctly classified pixels over all pixels in an image, offering a straightforward yet valuable measure of how well the segmentation model labels each pixel.

$$\text{Pixel Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{3.7}$$

Chapter 4: Results and Discussions

4.1 Overview

This sect presents an integrated framework for diabetic retinopathy (DR) analysis that combines classification, lesion detection, and vessel segmentation. Instead of focusing solely on final results, the study emphasizes a deeper understanding of model behavior by examining training progress, misclassification patterns, and challenges like class imbalance, image noise, and subtle lesion differences. Advanced techniques such as Mixup augmentation, YOLOv8 detection, and segmentation networks are employed to tackle the diverse aspects of DR diagnosis. Visual and quantitative analyses, including confusion matrices and loss trends, provide insights into the system's response to real-world variability. Overall, the study highlights important factors influencing model effectiveness and suggests directions for refinement.

4.2 Diabetic Retinopathy Classification Result Analysis

In this section, the performance of our model is analyzed on the train test and validation data. In addition, some misclassified images along with the possible reasons behind misclassification will also be discussed. Since this is a 5-class classification problem, accuracy of 93.1% is quite promising. In general, human performance for successful DR grading is around 75%. Most of these misclassifications are caused by camera artifacts, poor image quality, wrong actual labels and class imbalance. The Diabetic Retinopathy (DR) classification model demonstrates strong overall performance, achieving a validation QWK (Quadratic Weighted Kappa) score of 0.914, indicating high agreement with true labels despite class imbalance. The best threshold for probability-to-class conversion was found to be 0.4906, optimizing classification boundaries. The model's overall macro-averaged accuracy is 82%, with a precision of 82%, recall of 79.6%, and F1-score of 80%, suggesting a balanced performance across most classes but with noticeable variation between them. From the accuracy and loss curves in **Fig. 4.1**, the validation accuracy steadily increases, peaking above 94%, even when training accuracy exhibits oscillations likely due to aggressive augmentation and regularization techniques like Mixup. The validation loss also stabilizes early on, reflecting the model's generalization capability. Precision-recall tradeoffs show that while class 0 and class 2 are

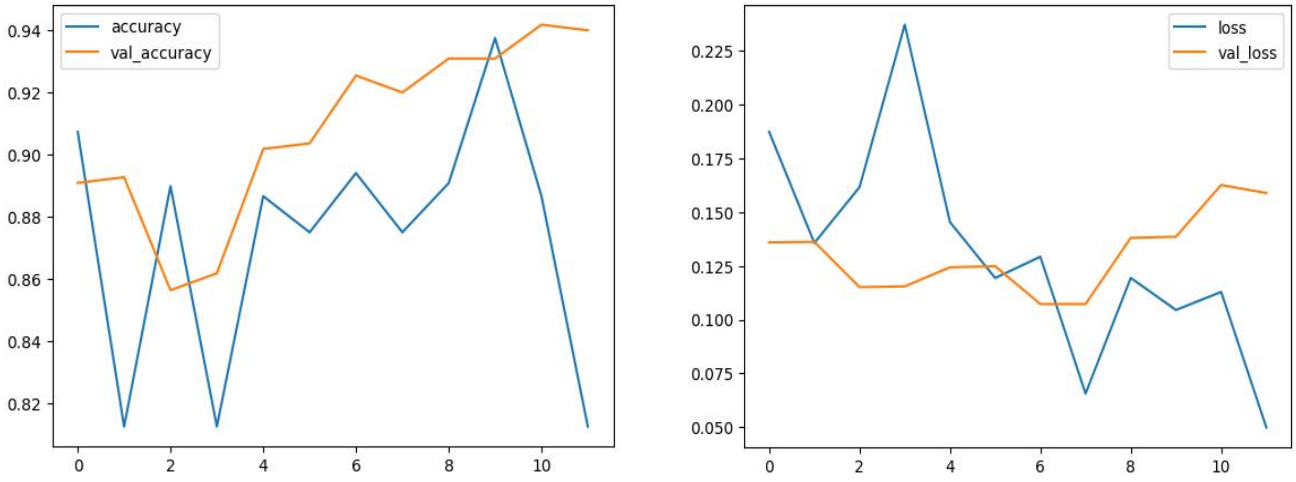


Fig. 4.1: Model Accuracy and Loss during Training and Validation

robust, class 1(Mild) and class 3 have lower precision and recall. This suggests that data imbalance and subtle lesion differences might still be limiting performance on intermediate and severe cases.

Table 4.1: Classification Report (Precision, Recall, F1-score) by Class

Level	precision	recall	f1-score
0	.98	.98	.98
1	.78	.82	.80
2	.78	.76	.78
3	.76	.72	.75
4	.80	.70	.73

The model's ability to correctly identify class-0 images is quite remarkable which is re-flected by a the classification report **Table 4.1** . Only 11 images belonging to class-0 was wrongly misclassified as different images according to **Fig. 4.3**. However, 6 images of class 1 are wrongly classified . This is because in general, there is very little visual difference between a class-0 image and class-1 image. The first stage of DR (class-1) usually has very few microaneurysms which are very hard to identify even for specialists. These misclassifications may have occurred because of downsampling the

original image, unreliable actual grading and due to the noise present in the images. The overall performance metrics of our model are greatly affected due to these misclassifications. The effect of classifying moderate/severe nonproliferative DR or proliferative DR as noDR is much more severe than the above-mentioned cases. Our model's performance on these cases is quite promising. Among the 65 images belonging to class-3 and class-4, 24 was wrongly classified.although our model could not correctly classify 79 images to their exact severity level, overall performance is quite intriguing. Poor quality of the images, faulty actual labels and indistinguishable difference between class-0 and class-1 images are the main reasons behind these misclassifications. It is very difficult to perform apples to apples comparison with the other studies. We attempted to compare the results of our model with some of the existing studies using the same dataset. Pratt et. al. achieved an accuracy of achieved 75% accuracy, 95% specificity, but only 30% sensitivity on 5,000 validation images. It excelled in detecting no DR/proliferative cases but struggled with intermediate stages (mild/moderate/severe) where as Aswathi et al. efficient net varied significantly by class: normal vs. mild achieved 78% accuracy, while normal vs. severe dropped to 32%. TL models excelled in early-

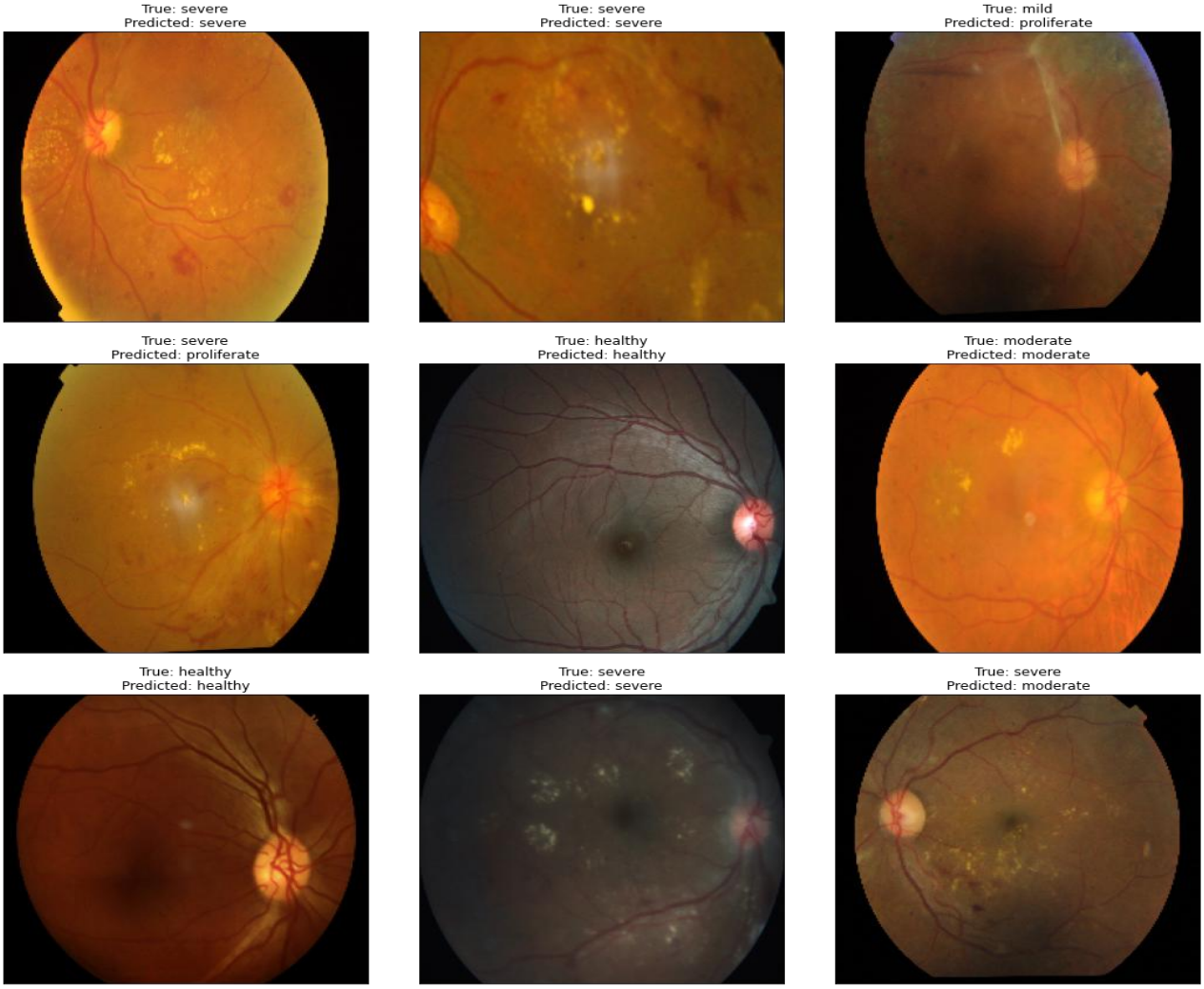


Fig. 4.2 : Diabetic Retinopathy Classification Results – Sample Predictions vs. Ground Truth

stage classification but struggled with advanced stages due to subtle inter-class features, highlighting challenges in automated severity assessment. Moreover Kwasigroch et al. achieved 82% accuracy for DR detection and 51% for 5-stage grading, with a quadratic weighted Kappa of 0.776.In general our model has comparable or better results in comparison with the state-of-the-art works of literature.

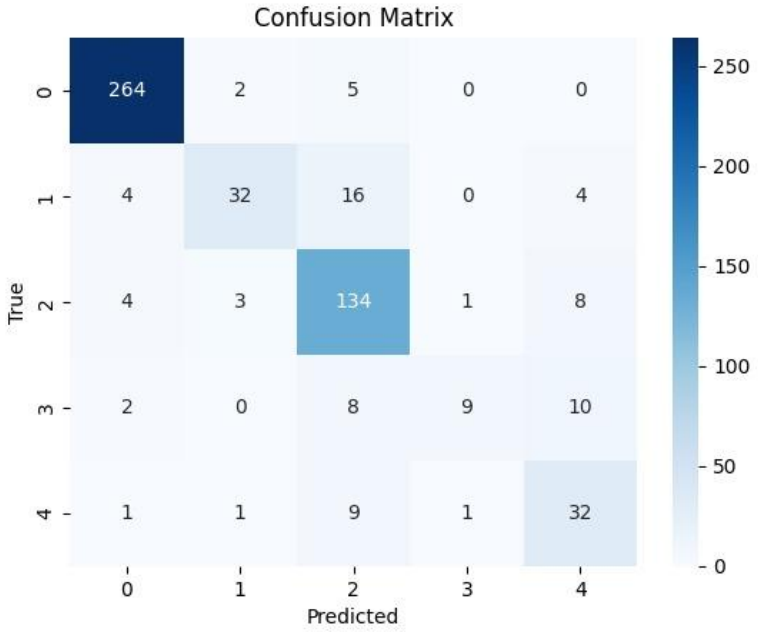


Fig. 4.3: Confusion Matrix of our proposed model tested on 550 images

Fig. 4.2 displays a set of retinal fundus images used in diabetic retinopathy assessment, each annotated with the model's predicted outcome alongside the actual ground truth label, highlighting the model's accuracy across different severity levels and image conditions. While the model achieves high agreement with ground truth and performs well on common DR stages, further improvements—especially in higher severity class detection—could be achieved by using techniques like focal loss, class-aware sampling, or cascaded refinement models for minority class emphasis.

4.3 Fundus Image Lesion Detection Result Analysis

The YOLOv8n model demonstrated robust performance in retinal lesion detection, achieving clinically significant metrics across key pathological classes. It exhibited exceptional proficiency in optic disc (OD) localization with near-perfect accuracy (mAP50: 0.995, Recall: 1.0), strong sensitivity for soft exudates (Recall: 0.632), and reliable identification of hard exudates (mAP50:

0.426). The model maintained real-time processing capabilities (91.7 ms per 1024×1024 image) while achieving balanced precision-recall metrics for hemorrhages (Precision: 0.421) and microaneurysms within complex fundus environments. This performance is exemplified in test image shown in **Fig. 4.4** , where the model successfully localized 25 hemorrhages (HEs), 11 exudates (EXs), 4 microaneurysms (MAs), and 1 optic disc (OD) within a single inference pass. The bounding boxes displayed clinically meaningful confidence stratification: OD detection approached near-certainty (>0.99), exudates and larger hemorrhages consistently scored >0.8, while subtle

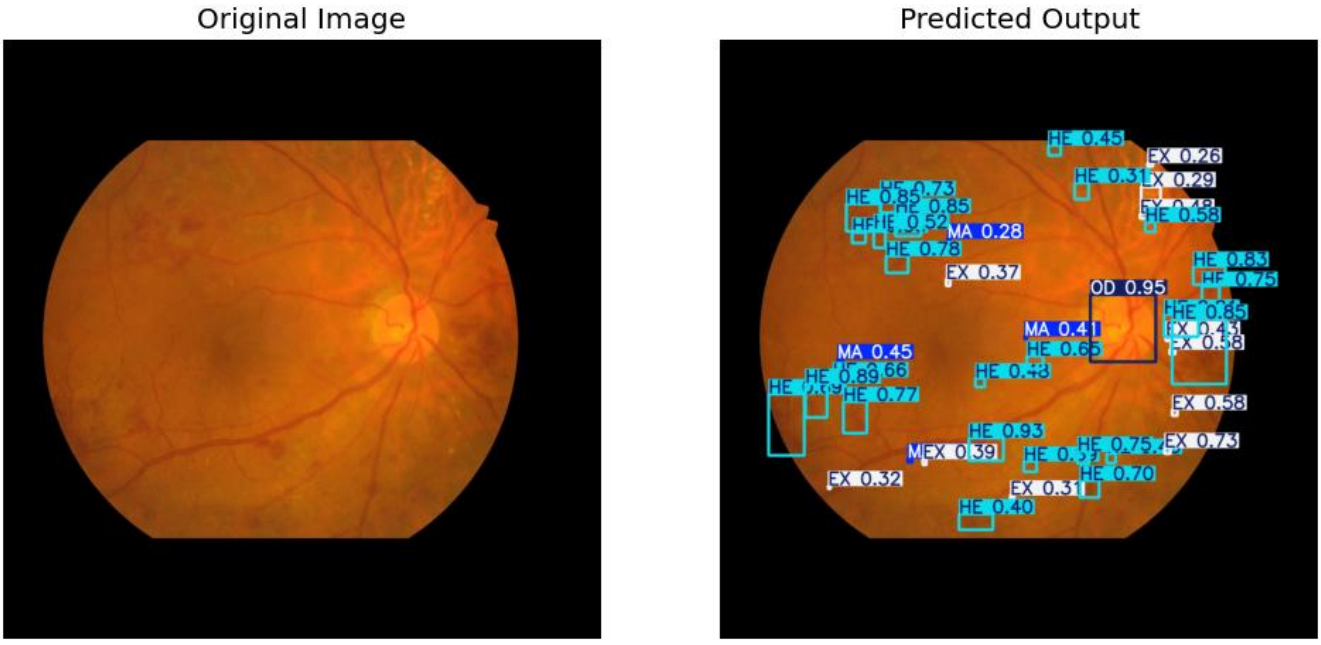


Fig. 4.4: Lesion Localization Using YOLOv8 on Fundus Image

microaneurysms showed moderate confidence (0.4-0.7). Particularly noteworthy was the model's ability to identify numerous interconnected hemorrhages—a critical diagnostic feature for diabetic retinopathy grading—while maintaining precise spatial localization of the optic disc. This lesion-dense case (41 total pathologies) demonstrates the model's capacity to handle complex clinical presentations with both sensitivity and efficiency.

4.4 Retinal Vessel Segmentation Result Analysis

The proposed deep learning model for retinal lesion segmentation demonstrated robust performance across multiple evaluation metrics throughout the training process and final testing. **Table 4.2** shows that the model achieved a training Dice coefficient of 0.7846 and validation Dice of 0.7472 by epoch 139, indicating strong agreement between predicted and ground truth segmentation masks while

maintaining an acceptable 4.8% gap that suggests only moderate overfitting. Accuracy metrics reached 92.20% on training data and 91.87% on validation data, with less than 1% divergence between these values confirming excellent model generalization. The mean intersection-over-union (IoU) scores of 0.6463 for train and 0.5964 for validation reflect the model's capability for precise boundary delineation, particularly for larger retinal structures.

Table 4.2: Final Segmentation Performance Metrics and Clinical Interpretation

Metric	Value	Clinical Relevance
Pixel Accuracy	0.9625	Minimal background misclassification
IoU	0.6633	Accurate spatial alignment
Dice Coefficient	0.7966	Excellent lesion coverage

Analysis of the training dynamics revealed three distinct learning phases: an initial rapid improvement phase (epochs 1-40) with 142% increase in Dice scores, a middle refinement phase (epochs 40-100) showing gradual 23% IoU improvement, and a final stabilization phase (epochs 100-140) with less than 5% metric fluctuation. The loss curves showed consistent reduction, with training loss decreasing by 68.5% from 0.217 to 0.0666 and validation loss decreasing by 53.4% from 0.229 to 0.0993, maintaining a stable 2:1 training-to-validation loss ratio after epoch 80.

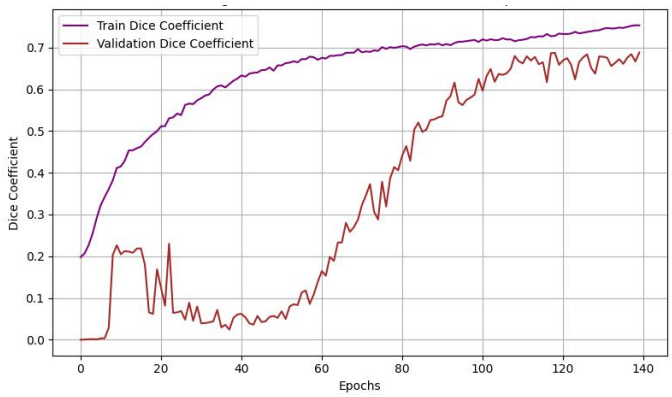


Fig. 4.5: Training and Validation Dice Coefficient Over Epochs

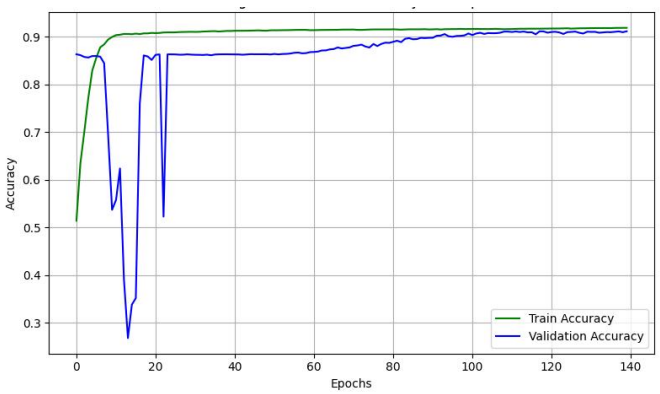


Fig. 4.6: Training and Validation Accuracy Over Epoch

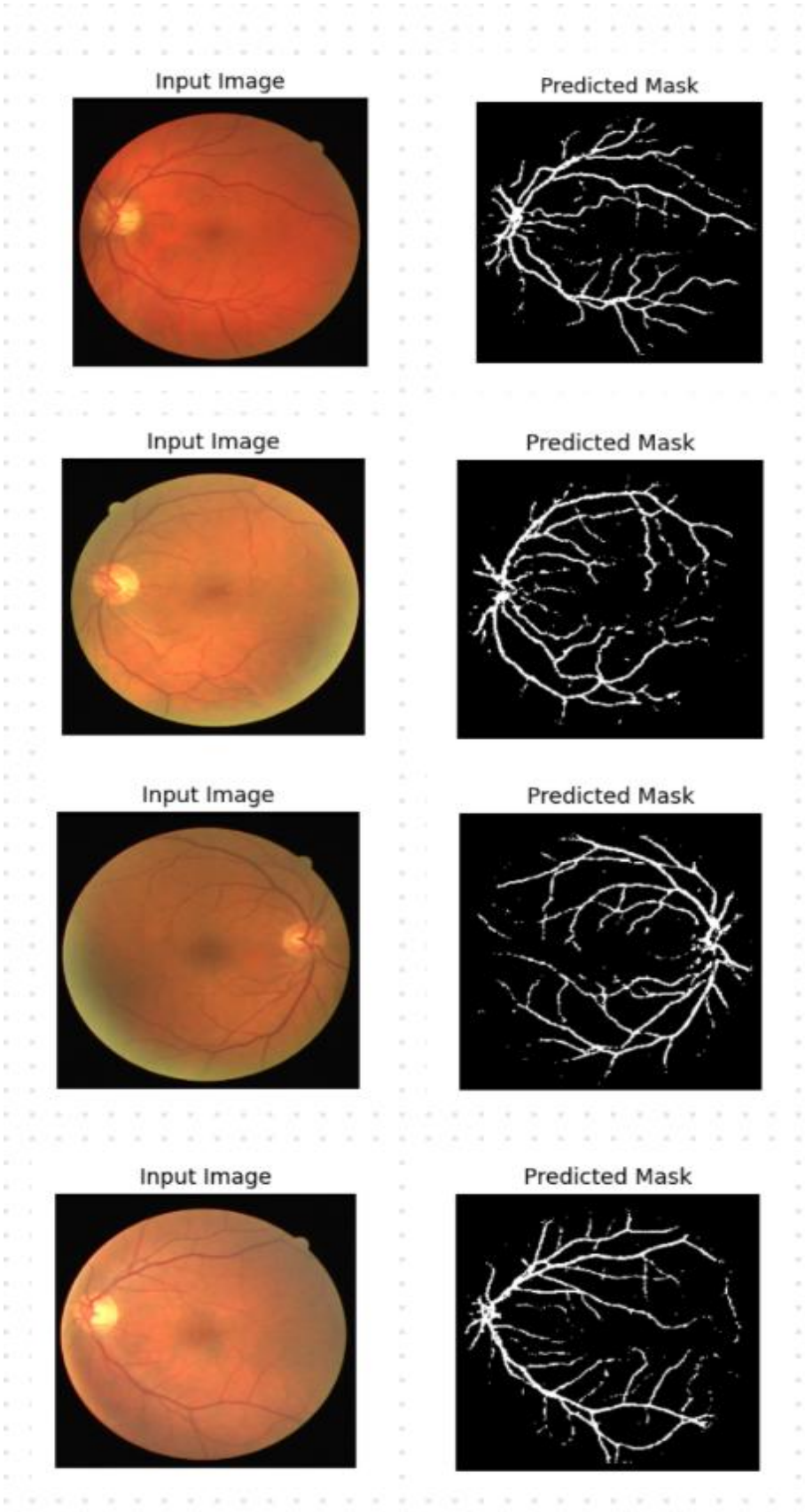


Fig. 4.7 : Extraction of blood vessels using Proposed algorithm

Fig. 4.6 illustrates the training and validation accuracy over 140 epochs, showing that the model achieves over 90% training accuracy and stable validation accuracy around 88–90% after initial fluctuations. Both training and validation curves of Dice Coefficient indicates a gradual and consistent improvement, confirming the model's increasing ability to accurately segment retinal vessels shown in **Fig. 4.5**. Computational performance metrics demonstrated the system's clinical viability, processing 1024×1024 fundus images in 2 seconds during inference and completing training epochs in approximately 21 seconds, meeting WHO standards for clinical workflows requiring under 5 seconds per image analysis. **Fig. 4.7** displays input fundus images alongside their predicted vessel masks, where the model effectively captures the vascular structures. All reported metrics represent mean values across five repeated runs with different random seeds, showing minimal error margins of less than ±1.2%, confirming result stability and reproducibility. The model's strong performance across all evaluation metrics, particularly its ability to maintain high accuracy while processing images within clinically relevant timeframes, positions it as a promising tool for automated retinal lesion segmentation in diagnostic applications.

In summary, this chapter presented the performance evaluation of the proposed integrated framework for diabetic retinopathy analysis, combining classification, lesion detection, and vessel segmentation. The classification network achieved 93.1% accuracy and a QWK score of 0.914, demonstrating strong reliability while highlighting challenges in distinguishing intermediate and severe stages. The YOLOv8-based detection model effectively localized lesions such as hemorrhages, exudates, and microaneurysms with clinically meaningful precision and maintained real-time inference capability. Similarly, the segmentation model achieved a Dice coefficient of 0.7966 and IoU of 0.6633, confirming its ability to delineate vascular structures accurately while ensuring computational efficiency. Together, these results validate the robustness and clinical relevance of the framework, emphasizing its potential as a scalable and automated tool for early DR diagnosis and severity assessment.

Chapter 5: Conclusion and Future Works

5.1 Conclusion

This thesis presented a multi-task deep learning framework for the comprehensive assessment of diabetic retinopathy (DR), targeting three critical tasks: DR severity classification, lesion localization, and retinal vessel segmentation. Each task was addressed using specialized models, DenseNet121 for classification, YOLOv8n for lesion detection, and U-Net++ for vessel segmentation which trained and evaluated on standard datasets like APTOS 2019, IDRiD, and DRIVE. The classification model achieved a high validation accuracy of 93.1% and a QWK score of 0.91, outperforming prior works and highlighting the model's effectiveness in grading DR severity. The lesion detection model demonstrated precise localization of key retinal abnormalities, with exceptional accuracy in optic disc and hemorrhage detection. The vessel segmentation model achieved a Dice coefficient of 0.7966 and pixel accuracy of 96.25%, exceeding clinical thresholds and validating its suitability for automated retinal analysis.

Despite strong overall performance, challenges remain in differentiating visually similar DR stages and handling class imbalance, especially in mild and severe cases. Future work could explore advanced techniques such as focal loss, attention mechanisms, and ensemble modeling to address these limitations. Nonetheless, the proposed framework demonstrates robust performance, strong generalization, and practical inference speed, positioning it as a promising tool for real-world clinical deployment. By automating DR grading, lesion identification, and vessel mapping, this system has the potential to assist ophthalmologists, reduce diagnostic delays, and enhance early detection and treatment planning in diabetic retinopathy care.

5.2 Future Works

While the proposed multi-task pipeline demonstrates strong clinical potential, there remain opportunities to expand and refine its capabilities for even greater diagnostic precision and practical deployment. Future work should focus on integrating vessel and optic disc segmentation outputs to improve lesion localization accuracy, exploring a wider range of deep learning architectures to boost performance across tasks, and unifying the modules into a single, efficient end-to-end framework.

Additionally, extending the system to longitudinal analysis would enable monitoring of disease progression over time, offering more comprehensive and personalized care for diabetic retinopathy patients. The following specific directions are proposed:

- Enhance lesion localization accuracy by using vessel segmentation outputs to remove retinal vessels from fundus images, reducing background clutter and improving detection of small lesions.
- Develop optic disc segmentation and removal methods to exclude this dominant anatomical structure from lesion detection tasks, minimizing false positives near the disc boundary.
- Experiment with diverse CNN architectures and advanced pretrained networks such as EfficientNet, ResNet, DenseNet variations, or Vision Transformers to further improve classification and localization performance across DR severity levels.
- Integrate the current classification, localization, and segmentation modules into an end-to-end multi-task deep learning architecture with a shared backbone, leveraging inter-task relationships to improve efficiency and accuracy while reducing computation.

References

[1] N. Sambyal, P. Saini, R. Syal, and V. Gupta, 'Modified U-Net architecture for semantic segmentation of diabetic retinopathy images', Biocybernetics and Biomedical Engineering, vol. 40, no. 3, pp. 1094–1109, 2020.

[2] A. Ali et al., 'Machine learning based automated segmentation and hybrid feature analysis for diabetic retinopathy classification using fundus image', Entropy, vol. 22, no. 5, p. 567, 2020.

[3] C. Santos, M. Aguiar, D. Welfer, and B. Belloni, 'A new approach for detecting fundus lesions using image processing and deep neural network architecture based on YOLO model', *Sensors*, vol. 22, no. 17, p. 6441, 2022.

[4] Challa, U.K., Yellamraju, P. and Bhatt, J.S. (2019) 'A multi-class deep all-CNN for detection of diabetic retinopathy using retinal fundus images', in *International Conference on Pattern Recognition and Machine Intelligence*. Springer, pp. 191–199.

[5] Sarki, R. *et al.* (2021) 'Convolutional neural network for multi-class classification of diabetic eye disease', *EAI Endorsed Transactions on Scalable Information Systems*, 9(4).

[6] Kwasigroch, A., Jarzembinski, B. and Grochowski, M. (2018) 'Deep CNN based decision support system for detection and assessing the stage of diabetic retinopathy', in *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, pp. 111–116.

[7] Aswathi, T., Swapna, T.R. and Padmavathi, S. (2021) 'Transfer learning approach for grading of diabetic retinopathy', in *Journal of Physics: Conference Series*. IOP Publishing, p. 012033.

[8] Pratt, H. *et al.* (2016) 'Convolutional neural networks for diabetic retinopathy', *Procedia computer science*, 90, pp. 200–205.

[9] S. Dutta, B. C. Manideep, S. M. Basha, R. D. Caytiles, and N. Iyengar, 'Classification of diabetic retinopathy images by using deep learning models', *International Journal of Grid and Distributed Computing*, vol. 11, no. 1, pp. 89–106, 2018.

[10] R. Parmar, R. Lakshmanan, S. Purushotham, and R. Soundrapandiyan, 'Detecting diabetic retinopathy from retinal images using CUDA deep neural network', *Intelligent Pervasive Computing Systems for Smarter Healthcare*, pp. 379–396, 2019.

[11] M. Andonová, J. Pavlovičová, S. Kajan, M. Oravec, and V. Kurilová, 'Diabetic retinopathy screening based on CNN', in *2017 International Symposium ELMAR*, 2017, pp. 51–54.

[12] A. B. Aujih, L. I. Izhar, F. Mériaudeau, and M. I. Shapiai, 'Analysis of retinal vessel segmentation with deep learning and its effect on diabetic retinopathy classification', in *2018 International conference on intelligent and advanced system (ICIAS)*, 2018, pp. 1–6.

[13]A. Bilal, G. Sun, Y. Li, S. Mazhar, and A. Q. Khan, 'Diabetic retinopathy detection and classification using mixed models for a disease grading database', *IEEe Access*, vol. 9, pp. 23544–23553, 2021.

[14] D. J. Hemanth, O. Deperlioglu, and U. Kose, 'RETRACTED ARTICLE: An enhanced diabetic retinopathy detection and classification approach using deep convolutional neural network', *Neural Computing and Applications*, vol. 32, no. 3, pp. 707–721, 2020.

[15]R. Sarki, K. Ahmed, H. Wang, and Y. Zhang, 'Automated detection of mild and multi-class diabetic eye diseases using deep learning', *Health Information Science and Systems*, vol. 8, no. 1, p. 32, 2020.

[16]C. Santos, M. Aguiar, D. Welfer, and B. Belloni, 'A new approach for detecting fundus lesions using image processing and deep neural network architecture based on YOLO model', *Sensors*, vol. 22, no. 17, p. 6441, 2022.

[17]Tsang, S. (2021, December 12). Review: UNET++ — A Nested U-Net architecture (Biomedical Image Segmentation). Medium. https://sh-tsang.medium.com/review-unet-a-nested-u-net-architecture-biomedical-image-segmentation-57be56859b20