# Leveraging Deep Learning for Diabetic Retinopathy Diagnosis

PHASE I REPORT

*Submitted by*


| | |
|---|---|
| **Lohesh M** | **21011101069** |
| **P Esvar Ram Kumar** | **21011101043** |
| **Rysha M Rafeeque** | **21011101103** |
| **P Rathna Sabapathy** | **21011101098** |


*in partial fulfilment for the award of the degree of*

**BACHELOR OF TECHNOLOGY**
**IN**
**ARTIFICIAL INTELLIGENCE**
**&**
**DATA SCIENCE**



**DEPARTMENT OF COMPUTER SCIENCE**
**AND ENGINEERING**
**SCHOOL OF ENGINEERING**
**SHIV NADAR UNIVERSITY CHENNAI**


**NOVEMBER 2024**

# SHIV NADAR UNIVERSITY CHENNAI

## BONAFIDE CERTIFICATE

Certified that this report titled "**Leveraging Deep Learning for Diabetic Retinopathy Diagnosisis**" is the bonafide work of **Lohesh M** (Reg. No: **21011101069**), **P Esvar Ram Kumar** (Reg. No: **21011101043**), **Rysha M Rafeeque** (Reg. No: **21011101103**), **P Rathna Sabapathy** (Reg. No: **21011101098**) who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.


**SIGNATURE**

Dr. T. Nagarajan

**Professor**

**Head of the Department**

Department of Computer Science & Engineering

School of Engineering

Shiv Nadar University Chennai,

Kalavakkam -603110.

**SIGNATURE**

Ms.Dhivya S

**Assistant Professor**

**Supervisor**

Department of Computer Science & Engineering

School of Engineering

Shiv Nadar University Chennai,

Kalavakkam -603110.

# ABSTRACT

Diabetic Retinopathy (DR) is a leading cause of blindness among diabetic patients, and its early detection plays a critical role in preventing vision loss. This project aims to develop a comprehensive deep learning-based solution for both the segmentation of retinal abnormalities and the classification of Diabetic Retinopathy severity. Using the IDRiD dataset, the segmentation task involves accurately identifying key retinal lesions such as microaneurysms, hemorrhages, and exudates. Additionally, the classification task categorizes retinal images into distinct DR stages, from no DR to proliferative DR, to support clinical decision-making. We experiment with various architectures, focusing on UNet and its hybrid variants, to achieve accurate and reliable segmentation. The project also integrates advanced preprocessing techniques like CLAHE and image augmentation to enhance model performance. While classification is initially approached using conventional CNNs, we plan to incorporate Explainable AI (XAI) techniques in the future to improve model interpretability and ensure its reliability for clinical applications. Our goal is to provide a deployable solution that aids healthcare professionals in diagnosing and managing Diabetic Retinopathy more effectively.

**Keywords** - Diabetic Retinopathy, Deep Learning, Segmentation, Classification, ExplainableAI

# ACKNOWLEDGEMENT

we would like to express our sincere gratitude to all those who have provided invaluable guidance, support, and encouragement throughout the course of this project.

First and foremost, We are deeply grateful to **Dr. T. Nagarajan**, Professor and Head, Department of Computer Science and Engineering, School of Engineering, Shiv Nadar University Chennai, for his steadfast guidance and unwavering encouragement. His insights into the field of artificial intelligence and cognitive science have been instrumental in shaping the direction of this research.

We owe a special debt of thanks to our supervisor, **Ms Dhivya S**, Assistant Professor, Department of Computer Science and Engineering, School of Engineering, Shiv Nadar University Chennai, for inspiring us, motivating us, and guiding us to perform our Major Project creatively and successfully. Her guidance and availability whenever we needed assistance were invaluable. We also acknowledge with gratitude the support of our family, whose understanding and encouragement have been crucial to the implementation of this project. Their unwavering belief in us has been a constant source of motivation and strength.

**Lohesh M**
**(21011101069)**

**P Esvar Ram Kumar**
**(21011101043)**

**Rysha M Rafeeque**
**(21011101103)**

**P Rathna Sabapathy**
**(21011101098)**

# TABLE OF CONTENTS

# List of Tables

# List of Figures

# Chapter 1

# INTRODUCTION

## 1.1 Overview

Diabetic Retinopathy (DR) is a severe complication of diabetes that can lead to blindness if not managed early. DR causes damage to retinal blood vessels, resulting in abnormalities such as microaneurysms, hemorrhages, and exudates. Early detection is vital for preventing further vision loss, and this project leverages machine learning to aid in the diagnosis and classification of DR severity, which can enhance clinical decision-making and treatment planning.

The project has two primary tasks: **Segmentation** and **Classification** of retinal images. In segmentation, the goal is to identify and outline specific abnormalities associated with DR within retinal images, providing localized insights into affected regions. For classification, the task is to categorize the overall severity of DR based on the identified abnormalities, allowing for streamlined diagnosis at different stages of the disease.

The dataset used is a combination of the **IDRiD**(Indian Diabetic Retinopathy Image Dataset) for segmentation tasks and the **Kaggle Diabetic Retinopathy 2019** dataset for classification. The IDRiD dataset includes detailed annotations for retinal lesions, making it suitable for precise segmentation, while the Kaggle dataset is labeled with five DR severity levels, facilitating multi-class classification.

This dual approach allows the project to both locate specific retinal abnormalities and assess the severity of DR, creating a comprehensive diagnostic tool that supports both general and detailed analysis of retinal images.

## 1.2 Problem Formulation

The formulation of this project is structured around these two core components: Segmentation of abnormalities and Classification of DR severity.

### 1.2.1 Segmentation

Using the IDRiD dataset, the project focuses on localizing specific abnormalities (microaneurysms, hemorrhages, exudates, and optic disc regions) within retinal images. Deep learning models, such as UNet and Global Convolutional Networks (GCNs), are employed to achieve high-precision segmentation. These architectures use an encoder-decoder approach, preserving spatial information to accurately define each abnormality's shape and location. To improve model robustness, data augmentation techniques are applied to increase variability, such as random rotation, cropping, and color adjustments, which ensure that the model generalizes well to unseen images.

### 1.2.2 Classification

Using the Kaggle Diabetic Retinopathy dataset, a deep learning classifier categorizes images into five DR severity levels: No DR, Mild, Moderate, Severe, and Proliferative DR. Transfer learning with pre-trained networks is applied to optimize feature extraction, improving classification accuracy even with limited labeled data. The classifier enhances treatment planning by distinguishing severity stages, enabling timely interventions for patients with advanced DR stages.

### 1.2.3 Explainable AI (XAI)

Since interpretability is crucial in medical diagnostics, Explainable AI techniques like Grad-CAM and Saliency Maps are incorporated. These methods highlight areas of interest in each image, allowing clinicians to verify model predictions. The added transparency from XAI builds trust by explaining the model's focus areas for each diagnosis, making it suitable for clinical applications.

## 1.3 Objectives

This project aims to build an advanced diagnostic tool for Diabetic Retinopathy (DR) by developing machine learning models for both segmentation of retinal abnormalities and classification of DR severity. The primary objectives include achieving high accuracy, interpretability, and clinical applicability, with a strong focus on robust real-world deployment. The objectives are as follows:

1. **Develop an Accurate Segmentation Model for Retinal Lesions:**
    - Build a model capable of precisely identifying and segmenting key retinal abnormalities associated with DR, including microaneurysms, hemorrhages, exudates, and the optic disc.
2. **Develop a Classification Model for DR Severity Stages:**
    - Design a classification model that accurately categorizes retinal images into the five clinically significant DR severity stages: No DR, Mild, Moderate, Severe, and Proliferative DR.
3. **Apply Explainable AI (XAI) Techniques to Enhance Model Transparency:**
    - Integrate Explainable AI methods to visualize the model's focus areas and highlight specific retinal regions contributing to both classification and segmentation outputs. These visualizations ensure clinicians can observe the reasoning behind model predictions, enhancing trust and interpretability in clinical contexts.
4. **Build a Deployable Solution with Clinical Utility:**
    - Structure the model's outputs—both segmented images and severity classifications—in a format suitable for integration into real-world clinical workflows. This includes clear, segmented masks that highlight abnormal areas and severity scores that support patient diagnosis and tailored treatment planning.

## 1.4 Motivation

The motivation for this project arises from the critical need for reliable and efficient diagnostic tools for diabetic retinopathy (DR), a leading cause of blindness in diabetic patients. While traditional diagnostic methods are effective, they often rely on the expertise of ophthalmologists and can be time-consuming. This makes them less accessible, especially in regions with limited healthcare resources. Early detection of DR is essential to prevent vision loss, and there is an increasing need for automated systems to assist healthcare professionals in providing accurate and timely diagnoses.

In response to these challenges, this project leverages advanced deep learning techniques to create an automated system that can segment retinal lesions and classify the severity of diabetic retinopathy. By experimenting with different model architectures, we aimed to improve upon existing systems. One key aspect of this project is the development of a hybrid UNet model. This model combines elements of UNet with other advanced techniques to enhance segmentation accuracy, especially in detecting subtle retinal abnormalities such as microaneurysms, hemorrhages, and exudates.

We tested multiple models—including standard UNet, Global Convolutional Networks (GCN), and a hybrid UNet—to improve the detection and segmentation of retinal lesions, which are often difficult to distinguish due to the complex nature of retinal images. The hybrid UNet model integrates the strengths of various architectures to address challenges like inter-class variability and class imbalance in the dataset, offering better performance in segmenting regions of interest.

Ultimately, the goal of this project is to build a robust, automated tool that can help detect and classify diabetic retinopathy in its early stages, providing clinicians with accurate and timely insights for better patient care. Through these innovations, we aim to improve the accessibility and efficiency of DR diagnosis, ensuring that patients, regardless of geographic location, receive the care they need to prevent vision loss.

In the future, we plan to integrate Explainable AI (XAI) techniques to enhance the model's transparency, making it more accessible to doctors and improving its reliability for clinical decision-making.

# Chapter 2

# Background

## 2.1 What is Diabetic Retinopathy ?

Diabetic Retinopathy (DR) is one of the most common complications of diabetes, affecting the blood vessels in the retina, the light-sensitive tissue at the back of the eye. As the disease progresses, it can cause vision impairment and ultimately blindness if not detected and treated early. Early diagnosis is crucial because diabetic retinopathy is asymptomatic in its early stages, meaning that individuals may not notice any symptoms until the disease has already caused significant damage. Therefore, detecting DR early through regular eye examinations is essential for preventing vision loss.

### 2.1.1 Diabetic Retinopathy and Its Severity

Diabetic retinopathy occurs in several stages, with increasing severity:

- **No Diabetic Retinopathy (No DR):** No observable abnormalities in the retina.
- **Mild Non-Proliferative Diabetic Retinopathy (Mild DR):** The first stage of DR where small microaneurysms form, which are tiny bulges in the blood vessels.
- **Moderate Non-Proliferative Diabetic Retinopathy (Moderate DR):** More blood vessels in the retina become blocked.
- **Severe Non-Proliferative Diabetic Retinopathy (Severe DR):** A significant number of blood vessels are blocked, leading to poor blood flow to the retina.

- **Proliferative Diabetic Retinopathy (PDR):** The most advanced stage, where new blood vessels grow abnormally on the retina and in the vitreous, increasing the risk of hemorrhage and vision loss.

These stages can be identified by ophthalmologists using retinal images. However, manual examination can be time-consuming and subject to human error, especially in resource-limited settings.

### 2.1.2 Traditional Methods of Diabetic Retinopathy Detection

Traditionally, the diagnosis of DR involves direct observation of retinal images, typically obtained through retinal fundus photography. Ophthalmologists examine these images for various signs of diabetic retinopathy, such as microaneurysms, hemorrhages, exudates, and neovascularization. However, this process is time-consuming and relies heavily on the experience of the specialist.

In recent years, the advent of machine learning (ML) and deep learning (DL) techniques has provided a promising alternative. These approaches can automate the detection and classification of DR with high accuracy, reducing the time and effort required for manual assessment and increasing the accessibility of DR diagnosis.

### 2.1.3 Deep Learning for Diabetic Retinopathy

Deep learning, particularly convolutional neural networks (CNNs), has emerged as one of the most effective approaches for medical image analysis, including retinal image analysis. CNNs can automatically learn hierarchical features from raw image data, making them suitable for tasks like image classification and segmentation. For DR, CNN-based models can be trained to identify retinal lesions and classify the severity of the disease based on labeled data.

Segmentation plays a critical role in this process. Accurate segmentation of retinal abnormalities—such as microaneurysms, hemorrhages, and exudates—is essential for both diagnosis and severity classification. Deep learning models like UNet have become popular for segmentation tasks in biomedical image analysis due to their ability to effectively segment images at pixel-level granularity.

### 2.1.4 Challenges in Diabetic Retinopathy Detection

While deep learning methods have shown promising results, several challenges persist:

- **Class Imbalance:** In most retinal images, the regions corresponding to lesions are relatively small compared to the background, leading to class imbalance. This makes it challenging for models to learn to detect these small abnormalities effectively.
- **Inter-Class Variability:** Different types of lesions can vary significantly in size, shape, and intensity, making it difficult for a single model to effectively detect all lesion types.
- **Data Quality:** The quality of retinal images can vary due to different imaging conditions, such as lighting, resolution, and focus, which can impact model performance.

### 2.1.5 Explainable AI in Medical Imaging

In the field of medical imaging, Explainable AI (XAI) has become increasingly important. XAI techniques aim to make machine learning models more transparent by providing insights into how models make their predictions. This is especially critical in healthcare, where clinicians need to trust and understand the decision-making process of AI models to use them in practice.

For DR detection, XAI methods such as Grad-CAM (Gradient-weighted Class Activation Mapping) and Saliency Maps are commonly used to highlight regions of the retina that contribute to the model's decision. By visualizing the areas of the image that are most influential in the model's classification or segmentation decision, these techniques provide transparency, which helps clinicians better understand and trust the AI system.

### 2.1.6 Relevant Datasets for Diabetic Retinopathy

Several datasets have been developed to aid in the research and development of automated diabetic retinopathy detection systems:

- **IDRiD (Indian Diabetic Retinopathy Image Dataset):** A dataset that contains retinal images annotated with information on various retinal abnormalities, such as

microaneurysms, hemorrhages, exudates, and optic disc.

- **Kaggle Diabetic Retinopathy 2019 Dataset:** A widely used dataset that contains retinal fundus images classified into five stages of diabetic retinopathy severity: No DR, Mild DR, Moderate DR, Severe DR, and Proliferative DR. This dataset allows for the training of both segmentation and classification models.

Both of these datasets provide a valuable resource for training and testing deep learning models aimed at automating DR diagnosis.

### 2.1.7 Related Work

Recent studies have focused on applying deep learning models, such as CNNs and UNet, for diabetic retinopathy detection. Some approaches focus on using CNNs for image classification to assess DR severity, while others apply segmentation models to detect specific lesions. Researchers have also explored hybrid models combining different architectures to improve segmentation performance, particularly for small or subtle retinal abnormalities.

Some models integrate techniques from XAI to enhance the interpretability of results. These advancements aim to provide more reliable and clinically useful AI-driven diagnostic tools for healthcare professionals.

This project builds upon the existing body of work by experimenting with different deep learning models, including a hybrid UNet architecture, to improve both the segmentation and classification of DR. Additionally, future work will incorporate Explainable AI techniques to ensure that the model's predictions are transparent, interpretable, and accessible to clinicians, improving trust in AI-based medical tools.

# Chapter 3

# Segmentation

## 3.1 Introduction: The Problem and Dataset

Diabetic Retinopathy (DR) is a complication of diabetes where high blood sugar levels damage the blood vessels in the retina, potentially leading to vision loss. Early detection of DR is crucial for preventing progression to more severe stages that could result in blindness. However, manual detection can be time-consuming and error-prone, particularly in large-scale screening.

To address this, the IDRiD (Indian Diabetic Retinopathy Image Dataset) is used in this project. This dataset contains annotated retinal images highlighting various abnormalities associated with DR, including:

- Microaneurysms (MA)
- Hemorrhages (HE)
- Exudates (EX)
- Soft Exudates (SE)
- Optic Disc (OD)

The goal is to segment these abnormalities using deep learning techniques to build an automated model capable of detecting DR and assessing its severity. Challenges such as high variability among abnormality classes and significant class imbalance—due to the large background-to-abnormality ratio in many images—need to be addressed.

## 3.2 Preprocessing and Data Augmentation

To prepare the data for training and enhance the model's ability to generalize, various preprocessing and augmentation techniques are applied.

**Data Preprocessing**

- **Image Resizing:** Retinal images are resized to a fixed dimension (e.g., $512 \times 512$ or $640 \times 640$ pixels) to ensure uniformity across the dataset.
- **Normalization:** Pixel values are scaled between 0 and 1 to stabilize training and prevent issues related to saturation in neural networks.
- **Mask Handling:** The ground-truth masks, which identify regions of interest (ROIs) corresponding to abnormalities, are binarized to distinguish between background (0) and abnormalities (1).

**Data Augmentation Techniques**

- **Random Rotation:** Images and their corresponding masks are rotated by random angles to increase model robustness against orientation.
- **Random Cropping:** Small regions are randomly cropped from each image to simulate different perspectives, enhancing localization performance.
- **CLAHE:** Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied to the green channel to enhance lesion visibility while reducing noise.
- **Brightness, Color, and Contrast Adjustments:** These augmentations simulate different imaging conditions, helping the model generalize better.

## 3.3 Model Architectures for Segmentation

### 3.3.1 Global Convolutional Network (GCN)

GCN helps capture broader contextual information for segmenting small lesions alongside large anatomical structures.

- **Large Kernel Convolutions:** Large kernel convolutions improve accuracy of segmentation for small lesions.
- **Boundary Refinement Module (BRM):** BRM refines segmentation boundaries for precise transitions.

### 3.3.2 UNet

UNet is widely used for biomedical image segmentation. It consists of two main parts:

- **Encoder-Decoder Structure:** The encoder downsamples the image, capturing contextual features, while the decoder upsamples it to reconstruct the original size, producing a segmentation map.
- **Skip Connections:** Connections link encoder and decoder layers, allowing the model to retain spatial information, improving segmentation accuracy.

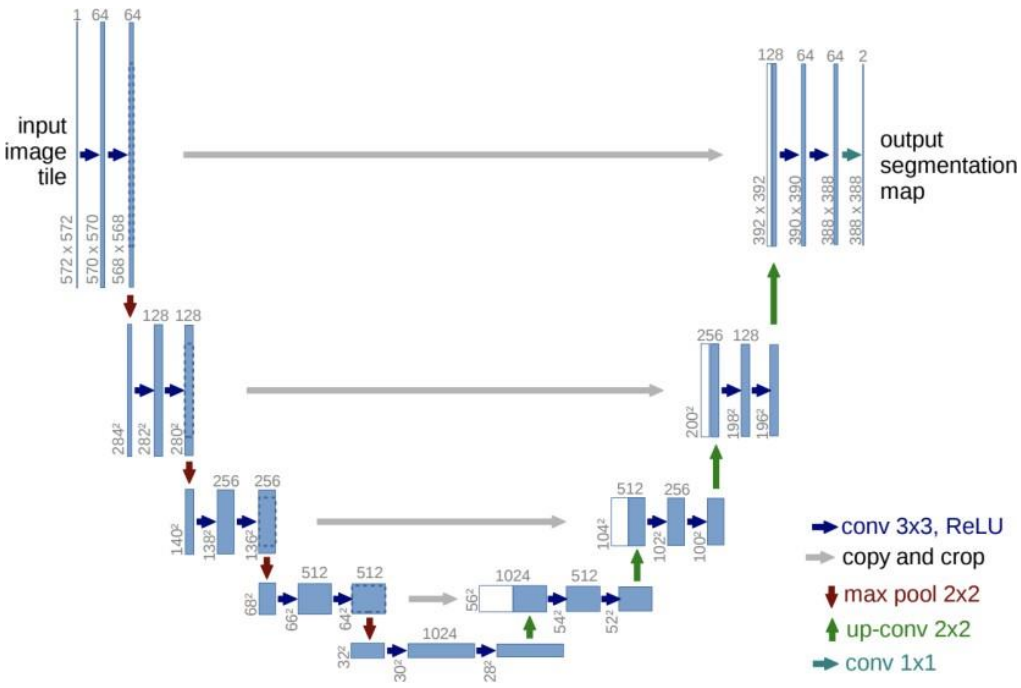


Figure 3.1: U-Net Architecture Overview

### 3.3.3 U-Net with VGG16 Backbone

The U-Net architecture is composed of an encoder-decoder pathway where VGG16, a well-established convolutional neural network model, serves as the encoder. This encoder,

pre-trained on extensive image data, captures crucial hierarchical features that highlight differences between healthy and abnormal retinal areas. Key components include:

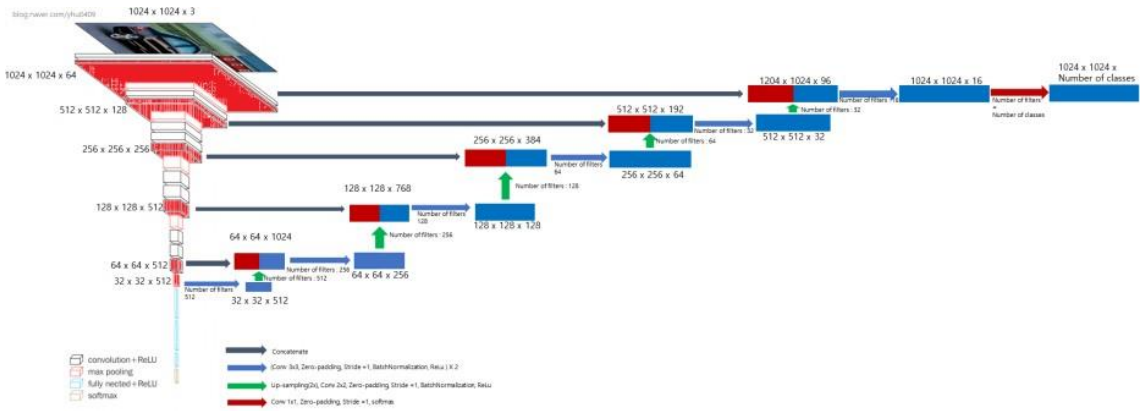


Figure 3.2: U-Net with VGG16 Backbone Architecture

- **Encoder-Decoder Pathway:** The VGG16-based encoder progressively downsamples the input image, while the decoder upsamples to reconstruct the image resolution.
- **Skip Connections:** Skip connections connect encoder and decoder layers, retaining spatial information and ensuring that fine details are preserved for precise segmentation.
- **Convolutional Blocks:** Each block within the encoder and decoder consists of two convolutional layers followed by batch normalization and activation functions.

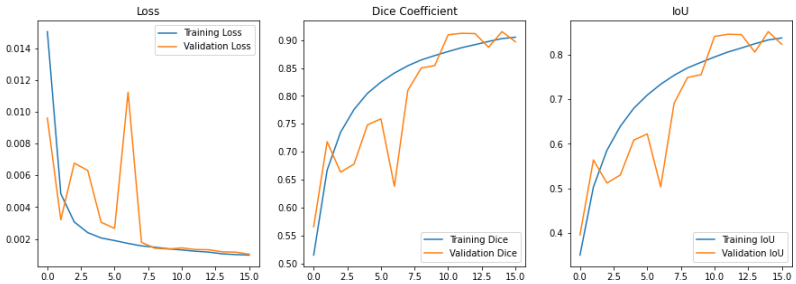


Figure 3.3: Model Performance Metrics Over Training

## 3.4 Loss Functions

Class imbalance between background and lesions necessitates appropriate loss functions.

- **Dice Loss:** Maximizes overlap between predicted and ground-truth masks.
- **Cross-Entropy Loss:** Used for pixel-wise classification in multi-class segmentation.
- **Focal Loss:** Reduces impact of background pixels, focusing on difficult-to-classify lesions.

## 3.5 Evaluation Metrics

The performance of the segmentation model is evaluated using:

- **AUPR:** Area Under Precision-Recall Curve measures precision vs. recall for each lesion class.
- **ROC AUC:** Area Under Curve for evaluating model's ability to distinguish healthy and diseased regions.

## 3.6 Training Strategy and Model Checkpointing

**Training Loop :**

- **Optimizer:** Adam optimizer with learning rate $10^{-4}$.
- **Learning Rate Scheduler:** Learning rate decay when validation loss plateaus.

**Checkpoints and TensorBoard Visualization :**

- **Model Checkpointing:** Models saved based on validation AUPR scores.
- **TensorBoard:** Real-time insights into model's training process.

## 3.7 Experimental Results and Observations

- The UNet model with MSE loss and CLAHE enhancement achieved an AUC of 91% and an AUPR of 55%.

- Green-channel-only models outperformed full RGB models due to increased sensitivity to retinal structures.
- The U-Net with VGG16 Backbone and Categorical Focal Loss (CFL) achieved a Dice Score of 94.959% and an IoU Score of 90.995%, indicating highly accurate and precise segmentation of retinal lesions.

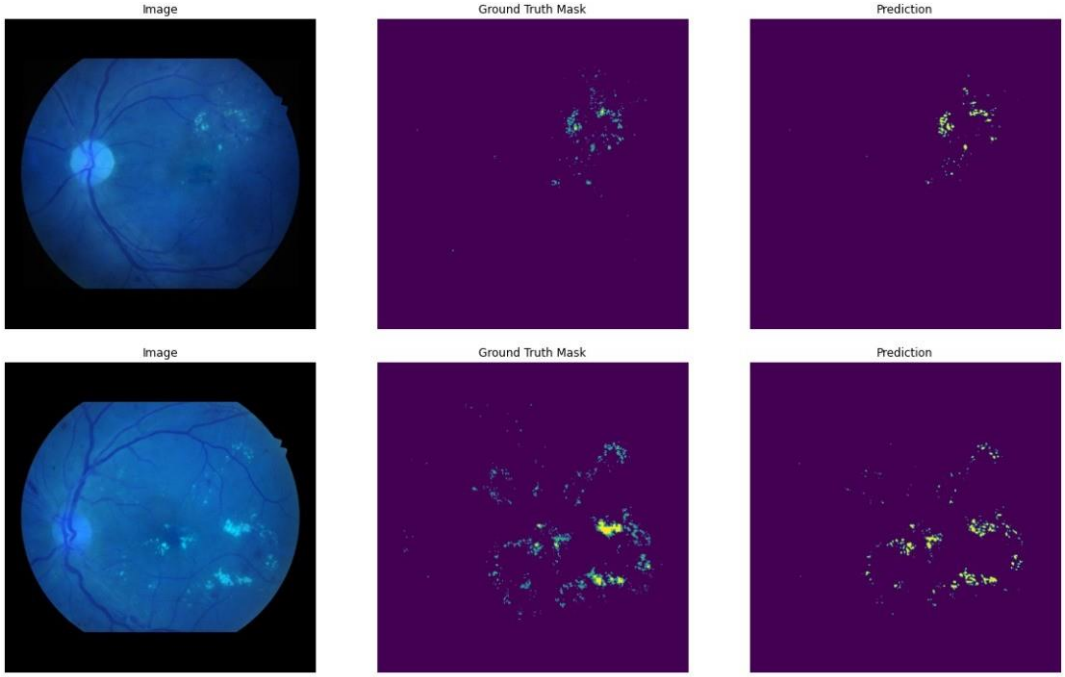


Figure 3.4: Segmentation results

| **Model** | **Loss Function** | **Green-Only AUPR** | **RGB AUPR** | **AUC** |
|---|---|---|---|---|
| UNet with No Random Image Enhancer | UNet MSELoss | 56% | 42% | 90% |
| UNet with With Random Image Enhancer (Color-Jitter) and CLAHE | UNet MSELoss | 56% | 34% | 87% |
| UNet with CLAHE Only | UNet MSELoss | 55% | 46% | 91% |
| U-Net with VGG16 Backbone | CFL | 68% | 53% | 93% |

Table 3.1: Results for segmentation using different models

# Chapter 4

# Classification

## 4.1 Dataset

The dataset used in this experiment is the APTOS 2019 Blindness Detection dataset, which contains retinal images labeled by five severity levels:

- 0: No Diabetic Retinopathy (DR)
- 1: Mild
- 2: Moderate
- 3: Severe
- 4: Proliferative

The images were resized to $512 \times 512$ pixels. The dataset was split into training and testing sets with a ratio of 67% for training and 33% for testing.

## 4.2 Data Preprocessing

To prepare the dataset, the following preprocessing steps were performed:

- **Resizing:** All images were resized to $256 \times 256$ for consistent input size.
- **Normalization:** Pixel values were scaled to the range $[0, 1]$.
- **Data Augmentation:** Random transformations such as rotations, flips, and brightness adjustments were applied to enhance model robustness.

- **Splitting:** The dataset was divided into training, validation, and test sets.

## 4.3 Models Used

### 4.3.1 Custom Convolutional Neural Network (CNN)

**Architecture:**

- **Convolutional Layers:** Three blocks of convolution layers with increasing filters (32, 64, 128, 256), each followed by ReLU activations and MaxPooling. These layers progressively capture complex features in the image.
- **Flattening:** After the convolutional layers, the output is flattened into a single vector.
- **Fully Connected Layers:** Three dense layers (with 1024, 512, and 6 units, respectively) further refine and condense the features for final classification. The last layer outputs scores for 6 classes.
- **Forward Pass:** Data flows sequentially through each layer, transforming input images into class probabilities for classification.

**Training Strategy:**

- **Loss Function:** Sparse categorical cross-entropy.
- **Optimizer:** Adam optimizer with early stopping.
- **Performance:** Effective in detecting mild and moderate cases, struggled with subtle differences in higher severity classes.

### 4.3.2 ResNet-34

**Architecture:**

- 34 layers with residual blocks to address the vanishing gradient problem.
- Convolutional layers with downsampling and a global average pooling layer.

**Training Strategy:**

- **Transfer Learning:** Pre-trained on ImageNet and fine-tuned on the DR dataset.
- **Loss Function:** Categorical cross-entropy with Adam optimizer and learning rate scheduling.
- **Performance:** Strong performance across all classes, particularly in distinguishing mild and moderate cases.

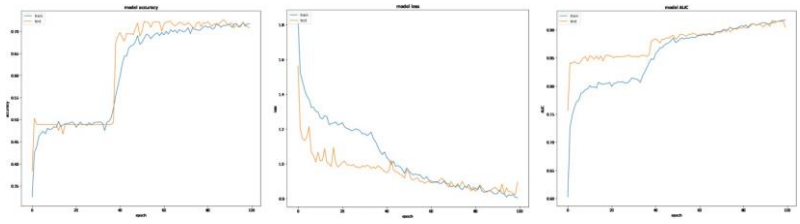

Figure 4.1: Evaluation Metric Plots

### 4.3.3 Vision Transformer (ViT) Model

**Architecture:**

- **Model Summary:** The model consists of 19 layers, including multiple multi-head attention, dense, and dropout layers. It combines a patch-based encoder-decoder architecture with attention mechanisms to effectively process image data.
- **Patch Creation:** The input image is divided into non-overlapping patches. Each patch is treated as a token and passed to the next layer.
- **Patch Encoding:** Each patch is projected into a higher-dimensional space and combined with positional embeddings to retain spatial information.
- **Transformer Blocks:** The encoded patches are processed through multiple Transformer layers. Each block consists of multi-head self-attention and MLP layers, with residual connections and layer normalization to stabilize learning.
- **Flattening and MLP Head:** After processing through Transformer blocks, the output is flattened. A final MLP head refines the features for classification.

- **Output Layer:** The final layer produces class probabilities using a softmax activation function.
- **Grad-CAM:** We use Grad-CAM to generate visual explanations of the model's decisions by highlighting regions in the input image that contribute most to the classification. This interpretability tool helps in understanding how the model processes and categorizes images.

**Training Strategy:**

- **Loss Function:** Categorical cross-entropy.
- **Optimizer:** Adam optimizer with dropout and data augmentation.
- **Performance:** ViT excelled in detecting severe cases where global context across the retina was critical.

## 4.4 Performance Results

Each model was evaluated using standard metrics:

- **Accuracy:** Overall accuracy across severity levels.
- **Precision, Recall, F1 Score:** Class-specific performance, essential for imbalanced classes.
- **Cohen's Kappa Score:** Agreement between predictions and labels, adjusted for chance.

### 4.4.1 Model Performance Summary

- **Custom CNN:** Moderate accuracy; effective for mild and moderate cases but limited success with severe cases.
- **ResNet-34:** Strong performance across all classes; effectively distinguished mild from moderate cases with complex pattern recognition.
- **Vision Transformer (ViT):**

- **Test Accuracy:** 87.62%
- **Test AUC:** 91.05%

ViT achieved the highest accuracy, particularly in recognizing severe and proliferative cases due to its self-attention mechanism, capturing global context effectively.

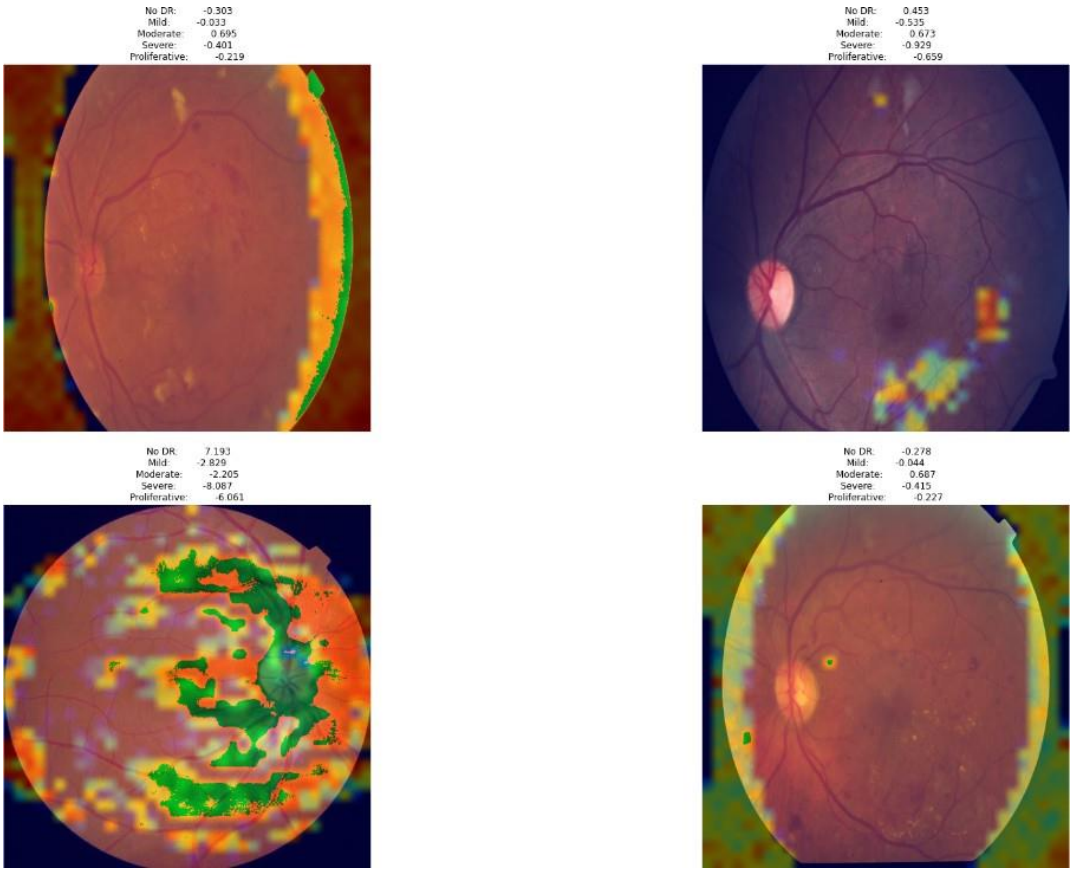


Figure 4.2: Classification results

## 4.5 Results

In summary:

- **Custom CNN:** Served as a baseline model with decent performance, but limited depth for severe cases.
- **ResNet-34:** Benefited from residual connections and pre-trained weights, allowing robust feature extraction.

- **Vision Transformer (ViT):** Achieved the highest accuracy, excelling in higher severity cases by leveraging self-attention to capture the global retinal context.

# Chapter 5

# CONCLUSION

In this project, we developed a deep learning-based model for the segmentation and classification of diabetic retinopathy (DR) abnormalities, aiming to assist in early detection and grading of DR severity. Using the IDRiD dataset, which contains annotated retinal images with key abnormalities, our work addressed challenges such as class imbalance and the wide variability in lesion sizes. We applied extensive preprocessing techniques—including CLAHE enhancement on the green channel and data augmentation methods—to improve model generalization and enhance lesion visibility, crucial for precise segmentation and accurate classification.

The segmentation component utilized architectures like UNet and GCN, selected for their ability to capture detailed retinal features. With targeted loss functions, such as Dice and Focal loss, we effectively managed the class imbalance, enabling high performance in detecting abnormalities like microaneurysms, hemorrhages, and exudates. For classification, we employed CNN-based models with transfer learning, focusing on grading the severity of DR by classifying different stages of the disease based on segmented features. This two-step approach—first segmenting abnormalities and then classifying DR severity—provided a comprehensive assessment framework that can serve as a powerful tool in DR screening.

Our experimental results showed that models trained with only the green channel often outperformed RGB models, highlighting the green channel's greater sensitivity to retinal structures. The segmentation models achieved high AUC and AUPR scores, while the classification component demonstrated robustness in identifying different DR stages, which is crucial for guiding treatment decisions. This work underscores the potential of

deep learning models in providing accurate, automated analysis for DR, paving the way for scalable, reliable screening solutions in clinical settings. Ultimately, our project advances the field of DR diagnostics, promoting early intervention and improved patient outcomes by leveraging both segmentation and classification for holistic disease assessment.

# REFERENCES


[1] Nitigya Sambyal, Poonam Saini, Rupali Syal, Varun Gupta, ”Modified U-Net architecture for semantic segmentation of diabetic retinopathy images,” *Author links open overlay panel*.

[2] Devaraj DRS, Prasanna Kumar SC, ”A survey on segmentation of exudates and microaneurysms for early detection of diabetic retinopathy,” *Mater Today Proc*, 2018; 10845–50.

[3] Joshi S, Karule PT, ”A review on exudates detection methods for diabetic retinopathy,” *Biomed Pharmacother*, 2017; 97: 1454–60.

[4] Wan S, Liang Y, Zhang Y, ”Deep convolutional neural networks for diabetic retinopathy detection by image classification,” *Comput Electr Eng*, 2018; 72:274–82.

[5] Shanthi T, Sabeenian RS, ”Modified Alexnet architecture for classification of diabetic retinopathy images,” *Comput Electr Eng*, 2019; 76: 56–64.

[6] Chudzik P, Al-diri B, Caliv F, Ometto G, Hunter A, ”Exudates segmentation using fully convolutional neural network and auxiliary codebook,” *40th Annu Int Conf IEEE Eng Biol Soc*, 2018.

[7] Qomariah D, Tjandrasa H, ”Exudate detection in retinal fundus images using combination of mathematical morphology and reny entropy thresholding,” *Int Conf Inf Commun Technol Syst*, 2017.

[8] Lin, K.Y., Hsih, W.H., Lin, Y.B., Wen, C.Y. and Chang, T.J., 2021. Update in the epidemiology, risk factors, screening, and treatment of diabetic retinopathy.

[9] S. Qummar, F. G. Khan, S. Shah, A. Khan, S. Shamshirband et al., “A deep

learning ensemble approach for diabetic retinopathy detection," IEEE Access, vol. 7, pp. 150530–150539, 2019.

[10] E. O. Rodrigues, A. Conci and P. Liatsis, "ELEMENT: Multi-modal retinal vessel segmentation based on a coupled region growing and machine learning approach," IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 12, pp. 3507–3519, 2020.

[11] L. Pan, Z. Zhang, S. Zheng and L. Huang, "MSC-Net: Multitask learning network for retinal vessel segmentation and centerline extraction," Applied Sciences, vol. 12, no. 1, pp. 1–10, 2022.