A project report on

Early Stage of Diabetic Retinopathy in Retinal Images Using Efficient Dense Net

Submitted in partial fulfillment for the award of the degree of

Bachelor of Technology in Computer Science and Engineering with Specialization in Artificial Intelligence and Machine Learning

by

Parasu Sai Nikhil (20BAI1217)

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

April 2024

Early Stage of Diabetic Retinopathy in Retinal Images Using Efficient Dense Net

Submitted in partial fulfillment for the award of the degree of

Bachelor of Technology in Computer Science and Engineering with Specialization in Artificial Intelligence and Machine Learning

by

Parasu Sai Nikhil (20BAI1217)

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

April 2024

VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

DECLARATION

I hereby declare that the thesis entitled **"Early Detection of Diabetic Retinopathy for Retinal images using Efficient Dense Net"** submitted by me, for the award of the degree of Bachelor of Technology in Computer Science and Engineering with Specialization in Artificial Intelligence and Machine Learning, Vellore Institute of Technology, Chennai is a recoed of bonafide work carried out by me under the supervision of Dr. Elakiya E**.**

I further declare that the work reported in this thesis has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma of this institute or of any other institute or university.

Place: Chennai

Date: Signature of the Candidate

PARASU SAI NIKHIL (20BAI1217)

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Computer Science and Engineering

CERTIFICATE

This is to certify that the report entitled **"Early Detection of Diabetic Retinopathy for Retinal images using Efficient Dense Net"** is prepared and submitted by **Parasu Sai Nikhil (20BAI1217)** to Vellore Institute of Technology, Chennai in partial fulfillment of the requirement for the award of the degree of **Bachelor of Technology in Computer Science and Engineering with Specialization in Artificial Intelligence and Machine Learning** programme is a bonafide record carried out under my guidance. The project fulfills the requirements as per the regulations of this University and in my opinion meets the standards for submission. The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma and the same is certified.

Signature of the Guide:

Name: Dr. Elakiya E.

Date:

Signature of the Internal Examiner

Name:

Date:

Signature of the External Examiner

Name:

Date:

Approved by Head of Department,

B.Tech.CSE with Specialization in Artificial Intelligence and Machine Learning

Name: Dr. Sweetlin Hemalatha C

Date:

(Seal of SCOPE)

ABSTRACT

Diabetic retinopathy, a common complication of diabetes, is a serious eye condition that impacts the retina, the back of the eye's light-sensitive tissue. With the progression of diabetes, prolonged exposure high blood sugar can damage the retina's tiny blood vessels, leading to various vision-related problems. Notably, diabetic retinopathy stands as a significant cause of blindness among adults of working age globally. Timely identification and intervention are essential to prevent irreversible vision loss, as once vision is compromised, recovery becomes challenging. Beyond vision impairment, diabetic retinopathy often serves as an indicator of potential complications such as macular edema and retinal detachment. This emphasizes the need for comprehensive management strategies to effectively guard against the multifaceted threats posed by this condition. In addressing the critical challenge of early detection and classification in diabetic retinopathy (DR), contemporary technological advancements offer promising solutions. Deep learning, a subset of artificial intelligence, emerges as a potent tool capable of revolutionizing the identification of DR at its nascent stages. Leveraging the power of deep learning algorithms, in this paper we aim to construct a highly accurate model for the early classification of DR.

ACKNOWLEDGEMENT

It is my pleasure to express with deep sense of gratitude to Dr. Elakiya E, Assistant Professor, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, for her constant guidance, continual encouragement, and understanding; more than all, she taught me patience in my endeavor. My association with her is not confined to academics only, but it is a great opportunity on my part of work with an intellectual and expert in the field of Deep Learning.

It is with gratitude that I would like to extend my thanks to the visionary leader Dr. G. Viswanathan our Honorable Chancellor, Mr. Sankar Viswanathan, Dr. Sekar Viswanathan, Dr. G V Selvam Vice Presidents, Dr. Sandhya Pentareddy, Executive Director, Ms. Kadhambari S. Viswanathan, Assistant Vice-President, Dr. V. S. Kanchana Bhaaskaran Vice-Chancellor, Dr. T. Thyagarajan Pro-Vice Chancellor, VIT Chennai and Dr. P. K. Manoharan, Additional Registrar for providing an exceptional working environment and inspiring all of us during the tenure of the course.

Special mention to Dr. Ganesan R, Dean, Dr. Parvathi R, Associate Dean Academics, Dr. Geetha S, Associate Dean Research, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai for spending their valuable time and efforts in sharing their knowledge and for helping us in every aspect.

In jubilant state, I express ingeniously my whole-hearted thanks to Dr. Sweetlin Hemalatha C, Head of the Department, B.Tech. CSE with Specialization in Artificial Intelligence and Machine Learning and the Project Coordinators for their valuable support and encouragement to take up and complete the thesis.

My sincere thanks to all the faculties and staffs at Vellore Institute of Technology, Chennai who helped me acquire the requisite knowledge. I would like to thank my parents for their support. It is indeed a pleasure to thank my friends who encouraged me to take up and complete this task.

Place: Chennai

Date:

Parasu Sai Nikhil

CONTENTS

DATASET DESCRIPTION

CHAPTER 4:

PROPOSED METHODOLOGY

CHAPTER 5

RESULTS AND DISCUSSIONS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1.1 INTRODUCTION TO DIABETIC RETINOPATHY:

Diabetes patients are at risk for developing diabetic retinopathy (DR), a progressive eye condition that impairs vision overall and eventually results in blindness. Diabetes is becoming more and more commonplace globally, making DR a serious public health issue. It is defined by damage to the retina's blood vessels, which leads to a variety of retinal abnormalities that impair vision. These lesions, which contribute to the disease's progression and visual impairment, include microaneurysms, hemorrhages, hard exudates, and neovascularization. Grading schemes are often used to determine the severity of diabetic retinopathy (DR). The Early Treatment Diabetic Retinopathy Study (ETDRS) severity scale is the most widely used approach. This scale divides diabetic retinopathy (DR) into five stages: stage 0 (no retinopathy) to stage 5 (advanced proliferative diabetic retinopathy (PDR) with high-risk characteristics).

The burden of DR is still significant even with advancements in treatment procedures including intravitreal injections and laser therapy, especially in low- and middle-income nations where access to eye care services may be restricted. Consequently, the demand for effective and scalable screening techniques to identify and handle DR is rising, particularly in environments with limited resources.

Artificial intelligence (AI) and imaging technology advancements in recent years have demonstrated encouraging promise to revolutionize DR detection and management. Deep learning algorithms for automated interpretation of retinal pictures have become a practical tool for early diagnosis and categorization of diabetic retinal lesions. By using fundus images, these algorithms may detect minute alterations in the retina that are suggestive of diabetic retinopathy.

Diabetic retinopathy (DR) is not only a concern for individual patients but also poses significant economic burdens on healthcare systems worldwide. The cost of managing DR-related complications, such as vision loss and blindness, is substantial and can strain healthcare resources, particularly in regions where access to specialized eye care services is limited.

Furthermore, the societal and personal costs associated with vision impairment are profound, affecting individuals' quality of life and productivity.

In response to these challenges, researchers and innovators are continuously striving to enhance the efficiency and accessibility of DR screening and management. One promising avenue is the integration of telemedicine and mobile health technologies. Teleophthalmology initiatives leverage digital platforms and remote imaging devices to facilitate screening and diagnosis, enabling healthcare providers to reach underserved populations and streamline the delivery of care. Mobile applications equipped with AI algorithms can analyze retinal images captured by smartphone cameras, empowering patients to monitor their eye health and seek timely intervention when needed.

Collaborative efforts between healthcare professionals, technologists, and policymakers are essential to ensure the successful implementation of these innovations. This includes addressing regulatory and reimbursement barriers, enhancing infrastructure for data sharing and communication, and promoting interdisciplinary training to equip healthcare workers with the skills needed to leverage new technologies effectively.

1.1.1 DEEP LEARNING TECHNIQUES IN DIABETIC RETINOPATHY DETECTION:

Medical image analysis has undergone a revolution because to deep learning techniques, including Convolutional Neural Networks (CNNs) and Residual Networks (ResNet). These techniques have revolutionized the identification and categorization of diabetic retinopathy (DR). These models are exceptionally good at learning hierarchical data representations, which enables them to accurately extract complex patterns and characteristics from fundus images. CNNs and ResNet can distinguish between normal and aberrant features associated with different stages of DR by examining the spatial connections and textures within retinal pictures.

The ability of deep learning models to effectively adapt and generalize to new data is one of its main advantages. These models are trained on sizable datasets annotated with ground truth labels through supervised learning, which enables them to discover intricate patterns and connections between picture features and disease symptoms. Consequently, retinal pictures can be reliably classified by CNNs and ResNets into various degrees of DR severity, giving doctors important information for prompt diagnosis and intervention. Deep learning techniques, particularly

Convolutional Neural Networks (CNNs) and Residual Networks (ResNets), have not only revolutionized the field of medical image analysis but have also significantly impacted the detection and management of diabetic retinopathy (DR). These advanced algorithms excel in extracting intricate features from complex retinal images, enabling accurate identification and categorization of DR lesions.

One of the key strengths of deep learning models lies in their ability to learn hierarchical representations of data. CNNs and ResNets are adept at capturing subtle spatial connections and textures within retinal images, allowing them to discern between normal retinal features and those indicative of DR pathology. By analyzing these features at multiple levels of abstraction, these models can effectively differentiate between different stages of DR severity, providing clinicians with valuable insights for timely intervention and management. A critical advantage of deep learning techniques is their adaptability and generalizability to new data. Through supervised learning on large annotated datasets, these models can learn complex patterns and correlations between image features and disease manifestations. This robust training process equips CNNs and ResNets with the capability to reliably classify retinal images into various degrees of DR severity, even when presented with unseen data. This capability is particularly valuable in clinical practice, where accurate and consistent diagnosis is essential for guiding patient care.

1.1.2 APPROACH AND INVESTIGATION:

In our approach and investigation, we aim to conduct a comprehensive evaluation of various deep learning models for automated diabetic retinopathy detection and classification. Specifically, we will assess the performance of Convolutional Neural Networks (CNNs) and ResNet architectures to determine their efficacy in extracting complex features from fundus images. Our methodology involves several key steps. First, we will curate a diverse dataset of retinal images annotated with ground truth labels indicating different stages of diabetic retinopathy severity. This dataset will be meticulously curated to ensure representation across various demographics and disease manifestations, enabling robust model training and evaluation.

Next, we will train multiple CNN and ResNet architectures using state-of-the-art techniques and optimization algorithms. Through supervised learning, these models will learn to identify and classify retinal abnormalities associated with diabetic retinopathy, leveraging the rich feature

representations learned from the input images. To assess the performance of each model, we will employ rigorous validation protocols, including cross-validation and hold-out validation. This will allow us to evaluate the models' accuracy, sensitivity, specificity, and other performance metrics on both training and validation datasets, ensuring robustness and generalizability.

Furthermore, we will explore the development of a hybrid model that combines the strengths of CNNs and ResNets. By leveraging complementary characteristics and architectural innovations, such as skip connections and residual learning, we aim to enhance the diagnostic resilience and accuracy of automated diabetic retinopathy detection. This hybrid approach seeks to capitalize on the unique capabilities of each model architecture, potentially surpassing the performance of individual models alone. Throughout the investigation, we will closely monitor model performance and iteratively refine our methodology to address any challenges or limitations encountered. Additionally, we will conduct comparative analyses to identify the strengths and weaknesses of each model architecture, informing future research directions and clinical applications.

Ultimately, our goal is to advance the field of diabetic retinopathy detection through cutting-edge deep learning techniques, ultimately improving patient outcomes and enhancing the efficiency of healthcare delivery.

1.2 OVERVIEW OF MEDICAL IMAGES

Medical imaging plays a pivotal role in modern healthcare, offering invaluable insights into the structure and function of the human body. These imaging techniques enable healthcare professionals to visualize internal organs, tissues, and physiological processes, facilitating accurate diagnosis and treatment planning across a wide spectrum of medical conditions. X-rays, one of the oldest and most widely used imaging modalities, provide detailed images of bones and dense tissues. They are particularly useful for detecting fractures, assessing joint abnormalities, and diagnosing conditions such as pneumonia and lung cancer. Computed tomography (CT) scans, on the other hand, offer cross-sectional images of the body, allowing for detailed visualization of soft tissues, blood vessels, and organs. CT scans are instrumental in diagnosing conditions ranging from head trauma and abdominal disorders to cardiovascular diseases.

Magnetic resonance imaging (MRI) utilizes powerful magnets and radio waves to produce detailed images of soft tissues, organs, and neurological structures. MRI is prized for its ability to provide high-resolution images without exposing patients to ionizing radiation, making it especially valuable for diagnosing brain and spinal cord disorders, musculoskeletal injuries, and tumors. Ultrasound imaging utilizes sound waves to generate real-time images of internal organs and tissues. It is widely used in obstetrics and gynecology for monitoring fetal development and detecting abnormalities during pregnancy. Additionally, ultrasound is employed in assessing cardiovascular health, diagnosing abdominal conditions, and guiding minimally invasive procedures such as biopsies and injections.

Optical coherence tomography (OCT) is a non-invasive imaging technique that produces cross-sectional images of tissues at micrometer resolution. It is commonly used in ophthalmology for assessing retinal health, diagnosing conditions like macular degeneration and glaucoma, and monitoring treatment outcomes in diabetic retinopathy. Each imaging modality offers unique advantages and is selected based on the clinical indication and specific requirements of the patient's condition. By leveraging these advanced imaging technologies, healthcare professionals can accurately diagnose diseases, monitor treatment progress, and optimize patient outcomes, thereby improving the quality of care and advancing medical knowledge.

1.2.1 TYPES OF MEDICAL IMAGES

Medical imaging plays a crucial role in diagnosing and monitoring various health conditions, providing clinicians with valuable insights into the internal structures and functions of the human body. There are several modalities of medical imaging, each offering unique advantages and insights into different aspects of anatomy and pathology. X-ray imaging, for example, is commonly used to visualize skeletal structures and detect abnormalities such as fractures, dislocations, and pulmonary diseases. Magnetic Resonance Imaging (MRI) utilizes powerful magnets and radio waves to generate detailed images of soft tissues, organs, and blood vessels, making it particularly useful for diagnosing neurological disorders, musculoskeletal injuries, and cardiovascular conditions. Computed Tomography (CT) imaging combines X-ray technology with computer processing to create cross-sectional images of the body, enabling precise localization of tumors, internal bleeding, and other abnormalities. In the context of diabetic retinopathy diagnosis, fundus photography is the primary imaging modality used to capture high-resolution images of the retina, allowing clinicians to assess the presence and severity of retinal lesions associated with the disease. Each modality offers distinct advantages and limitations, and the choice of imaging technique depends on factors such as the suspected pathology, patient's condition, and availability of equipment and expertise.

1.2.2 PREPROCESSING FOR IMPROVED RESULTS:

Preprocessing plays a crucial role in enhancing the quality and reliability of medical image analysis, particularly when leveraging deep learning algorithms for interpretation. By applying various preprocessing techniques, we can mitigate common challenges associated with medical images, ensuring accurate and robust performance of deep learning models. One essential preprocessing step is image registration, which involves aligning multiple images acquired from different modalities or time points to a common coordinate system. This process is essential for comparing images, tracking disease progression, and facilitating accurate diagnosis and treatment planning. Normalization is another critical preprocessing technique used to standardize the intensity values and spatial resolution of medical images. Normalization ensures consistency across images, reducing variability and improving the reliability of deep learning algorithms for

image analysis.

Contrast enhancement techniques are employed to improve the visibility of anatomical structures and pathological features in medical images. By adjusting the image's contrast and brightness levels, we can enhance subtle details and improve the interpretability of deep learning models. Noise removal is essential for reducing artifacts and enhancing image clarity. Medical images are often affected by various sources of noise, including electronic noise, motion artifacts, and environmental interference. Preprocessing methods such as filtering and denoising algorithms can effectively suppress noise while preserving important image features. Furthermore, preprocessing methods can address specific challenges encountered in medical imaging, such as motion artifacts, uneven lighting, and variations in image parameters. By carefully preprocessing images to correct these issues, we can ensure the optimal performance of deep learning models for tasks such as disease detection, segmentation, and classification.

Overall, preprocessing plays a critical role in preparing medical images for deep learning analysis. By applying techniques such as image registration, normalization, contrast enhancement, and noise removal, we can improve the quality, consistency, and interpretability of medical images, ultimately enhancing the efficacy of deep learning algorithms for medical image analysis and clinical decision-making.

1.2.3 ADVANTAGES OF USING DEEP LEARNING ON MEDICAL IMAGES:

Deep learning offers numerous advantages for the analysis of medical images, particularly in tasks such as segmentation, classification, and anomaly detection. One key advantage is the ability of deep learning models to automatically learn hierarchical representations of image data. Unlike traditional machine learning techniques that require manual feature extraction, deep learning models can extract complex features directly from raw image data, enabling more accurate and efficient analysis. Furthermore, deep learning models exhibit state-of-the-art performance in medical image analysis tasks due to their capacity to capture intricate patterns and structures within images. These models can learn from large datasets of annotated medical images, allowing them to identify subtle abnormalities and variations that may be challenging for human observers to detect.

Moreover, deep learning models offer scalability and flexibility, making them well-suited for handling diverse medical imaging modalities and applications. Whether it's X-rays, CT scans,

MRI images, or histopathological slides, deep learning algorithms can be adapted and trained to analyze various types of medical images, providing valuable insights across different clinical specialties. Another advantage of deep learning in medical image analysis is its potential for continuous improvement and adaptation. As new data becomes available and models are refined, deep learning algorithms can evolve to enhance their performance and accuracy. This adaptability is particularly valuable in the dynamic field of medicine, where new imaging technologies and diagnostic challenges constantly emerge.

Additionally, deep learning-based approaches often offer faster processing times compared to traditional methods, enabling real-time or near-real-time analysis of medical images. This rapid analysis can expedite diagnosis, treatment planning, and patient management, leading to improved clinical outcomes and efficiency in healthcare delivery. Overall, the advantages of using deep learning on medical images extend beyond improved accuracy and efficiency. These models have the potential to revolutionize medical imaging by automating complex tasks, enhancing diagnostic capabilities, and ultimately improving patient care across a wide range of clinical applications.

1.2.4 CHALLENGES IN MEDICAL IMAGES

Medical imaging, while indispensable in modern healthcare, presents a multitude of challenges that can impede its effectiveness and reliability in clinical practice. One of the foremost challenges is the inherent variability in image quality across different modalities and imaging devices. Factors such as equipment calibration, patient movement during scanning, and variations in tissue properties can introduce noise and artifacts into images, complicating interpretation and analysis. Moreover, the complexity and diversity of anatomical structures among patients pose challenges in image interpretation. Normal anatomical variations, as well as pathological changes, may manifest differently in different individuals, leading to potential misdiagnosis or oversight of clinically significant findings. Another critical challenge lies in the scarcity of annotated data required for training and validating deep learning models, particularly in specialized medical domains. Annotating medical images with ground truth labels is a time-consuming and labor-intensive process that often requires expert knowledge and consensus among medical professionals. As a result, datasets for training deep learning algorithms may be limited in size and diversity, hindering the generalizability and robustness of the models.

Moreover, the interpretation of medical images relies heavily on the expertise and experience of radiologists, pathologists, and other healthcare professionals, leading to variability in diagnostic accuracy and treatment recommendations.

The integration of medical imaging technologies into existing healthcare systems also presents significant logistical and technical challenges. Issues related to data management, storage, and sharing must be addressed to ensure the seamless exchange of medical images and patient information between healthcare facilities. Interoperability standards and protocols play a crucial role in enabling the integration of diverse imaging modalities and electronic health record systems, but achieving widespread adoption and compliance can be complex and time-consuming. Furthermore, ensuring the privacy and security of patient data in medical imaging is paramount, given the sensitive nature of healthcare information. Compliance with regulatory frameworks such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in the European Union adds another layer of complexity to the development and deployment of medical imaging systems.

1.3 CURRENT DIAGNOSIS METHODS

The diagnosis of diabetic retinopathy traditionally relies on manual screening by ophthalmologists through the evaluation of fundus photographs. This process involves visually inspecting retinal images to identify characteristic lesions such as microaneurysms, hemorrhages, exudates, and neovascularization. However, manual screening is time-consuming, labor-intensive, and subject to inter-observer variability due to differences in expertise and subjective interpretation. Moreover, the growing prevalence of diabetes worldwide has resulted in an increased demand for diabetic retinopathy screening, exacerbating the challenges associated with the limited availability of trained ophthalmologists. In addition to manual screening, automated image analysis techniques based on traditional image processing methods have been developed to assist in diabetic retinopathy diagnosis. These techniques often rely on feature extraction and classification algorithms to detect and classify retinal lesions. While automation has the potential to improve efficiency and reduce variability, these methods may still suffer from limitations such as suboptimal accuracy, especially in the presence of complex pathologies or image artifacts. As such, there is a pressing need for more robust and accurate diagnostic approaches that can overcome the limitations of existing methods and facilitate early detection and management of diabetic retinopathy.

1.4 PROJECT STATEMENT

- Diabetic retinopathy (DR) poses a serious threat of vision loss for people with diabetes. Late detection often leads to irreversible damage.
- Traditional diagnosis methods can be slow and subjective, increasing the risk of missed cases and delayed treatment.
- So it is very important to identify the diabetic retinopathy in the early stage itself.
- So an advanced AI method is necessary for the detection of diabetic retinopathy in the early stage.

1.5 OBJECTIVES:

- To create and train deep learning models tailored for diabetic retinopathy detection using labeled retina images.
- To implement image enhancement techniques to improve the clarity and quality of retina images, optimizing them for effective model identification.
- Enhance the models to specifically identify early signs of diabetic retinopathy, leveraging improved image clarity for accurate detection in the initial stages of the disease.
- To design a hybrid model based on the top performing models to get the accurate results.

1.6 SCOPE OF THE PROJECT:

The project's scope extends beyond the creation of an AI-based technique for early diabetic retinopathy diagnosis to encompass a comprehensive approach to dataset curation, algorithm development, and evaluation, integration with clinical workflows, regulatory compliance, and scalability. Dataset assembly is a crucial initial step, involving the collection of a diverse range of retinal images representing various degrees of retinopathy across different demographic groups. Preprocessing methods, such as contrast enhancement and noise reduction, will be applied to ensure optimal image quality and facilitate accurate analysis.

Deep learning techniques, particularly convolutional neural networks (CNNs) and their variations, will be employed to analyze retinal images and detect early-stage abnormalities indicative of diabetic retinopathy. Rigorous evaluation of the proposed model's performance will be conducted using various metrics to ensure robustness and generalizability across diverse patient populations and imaging settings. The project also aims to explore the development of a hybrid model that integrates the best features from multiple architectures to further improve diagnostic accuracy. Integration with current clinical workflows is essential, requiring seamless incorporation of the automated detection system into standard screening and diagnostic procedures.

Efforts will be made to obtain approvals for clinical use and ensure regulatory compliance, addressing any legal and ethical considerations associated with deploying the solution in healthcare settings. Additionally, scalability and accessibility will be prioritized to make the solution widely available and user-friendly in various healthcare environments. Ultimately, the overarching goal of the project is to enhance patient outcomes by enabling early detection of diabetic retinopathy, facilitating prompt intervention and treatment. By leveraging advanced AI techniques, robust evaluation methodologies, and integration with clinical workflows, the project aims to make a meaningful impact on the diagnosis and management of this sight-threatening complication, ultimately improving patient care and quality of life.

Chapter 2

Background

2.1 RELATED WORK:

In recent research endeavors aimed at diabetic retinopathy (DR) detection, a wide array of methodologies have been explored, each contributing to the ongoing quest for accurate and automated diagnoses. One notable study, referenced as [1], adopts a convolutional neural network (CNN) approach for DR detection. This study leverages pre-trained models, specifically VGG16 and ResNet50, which are fine-tuned for DR detection tasks. The outcomes of this study are commendable, showcasing impressive accuracy in the binary classification of both normal and DR cases. By harnessing the power of CNNs and fine-tuning pre-trained models, this approach demonstrates promise in accurately identifying cases of DR, thereby facilitating timely intervention and treatment.

Another noteworthy contribution, cited as [2], introduces a sophisticated two-stage classification system for DR detection. This approach employs multiple pre-trained CNNs and utilizes ensemble stacking techniques to enhance classification performance. By leveraging the strengths of ensemble learning, which combines predictions from multiple models to improve accuracy and robustness, this methodology offers a comprehensive solution for DR classification. The two-stage classification system enhances the model's ability to discern subtle differences in retinal images, thereby improving diagnostic accuracy and reliability. Furthermore, reference [3] addresses DR classification using two distinct CNN architectures, namely DenseNet121 and VGG16, on the APTOS dataset. This study highlights the comparative performance of these architectures, with DenseNet121 outperforming VGG16 in terms of accuracy. The superiority of DenseNet121 underscores the importance of selecting appropriate model architectures for DR classification tasks, as the inherent design characteristics of CNN architectures can significantly impact performance.

On a different front, [4] adopts a texture-based approach, combining green channel extraction and Support Vector Machines, with Local Energy-based Shape Histogram (LESH) outperforming Local Binary Pattern (LBP). [5][20] emphasizes the importance of preprocessing techniques, employing resizing and normalization, and finds that the DenseNet 169 is better

performing CNN for DR severity classification. In [6][19] they have created a A patch-based approach using an encoderdecoder neural network. The fundus image is divided into patches, which are then fed into the network for classification, also they have used CLACHE preprocessing technique to enhance the patches in the image.Meanwhile, [7] they have trained their modified CNN model on the private dataset and showed better results. The modifications they made to the CNN are included fine-tuning pre-trained networks and adding custom layers for feature extraction.

In [9], convolutional neural networks (CNNs) and a VGG classifier are utilized for feature extraction, with hyperparameters optimized using genetic algorithms (GA) to enhance performance. This approach leverages the capabilities of deep learning models for extracting discriminative features from retinal images, while the use of GA facilitates the fine-tuning of model parameters to maximize classification accuracy. By combining CNNs and VGG classifiers with GA optimization, this study offers a comprehensive framework for accurate and efficient diabetic retinopathy detection. Meanwhile, [10] focuses on preprocessing techniques to enhance the quality of retinal images before feeding them into a CNN model for classification. Techniques such as green channel extraction, normalization, and morphological operations like sharpening are employed to improve image clarity and feature representation. By optimizing image quality through preprocessing, the CNN model can better capture relevant features for accurate classification, thereby improving overall performance in diabetic retinopathy detection tasks.

Similarly, [11] also emphasizes the importance of preprocessing techniques in improving model performance. Here, images are resized to 256x256 and subjected to histogram equalization to enhance contrast and improve visibility of retinal structures. Fine-tuned VGG and CNN architectures are then employed for model training, leveraging the enhanced image quality to extract meaningful features for diabetic retinopathy classification. This approach underscores the significance of image preprocessing in maximizing the effectiveness of deep learning models for medical image analysis tasks. In contrast, [12] adopts a unique approach by utilizing five pre-trained CNNs - AlexNet, NASNetLarge, VGG-19, Inception V3, and ShuffleNet - to extract features from fundus images. A genetic algorithm is employed to select the most relevant and optimal features from the extracted set, reducing model complexity and enhancing performance. Subsequently, support vector machine (SVM) is used for classification based on the selected

features. This methodological variation showcases the flexibility and versatility of different approaches in addressing the challenges of diabetic retinopathy detection, highlighting the importance of feature selection and model optimization in achieving accurate classification results.

In [13], the authors focus on preprocessing optical coherence tomography (OCT) images to address inconsistencies such as varying brightness and size. These preprocessing steps are crucial for standardizing the input data and ensuring consistency across different OCT images, which may exhibit variations due to factors like imaging conditions and patient characteristics. Additionally, the authors employ a pre-trained deep learning model as a feature extractor, leveraging its ability to capture informative features from the OCT images. However, instead of fine-tuning the entire model, only the final layers are re-trained on the OCT image dataset. This strategy allows for efficient utilization of pre-trained representations while adapting the model to the specific characteristics of OCT images. By selectively retraining the final layers, the model can learn task-specific features relevant to OCT image analysis, ultimately improving performance in subsequent classification tasks.

In [15], the authors undertake a comprehensive investigation into different convolutional neural network (CNN) architectures for diabetic retinopathy (DR) classification. Specifically, they compare the performance of VGG16, ResNet50, and InceptionV3 architectures, fine-tuned for the DR classification task. Notably, the study is limited to binary classification, where the model predicts whether the eye is affected by DR or not, without further subclassifying the severity stages of the disease. Despite this limitation, the research provides valuable insights into the comparative effectiveness of different CNN architectures for DR detection. By fine-tuning these architectures on DR-specific datasets, the authors aim to optimize model performance and evaluate the suitability of each architecture for automated DR screening applications. Overall, the study contributes to the ongoing efforts to develop accurate and efficient deep learning-based systems for diabetic retinopathy diagnosis, laying the groundwork for future advancements in this critical area of medical imaging research.

In [16], a novel method is proposed for classifying retinal diseases in fundus images and identifying affected regions, even in scenarios with limited labeled data. This innovative approach addresses a common challenge in medical image analysis, where obtaining large annotated datasets for training deep learning models may be impractical or cost-prohibitive. By

leveraging techniques such as transfer learning and data augmentation, the proposed method enables effective classification of retinal diseases while mitigating the need for extensive labeled data. Furthermore, the method incorporates region-based analysis to pinpoint affected regions within the fundus images, providing valuable insights for diagnosis and treatment planning.

In [17], a unified framework is introduced for enhancing blood vessel contrast in various medical images. This framework represents a significant advancement in image processing techniques, offering a versatile solution applicable to diverse imaging modalities beyond retinal imaging. By focusing on enhancing vessel contrast through the separation of vessel and non-vessel pixels for individual processing, the method effectively improves the visibility and delineation of blood vessels within medical images. This enhancement is crucial for accurate diagnosis and analysis, particularly in fields such as radiology and cardiology, where blood vessel morphology plays a critical role in disease assessment and treatment planning.

In [18] proposed a two-stage deep learning framework for segmenting the optic disc (OD) and estimating the cup-to-disc ratio (CDR) in retinal fundus images. The first stage segments the OD using a modified U-Net architecture, and the second stage estimates the CDR within the segmented OD region. This method achieved promising results for OD segmentation and CDR estimation, potentially aiding in the diagnosis of glaucoma. Coming to the metrics [21] has receiver operating characteristic (ROC) curve is a graph used to assess the performance of a binary classification model at various thresholds. It plots the true positive rate (TPR), also known as recall, against the false positive rate (FPR) at different points, allowing you to visualize the trade-off between correctly classifying positive and negative cases. This helps you evaluate the model's ability to discriminate between the two classes.

In [22], the authors propose a novel method based on the U-net architecture, augmented with inception modules and residual connections, specifically tailored for automatic hard exudate segmentation in fundus images. This innovative approach targets a significant challenge in medical image analysis – accurately segmenting small and intricate structures, such as hard exudates, within fundus images. By leveraging the U-net framework along with inception modules and residual connections, the proposed method aims to enhance segmentation accuracy, particularly in scenarios with limited data and small target regions. Compared to existing methods, which may struggle to accurately delineate such fine structures, the U-net based approach offers a promising solution for improving segmentation precision and reliability. In

[23], the authors present a hybrid deep learning approach for diabetic retinopathy (DR) classification in retinal images. This approach integrates a convolutional neural network (CNN) with pre-trained VGG models to extract features and classify images. By combining the strengths of both CNNs and VGG models, the proposed hybrid architecture aims to create a robust system for automatic DR detection and classification. The CNN component enables the model to learn hierarchical features directly from the raw image data, while the pre-trained VGG models contribute high-level feature representations learned from large-scale image datasets. Through this synergistic fusion of deep learning techniques, the proposed approach offers a comprehensive and effective solution for addressing the challenges associated with DR detection and classification.

In their study, as detailed in [25], researchers proposed a novel approach for diabetic retinopathy (DR) classification by integrating deep convolutional neural networks (D-CNNs) with image processing techniques. The primary objective of employing image processing techniques was to enhance the performance of D-CNNs in accurately classifying diabetic retinopathy severity levels. The image processing techniques implemented in the study were carefully selected and tailored to address specific challenges encountered in diabetic retinopathy diagnosis. These techniques likely encompassed preprocessing steps such as image enhancement, noise reduction, and feature extraction, aimed at optimizing the quality and relevance of input images for the subsequent D-CNN-based classification task. The authors of the study reported promising results for DR classification, suggesting that the combination of D-CNNs and image processing techniques yielded notable improvements in classification accuracy. However, specific accuracy values were not explicitly mentioned in the study.

2.2 THEMES DISCOVERED

2.2.1 DETECTION OF DR (BINARY OR MULTI CLASS)

A distinction exists among the papers in terms of the classification approach adopted. Some papers focus on binary classification, distinguishing between normal cases and diabetic retinopathy (DR) cases. This simplifies the task to a yes/no decision. On the other hand, other papers delve into multi-class classification, categorizing fundus images into multiple severity stages of DR, typically ranging from 0 to 4.

Binary classification, which distinguishes between normal cases and those with diabetic retinopathy, simplifies the task into a straightforward yes/no decision. This approach is particularly suitable for screening purposes, where the primary goal is to identify individuals at risk of diabetic retinopathy who may require further evaluation or intervention. By focusing on the presence or absence of diabetic retinopathy, binary classification models can efficiently triage patients and prioritize those in need of closer examination by healthcare professionals. On the other hand, multi-class classification involves categorizing fundus images into multiple severity stages of diabetic retinopathy, typically ranging from 0 to 4 according to established grading scales such as the Early Treatment Diabetic Retinopathy Study (ETDRS) severity scale. This approach provides more granular information about the progression and severity of diabetic retinopathy, allowing for more nuanced patient management and treatment planning. Multi-class classification models enable clinicians to stratify patients based on the severity of retinopathy, guiding decisions regarding referral for treatment, monitoring intervals, and intervention strategies.

Both binary and multi-class classification approaches have their advantages and applications. Binary classification is well-suited for population-based screening programs and initial risk stratification, offering simplicity and efficiency. In contrast, multi-class classification provides detailed insights into the severity and progression of diabetic retinopathy, supporting individualized patient care and treatment optimization.

2.2.2 PREPROCESSING TECHNIQUES:

A common theme across papers involves the application of preprocessing techniques to enhance the quality of fundus images. Techniques include color channel extraction, resizing, image sharpening, and data augmentation. Preprocessing techniques play a crucial role in optimizing the quality of fundus images for subsequent analysis in diabetic retinopathy detection studies. These techniques aim to enhance image clarity, reduce noise, and standardize image properties, thereby improving the performance of classification models. Color channel extraction is a fundamental preprocessing step that involves separating the red, green, and blue channels of the fundus images. By isolating specific color channels, researchers can focus on the channel(s) containing the most relevant information for diabetic retinopathy detection. This selective extraction helps enhance the visibility of retinal features and improves the discriminative power of classification models.

Resizing is another essential preprocessing technique used to standardize the dimensions of fundus images. Resizing ensures that all images have uniform spatial resolution, facilitating consistency in feature extraction and model interpretation. Additionally, resizing helps mitigate computational constraints and memory limitations associated with processing large-scale image datasets. Image sharpening techniques are applied to enhance the contrast and sharpness of fundus images, making subtle features and abnormalities more distinguishable. This preprocessing step can involve the application of filters, such as the Laplacian or unsharp mask filter, to accentuate edges and fine details in the images. By enhancing image sharpness, researchers can improve the accuracy of feature extraction and increase the discriminative power of classification models.

2.2.3 DEEP LEARNING ARCHITECTURES:

The use of deep learning architectures, especially convolutional neural networks (CNNs), is there in the literature. Different variations and combinations of CNN architectures, such as VGG16, ResNet, DenseNet, InceptionV3, and custom-designed networks, are employed for feature extraction and classification. VGG16 is a widely used CNN architecture known for its simplicity and effectiveness. It consists of multiple convolutional layers followed by fully connected layers, enabling hierarchical feature extraction and representation learning. Researchers often employ

VGG16 as a baseline model or as part of more complex ensembles for diabetic retinopathy detection tasks. ResNet, or residual networks, introduced a breakthrough architectural innovation with skip connections, enabling the training of very deep neural networks while mitigating the vanishing gradient problem. ResNet architectures, such as ResNet50 and ResNet101, have been successfully applied to diabetic retinopathy detection, demonstrating superior performance in feature extraction and classification tasks.

DenseNet is another CNN architecture that emphasizes feature reuse and dense connections between layers. By connecting each layer to every other layer in a feed-forward manner, DenseNet facilitates feature propagation and gradient flow, leading to efficient parameter utilization and enhanced model performance. Researchers have explored the use of DenseNet architectures for diabetic retinopathy detection, achieving competitive results in classification accuracy. InceptionV3, part of the Inception family of CNN architectures, incorporates parallel convolutional pathways with varying receptive field sizes to capture multi-scale features. This design enables the network to effectively extract both local and global image features, enhancing its discriminative power in classification tasks. InceptionV3 has been utilized in diabetic retinopathy detection studies, demonstrating robust performance across diverse datasets. In addition to pre-designed CNN architectures, researchers also develop custom-designed networks tailored to specific requirements and challenges of diabetic retinopathy detection. These custom architectures may incorporate novel architectural elements, regularization techniques, or domain-specific knowledge to optimize performance and address unique characteristics of retinal images.

2.2.3 MODEL EVALUATION METRICS:

In the evaluation of diabetic retinopathy detection models, researchers commonly employ a range of metrics to assess performance comprehensively. These metrics provide insights into the models' ability to accurately detect and classify diabetic retinopathy across different severity levels. Accuracy is a fundamental metric that measures the overall correctness of the model's predictions, representing the proportion of correctly classified instances among all instances in the dataset. While accuracy provides a general indication of model performance, it may not adequately capture the performance imbalances between classes, especially in the presence of class imbalance.

Sensitivity, also known as recall, quantifies the proportion of true positive cases correctly

identified by the model out of all actual positive cases. It reflects the model's ability to correctly detect diabetic retinopathy cases, minimizing false negatives and ensuring that individuals with the condition are not overlooked. Specificity measures the proportion of true negative cases correctly identified by the model out of all actual negative cases. High specificity indicates that the model effectively identifies non-diabetic retinopathy cases, reducing false positive errors and minimizing unnecessary interventions or referrals.

The F1-score, which is the harmonic mean of precision and recall, provides a balanced measure of a model's performance by considering both false positives and false negatives. It captures the trade-off between precision (the proportion of true positive cases among all predicted positive cases) and recall (sensitivity), offering a single metric to assess overall classification performance. The area under the receiver operating characteristic (ROC) curve (AUC) evaluates the model's ability to discriminate between positive and negative cases across different threshold values. A higher AUC indicates better discrimination performance, with values closer to 1 indicating optimal performance. The ROC curve visualizes the trade-off between true positive rate (sensitivity) and false positive rate across various threshold levels.

2.3 RESEARCH GAPS

2.3.1 LIMITED EXPLORATION OF HYBRID MODELS:

Many papers in the literature review focus on individual deep learning architectures. However, there is a research gap in exploring hybrid models that combine the strengths of multiple deep learning architectures which can provide better results. The utilization of hybrid models allows researchers to leverage the unique strengths and characteristics of different architectures to improve overall performance. By combining complementary features and learning mechanisms from diverse architectures, hybrid models can achieve superior performance in feature extraction, representation learning, and classification tasks. Despite the promising potential of hybrid models, their exploration in the context of diabetic retinopathy detection has been limited. This gap presents an opportunity for researchers to innovate and develop novel approaches that capitalize on the synergies between various deep learning architectures.

Hybrid models can be constructed through various techniques, including ensemble methods, feature fusion, and multi-task learning. Ensemble methods combine predictions from multiple base models to improve robustness and generalization. Feature fusion techniques integrate features extracted from different architectures to enhance representation learning and discriminative power. Multi-task learning frameworks jointly train models on multiple related tasks, leveraging shared representations to improve performance on individual tasks. By exploring hybrid models in the context of diabetic retinopathy detection, researchers can unlock new avenues for improving diagnostic accuracy, scalability, and clinical utility. These models have the potential to enhance early detection, enable personalized treatment strategies, and ultimately improve patient outcomes in diabetic retinopathy management.

2.3.2 LIMITED UTILIZATION OF PREPROCESSING TECHNIQUES:

While many research papers employed common preprocessing techniques such as image resizing, sharpening, and data augmentation, the incorporation of diverse preprocessing methods has the potential to enhance image quality and improve result accuracy. While traditional preprocessing techniques address fundamental aspects of image quality, such as resolution, contrast, and noise, the incorporation of diverse preprocessing methods can offer additional

benefits. For example, techniques such as histogram equalization and gamma correction can enhance the visibility of subtle details and improve overall image contrast, particularly in cases with uneven lighting or low contrast.

Furthermore, advanced denoising algorithms, such as wavelet denoising and non-local means filtering, can effectively suppress noise while preserving important image features. By reducing noise artifacts, these techniques can enhance the clarity and interpretability of retinal images, leading to more accurate detection and classification results. Texture analysis methods, such as Gabor filtering and local binary patterns, offer insights into the textural characteristics of retinal images, enabling the extraction of discriminative features for diabetic retinopathy detection. By incorporating texture-based preprocessing techniques, researchers can capture additional information about the spatial organization and structural patterns present in retinal images, enhancing the robustness and discriminative power of detection models.

2.3.3 INADEQUATE ANALYSIS OF MODEL INTERPRETABILITY:

The interpretability of the developed models might be overlooked in the literature. There is a research gap in providing insights into how interpretable the models are, especially in a clinical context. Investigating methods for explaining model decisions and ensuring transparency in the decision-making process could be a valuable contribution. Interpretability is essential for understanding how deep learning models arrive at their predictions, especially when these predictions influence clinical decisions. In the context of diabetic retinopathy detection, interpretable models can provide valuable insights into the features and patterns driving classification decisions, enabling clinicians to validate model outputs and make informed decisions about patient care.

One approach to improving model interpretability is through the use of attention mechanisms, which highlight regions of input images that are most relevant for making predictions. Attention-based methods allow clinicians to visualize the areas of the retina that contribute most significantly to the classification of diabetic retinopathy, aiding in the interpretation of model decisions and enhancing trust in automated detection systems. Additionally, post-hoc interpretability techniques, such as gradient-based attribution methods and saliency maps, can provide explanations for model predictions by attributing importance scores to input features. These methods offer insights into the factors influencing model decisions and help identify areas

of uncertainty, enabling clinicians to assess the reliability of model outputs.

Chapter 3

Dataset Description

3.1 DATASET DESCRIPTION:

The dataset comprises a substantial collection of high-resolution retina images captured under diverse imaging conditions. Each image has a rating on a scale ranging from 0 to 4, delineating the spectrum of DR severity. The rating scale encompasses the following classes: "NO DR" for the absence of diabetic retinopathy, "Mild" indicating a mild manifestation. Likewise "Moderate" denoting a moderate level of severity, "Severe" signifying a pronounced severity, and "Proliferate DR" characterizing an advanced and proliferative stage of diabetic retinopathy.

Table 1: Distribution of images in each class

Stage name	No of images
No DR (0)	1805
Mild DR (1)	370
Moderate DR (2)	999
Severe DR (3)	193
Proliferative DR (4)	295

Table 1: No of Images in each class

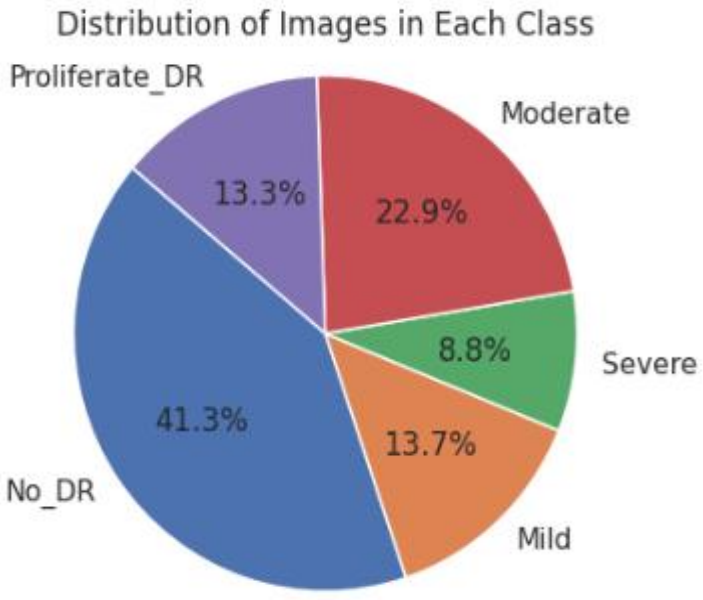

Fig 1: Distribution of images in each class

Figure 2 displays random images representing each level of diabetic retinopathy from the dataset. This curated selection provides a visual representation of the diverse manifestations of diabetic retinopathy across different severity levels. By showcasing images from each stage, the figure offers insight into the varying degrees of retinal pathology associated with diabetic retinopathy, ranging from mild to severe. Each image encapsulates distinct features indicative of its corresponding diabetic retinopathy level, including microaneurysms, hemorrhages, exudates, and neovascularization. These visual cues serve as valuable reference points for clinicians and researchers alike, aiding in the recognition and interpretation of diabetic retinopathy manifestations.

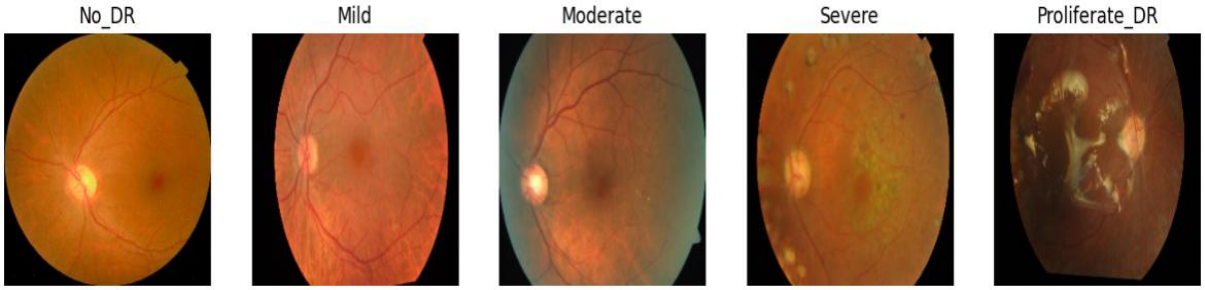

Fig 2: Images from each class

Chapter 4

Proposed Methodology

4.1 IMAGE ENHANCEMENT TECHNIQUES USED

4.1.1GAUSSIAN BLURRING:

Gaussian blurring, a widely used image processing technique, plays a crucial role in reducing noise and enhancing important features in digital images. By convolving the image with a Gaussian filter, Gaussian blurring effectively smooths the image by attenuating high-frequency components while preserving low-frequency information. The Gaussian filter, derived from the Gaussian function, is characterized by its bell-shaped curve, with the central pixel values having the highest weights and gradually decreasing weights towards the edges of the filter. This characteristic ensures that the blurring effect is more pronounced near the central pixel, gradually diminishing towards the surrounding pixels.

One of the key advantages of Gaussian blurring is its ability to reduce noise without significantly affecting image details or edges. Unlike uniform blurring kernels, which can introduce undesirable artifacts and distortions, Gaussian blurring produces more natural-looking results by mimicking the gradual transition of light intensity observed in real-world scenes. Gaussian blurring is particularly effective in applications where noise reduction is essential, such as medical imaging, photography, and computer vision. In medical imaging, for example, Gaussian blurring can improve the visibility of anatomical structures and pathological features by reducing the effects of noise and enhancing contrast. Moreover, Gaussian blurring is often used as a preprocessing step in various image analysis tasks, including edge detection, segmentation, and object recognition. By smoothing the image and reducing noise, Gaussian blurring helps to enhance the performance and robustness of subsequent algorithms, leading to more accurate and reliable results.

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2/2\sigma\hat{}2)}$$

Where:

- x, y: represent the distance from the origin (center of the kernel) in the horizontal and

vertical axes, respectively.

- σ (sigma): represents the standard deviation of the Gaussian distribution, which controls the "spread" of the blur. A larger σ leads to a stronger blur effect.
- π (pi): mathematical constant representing the ratio of a circle's circumference to its diameter, approximately 3.14159.

4.1.2 CANNY EDGE DETECTION:

Canny edge detection is a powerful image processing technique widely used for detecting edges in digital images. It employs a multi-stage algorithm that systematically identifies significant changes in pixel intensity, thereby highlighting the boundaries of objects and features within the image. The first stage of the Canny edge detection algorithm involves smoothing the image to reduce noise and eliminate irrelevant details that could interfere with edge detection. This is typically achieved using a Gaussian filter, which blurs the image while preserving important edges and structures.

Next, the algorithm computes the gradient magnitude and direction at each pixel to identify regions of rapid intensity change, which are indicative of potential edges. This gradient calculation is often performed using techniques such as Sobel or Prewitt operators, which estimate the rate of change of pixel intensity in both horizontal and vertical directions. Following gradient calculation, non-maximum suppression is applied to thin the detected edges and retain only the most significant ones. This step involves suppressing non-maximum pixels in the gradient direction, ensuring that only local maxima along the edges are preserved. Finally, edge tracking by hysteresis is performed to link adjacent edge pixels and form continuous edge contours. This process involves defining two threshold values: a high threshold to identify strong edge pixels and a low threshold to detect weak edge pixels. Pixels with gradient magnitudes above the high threshold are considered strong edges, while those between the high and low thresholds are classified as weak edges. Weak edge pixels are retained only if they are connected to strong edge pixels, ensuring the continuity of edge contours.

The result of the Canny edge detection algorithm is a binary image where edge pixels are represented as white pixels on a black background. These edges serve as key features for various computer vision applications, including object recognition, image segmentation, and boundary detection. Canny edge detection is valued for its robustness, accuracy, and ability to detect edges

even in the presence of noise or cluttered backgrounds. Its versatility and effectiveness make it a fundamental tool in image processing and computer vision, enabling a wide range of applications in fields such as medical imaging, autonomous driving, and robotics.

4.1.3 CLAHE (CONTRAST LIMITED ADAPTIVE HISTOGRAM EQUALIZATION):

CLAHE (Contrast Limited Adaptive Histogram Equalization) is a sophisticated image enhancement technique designed to address the limitations of traditional histogram equalization methods. By intelligently adjusting the intensity levels in local regions of an image, CLAHE effectively enhances contrast while minimizing the risk of over-amplifying noise. The key principle behind CLAHE lies in its adaptive nature, which allows it to tailor the enhancement process to the specific characteristics of each region within the image. To achieve this, CLAHE divides the image into small tiles or patches, typically measuring 8x8 pixels. Histogram equalization is then applied independently to each of these tiles, redistributing the intensity values to achieve a more balanced histogram.

However, to prevent the over-amplification of noise that can occur with traditional histogram equalization, CLAHE incorporates a crucial contrast-limiting mechanism. This mechanism ensures that the enhancement process does not excessively amplify intensity variations, particularly in regions with high local contrast. By imposing a maximum limit on the contrast enhancement for each tile, CLAHE effectively preserves the overall integrity of the image while enhancing visibility of details in both dark and bright areas. Overall, CLAHE's adaptive approach to contrast enhancement makes it particularly well-suited for improving image quality in scenarios where lighting conditions vary or when images contain regions with significantly different contrast levels. By providing enhanced visibility of details across the entire image, CLAHE contributes to more accurate and reliable image analysis in various applications, including medical imaging, remote sensing, and digital photography.

4.1.4 LTP (LOCAL TERNARY PATTERN):

Local Ternary Pattern (LTP) is an image processing technique used for texture analysis and feature extraction. In LTP, each pixel in an image is compared to the average intensity of its neighboring pixels within a defined neighborhood. The comparison results in a ternary pattern for each pixel, indicating whether the pixel is brighter, darker, or similar to the local average intensity. This ternary representation captures important information about the local texture characteristics of the image, including variations in brightness and contrast. By encoding these texture features into ternary patterns, LTP enables robust and discriminative feature extraction, which is valuable for tasks such as object recognition, image classification, and texture segmentation. Additionally, LTP offers computational efficiency and simplicity, making it suitable for real-time applications and large-scale image analysis tasks. Its versatility and effectiveness have led to its widespread adoption in various fields, including computer vision, pattern recognition, and medical imaging, where capturing and analyzing texture information is essential for accurate image interpretation and understanding.

In addition to comparing pixel intensities with the local average, LTP also incorporates a thresholding mechanism to determine whether a pixel is brighter, darker, or similar to its neighbors. This thresholding step enhances the discriminative power of LTP by distinguishing meaningful texture variations from noise or fluctuations in pixel values. Furthermore, LTP can be adapted to different neighborhood sizes and configurations, allowing for flexibility in capturing texture patterns at various scales and orientations. This adaptability makes LTP suitable for analyzing textures of varying complexities, ranging from fine-grained textures in natural scenes to coarse textures in industrial materials. Moreover, LTP is robust to illumination changes and image transformations, ensuring consistent performance across different imaging conditions and environments. Overall, LTP offers a versatile and efficient approach to texture analysis, enabling the extraction of informative features that contribute to the accurate characterization and recognition of textures in digital images. Its simplicity, computational efficiency, and effectiveness make LTP a valuable tool in various applications where texture information plays a crucial role in image understanding and analysis.

$$LTP\ (P, R, \tau) = \sum (i\ =\ 0\ to\ P - 1)\ s(p\_i\ -\ p\_c)\ \times\ 3\^{}i$$

Where:

- P: Number of neighbors considered (typically 8).
- R: Radius of the neighborhood (distance from the center pixel).
- τ (tau): Threshold value.
- p_i: Gray level value of the i-th neighbor.
- p_c: Gray level value of the center pixel.
- s(x): Sign function ($s(x) = \{ 1, \text{ if } x \geq \tau : 0, \text{ if } |x| < \tau : 1, \text{ if } x \leq -\tau\}$)

4.1.5 AHE (ADAPTIVE HISTOGRAM EQUALIZATION):

Adaptive Histogram Equalization (AHE) is a technique employed in image processing to enhance contrast by adjusting the intensity distribution of an image. Unlike traditional histogram equalization methods, which operate globally on the entire image, AHE adapts its transformation locally, considering smaller regions or patches within the image. This adaptive approach allows AHE to effectively address variations in illumination and contrast across different parts of the image. By analyzing and enhancing local regions individually, AHE avoids over-amplification of noise in homogeneous areas while simultaneously improving the visibility of details and edges in both dark and bright regions. This is particularly advantageous in images with uneven illumination or complex textures, where traditional histogram equalization may result in undesirable artifacts or loss of image information.

The adaptive nature of AHE makes it well-suited for a wide range of applications, including medical imaging, satellite imagery, and digital photography. In medical imaging, for example, AHE can enhance the visibility of subtle features in X-ray or MRI scans, leading to improved diagnostic accuracy and interpretation. Similarly, in satellite imagery, AHE can enhance the contrast of aerial photographs, making it easier to identify and analyze land features or objects of interest. Moreover, AHE can be further refined through variations such as Contrast Limited Adaptive Histogram Equalization (CLAHE), which introduces constraints to limit the amplification of local contrast enhancements. This helps to prevent over-enhancement of noise while still effectively improving contrast and visibility. Overall, AHE is a powerful tool for contrast enhancement in digital images, offering a flexible and adaptive approach to addressing variations in illumination and texture. Its ability to locally adjust contrast while preserving image details makes it a valuable technique in various image processing applications, contributing to improved visualization, analysis, and interpretation of digital images.

4.2 DEEP LEARNING ARCHITECTURES USED

4.2.1 CNN (CONVOLUTIONAL NEURAL NETWORK):

Convolutional Neural Networks (CNNs) are a class of deep learning algorithms specifically designed for processing and analyzing visual data. They are inspired by the organization of the visual cortex in animals and are particularly adept at learning hierarchical representations of features from images. CNNs leverage a hierarchical structure of layers, including convolutional layers, pooling layers, and fully connected layers, to automatically extract and learn relevant features from input images. The key innovation of CNNs lies in their ability to learn spatial hierarchies of features through the application of convolutional filters across the input image. These filters, also known as kernels, convolve over the input image, capturing local patterns and features at different spatial scales. Subsequent layers in the network combine and abstract these features, allowing the CNN to progressively learn more complex and abstract representations of the input data.

CNNs have demonstrated remarkable success in a wide range of computer vision tasks, including image classification, object detection, and facial recognition. Their hierarchical architecture enables them to learn discriminative features directly from raw pixel values, without the need for manual feature engineering. In the context of diabetic retinopathy (DR) detection, CNNs have emerged as powerful tools for automated diagnosis and classification. By treating retinal images as grids of pixels, CNNs can effectively capture the spatial relationships between features and learn to distinguish between normal and abnormal retinal characteristics. CNNs automatically learn relevant features associated with DR, such as microaneurysms, hemorrhages, and exudates, which are indicative signs of the disease.

Through extensive training on large datasets of annotated retinal images, CNNs can generalize well to unseen data and accurately classify retinal images into different stages of DR severity. This capability enables early detection of DR, facilitating timely intervention and treatment to prevent vision loss and improve patient outcomes. Overall, CNNs represent a powerful and versatile approach to image analysis, with significant potential for advancing the field of diabetic retinopathy detection and management. Their ability to automatically learn relevant features from retinal images holds promise for improving the efficiency and accuracy of DR screening

programs, ultimately benefiting patients worldwide.

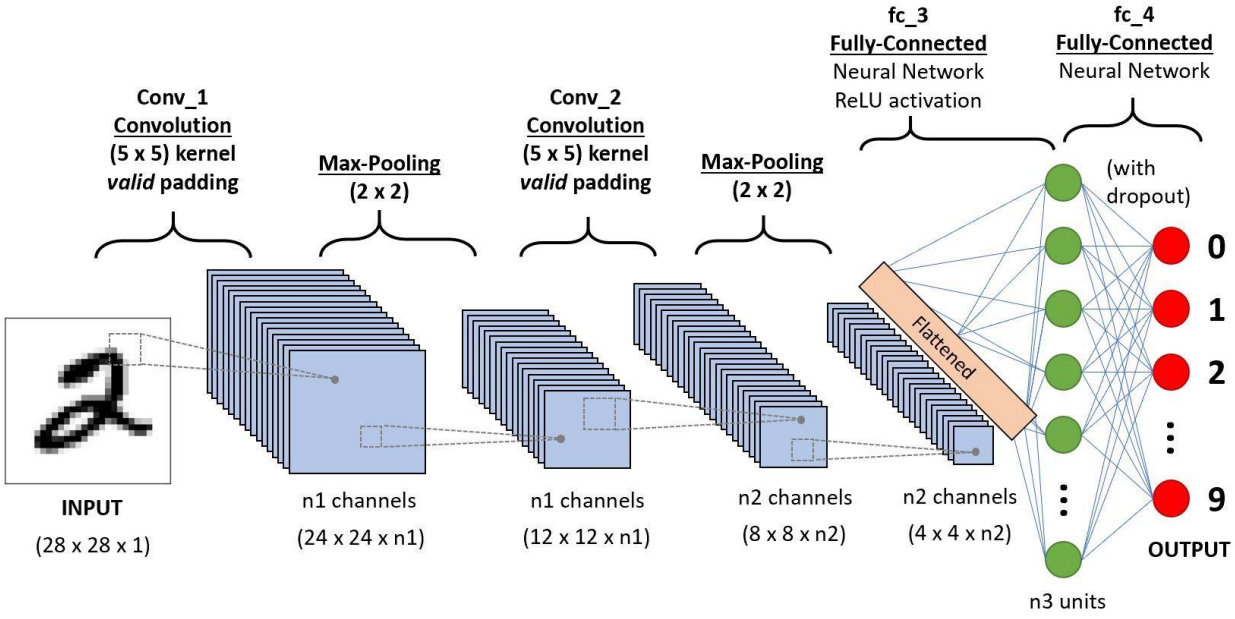

Fig 3: CNN Architecture

4.2.2 RESNET 50:

ResNet-50 is a specific variant of the ResNet architecture, characterized by its depth and effectiveness in training very deep neural networks. ResNet, short for Residual Network, introduced a groundbreaking architectural innovation with the concept of residual connections. These connections allow information to bypass certain layers and be directly propagated to deeper layers, facilitating the training of networks with tens or even hundreds of layers. The key idea behind residual connections is to address the degradation problem that arises when increasing the depth of neural networks. As networks become deeper, they become increasingly difficult to train due to issues such as vanishing gradients and the risk of overfitting. ResNet's residual connections help alleviate these challenges by enabling the model to learn residual mappings, allowing it to skip over certain layers and focus on learning the residual information.

By learning residual mappings, ResNet-50 can effectively train very deep networks while mitigating the risk of overfitting. Overfitting occurs when a model memorizes the training data too well and fails to generalize to unseen data. Residual connections provide shortcuts for gradient flow during training, allowing the model to learn more robust representations and

reducing the likelihood of overfitting. In the context of medical applications like diabetic retinopathy (DR) detection, ResNet-50's ability to effectively train deep networks is particularly valuable. The complex and varied nature of retinal images requires models that can learn hierarchical representations of features across multiple layers. ResNet's residual connections enable the model to capture intricate patterns and abnormalities associated with DR, improving the accuracy and robustness of detection algorithms.

Through extensive training on large datasets of annotated retinal images, ResNet-50 can learn to recognize subtle signs of DR, such as microaneurysms, hemorrhages, and exudates, with high precision and recall. The model's ability to generalize well to unseen data ensures reliable performance in real-world clinical settings, facilitating early detection and intervention to prevent vision loss in patients with diabetes. In summary, ResNet-50's depth and residual connections make it a powerful architecture for diabetic retinopathy detection, enabling the development of accurate and reliable automated screening systems. Its effectiveness in training deep networks is crucial for addressing the complexities of medical image analysis and improving patient outcomes in the diagnosis and management of diabetic retinopathy. The residual blocks represented as:

$$y = F(x, W\_1) + x$$

Where:

- y: Output of the block.
- x: Input of the block.
- F(x, W_1): Represents the transformation applied by the block (e.g., convolution with weights W_1).

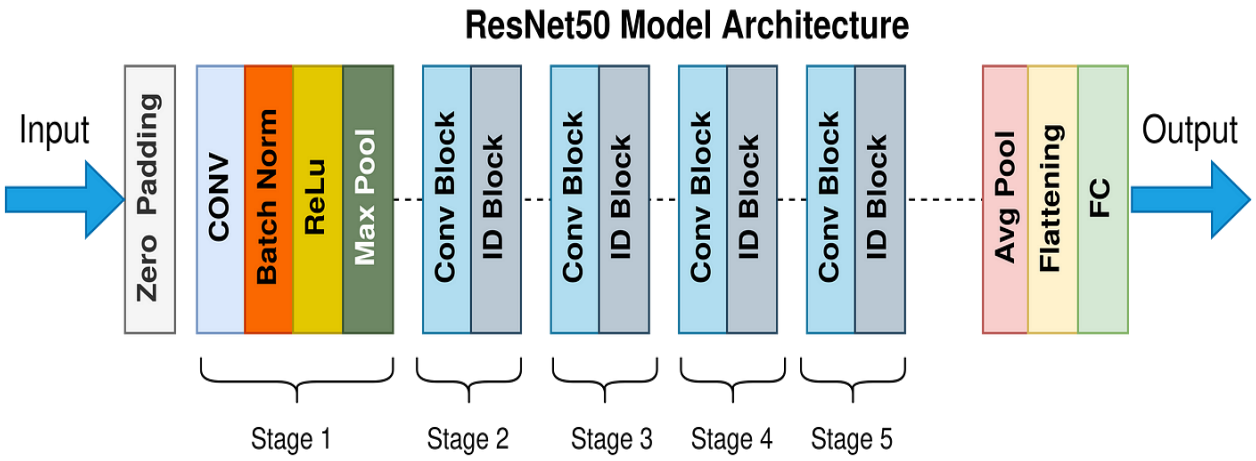

Fig 4: ResNet 50 Architecture

4.2.3 DENSENET121:

DenseNet121 is a variant of the DenseNet architecture, which is characterized by its use of dense blocks. In DenseNet121, each layer receives input from all preceding layers within the same dense block, promoting feature reuse and facilitating efficient learning. This architectural design addresses challenges related to vanishing gradients commonly encountered in very deep neural networks. The key innovation of DenseNet is its dense connectivity pattern, where each layer is directly connected to every other layer within the same dense block. By leveraging dense connections, DenseNet121 promotes feature reuse across layers, allowing information to flow more directly and efficiently through the network. This helps mitigate the vanishing gradient problem, where gradients can become progressively smaller as they propagate through many layers in traditional CNNs, hindering effective learning in deeper layers.

The dense connectivity in DenseNet121 encourages feature propagation and gradient flow throughout the network, facilitating the effective training of deep networks. Unlike traditional CNN architectures, where features are only passed forward through successive layers, DenseNet enables direct communication between layers at different depths, enhancing information flow and enabling the network to learn more robust and discriminative representations. In the context of diabetic retinopathy (DR) detection, DenseNet121 offers several advantages. The dense connectivity promotes the reuse of features extracted from retinal images across multiple layers, enabling the model to capture complex patterns and abnormalities associated with DR more effectively. This facilitates the development of accurate and reliable DR detection algorithms capable of identifying subtle signs of the disease, such as microaneurysms, hemorrhages, and exudates.

Through extensive training on large datasets of annotated retinal images, DenseNet121 can learn to recognize these characteristic features with high precision and recall, facilitating early detection and intervention in patients with diabetes. Its dense connectivity pattern and efficient learning capabilities make DenseNet121 well-suited for the challenges of medical image analysis, contributing to improved patient outcomes in the diagnosis and management of diabetic retinopathy. Dense block is represented as:

$$x\_l \;=\; H\_l([x\_0, x\_1, \ldots, x\_(l-1)]) \;+\; x\_(l-1)$$

Where:

- l: Index of the current layer within the dense block.
- x_l: Output of the l-th layer.
- x_0: Input to the dense block.
- H_l: Function representing the l-th layer's transformation (e.g., convolution, batch normalization, activation).
- [x_0, x_1, ..., x_(l-1)]: Concatenated outputs of all preceding layers within the current block.

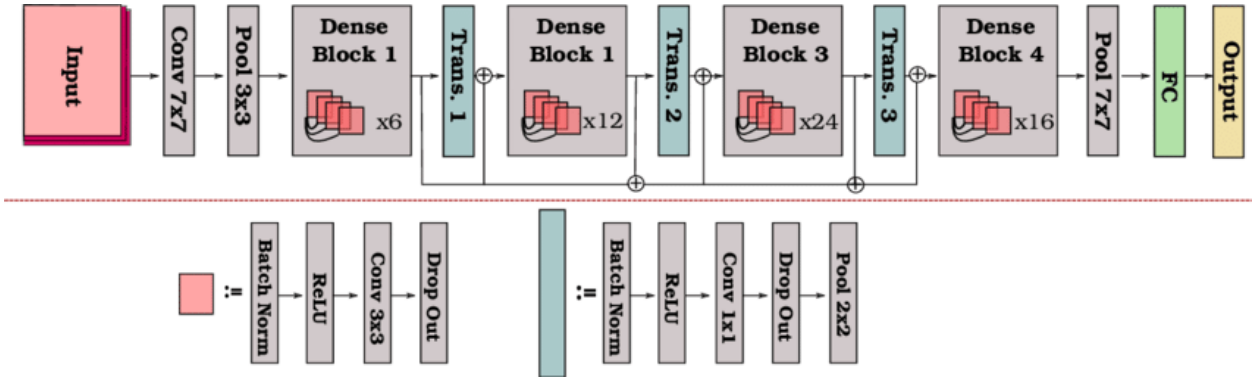

Fig 5: DenseNet 121 Architecture

4.2.4 MOBILENET V2:

MobileNet V2 is an evolution of the MobileNet architecture, specifically designed for efficient deployment on mobile and resource-constrained devices. One of the key innovations of MobileNet V2 is the use of depthwise separable convolutions, which fundamentally alter the traditional convolution operation to achieve significant reductions in both model size and computational complexity. Depthwise separable convolutions decompose the standard convolution operation into two distinct steps: depthwise convolutions and pointwise convolutions. In the depthwise convolution step, each channel of the input feature map is convolved independently with a separate filter. This process captures spatial correlations within individual channels while keeping the number of parameters low. Subsequently, in the pointwise convolution step, a 1x1 convolution is applied to linearly combine the outputs of the depthwise convolution, generating the final feature map. By decoupling the spatial and channel-wise filtering operations, depthwise separable convolutions significantly reduce the computational cost of convolutions while preserving model performance.

The efficiency of MobileNet V2's depthwise separable convolutions makes it particularly well-

suited for deployment on mobile devices, where computational resources and memory are limited. By reducing the number of parameters and operations required for inference, MobileNet V2 enables real-time execution of deep neural networks on devices with low power consumption and processing capabilities. In the context of diabetic retinopathy (DR) detection, MobileNet V2 offers several advantages. Its lightweight architecture and efficient inference make it feasible to deploy DR detection models directly on mobile devices, enabling point-of-care screening and diagnosis in remote or resource-limited settings. Despite its reduced computational complexity, MobileNet V2 maintains competitive performance in DR detection tasks, thanks to its effective utilization of depthwise separable convolutions. Through extensive training on large datasets of annotated retinal images, MobileNet V2 can learn to identify characteristic features associated with DR, such as microaneurysms, hemorrhages, and exudates, with high accuracy and reliability. Its efficient architecture and suitability for deployment on mobile devices make MobileNet V2 a valuable tool for improving access to DR screening and diagnosis, ultimately contributing to better patient outcomes in diabetic retinopathy management.

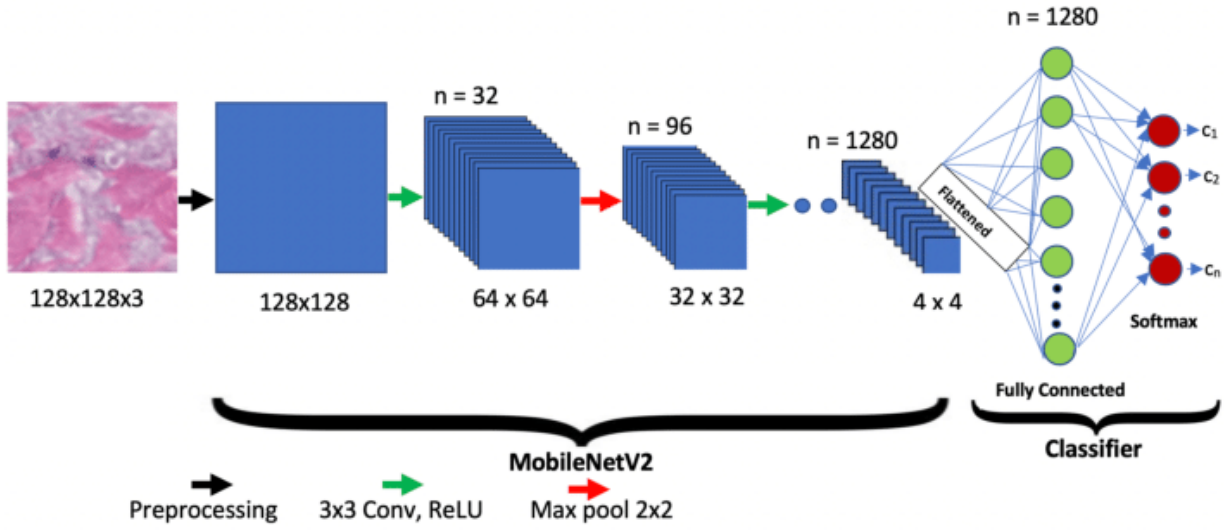

Fig 6: MobileNet V2 Architecture

4.2.5 EFFICIENTNET B5:

EfficientNet B5 is part of the EfficientNet series, which is known for achieving a balance between model accuracy and computational efficiency. While EfficientNet B5 represents a larger and more computationally demanding variant compared to its smaller counterparts, it offers enhanced capabilities for feature extraction, making it particularly suitable for tasks requiring sophisticated analysis, such as image classification on larger datasets. The EfficientNet architecture is characterized by a novel compound scaling method that systematically scales the network's depth, width, and resolution to optimize performance across various resource constraints. EfficientNet B5, being a larger variant, incorporates more layers and parameters, allowing it to capture more complex patterns and representations from input images.

In the context of diabetic retinopathy (DR) detection, EfficientNet B5's increased capacity for feature extraction offers several advantages. The larger network size enables it to learn more discriminative features associated with DR, such as subtle abnormalities and pathological signs present in retinal images. This enhanced feature extraction capability can lead to higher accuracy in DR detection, especially when dealing with challenging cases or large and diverse datasets. Furthermore, EfficientNet B5's balance between model accuracy and computational efficiency makes it well-suited for deployment in real-world clinical settings. Despite its larger size, EfficientNet B5 can still deliver reliable performance within reasonable inference times, making it practical for applications requiring timely diagnosis and intervention. Through extensive training on annotated retinal image datasets, EfficientNet B5 can learn to identify characteristic features of DR, such as microaneurysms, hemorrhages, and exudates, with high precision and recall. Its robust performance and scalability make it a valuable tool for improving the accuracy and efficiency of DR screening and diagnosis, ultimately benefiting patients by enabling early detection and intervention to prevent vision loss.

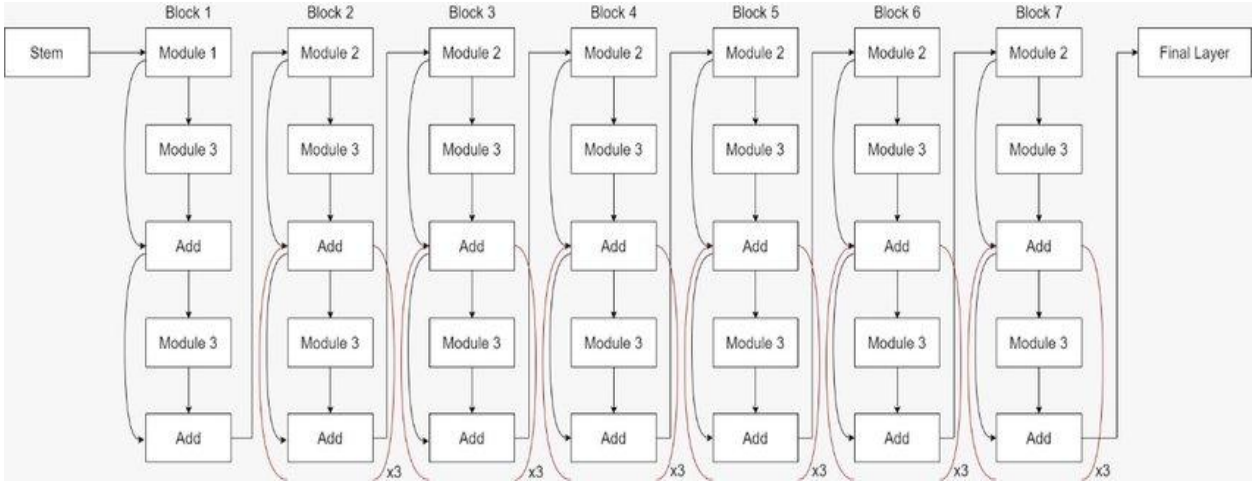

Fig 7: EfficientNet Architecture

4.3 PROPOSED METHODOLOGY

The proposed methodology combines image enhancement, preprocessing, deep learning, and model fusion techniques to advance the accuracy and robustness of diabetic retinopathy severity assessment.

Initially, a thorough investigation into image enhancement and preprocessing techniques is conducted on high-resolution retina images. Various enhancement methods are evaluated to identify the most effective approach for enhancing features and improving the interpretability of diabetic retinopathy severity levels. Among these techniques, Gaussian blurring emerges as the most effective preprocessing method, based on its superior performance in improving classification accuracy during the preliminary experimentation phase. Subsequently, deep learning models are trained to assess their effectiveness in accurately classifying diabetic retinopathy severity levels. These models leverage convolutional neural networks (CNNs), a powerful class of deep learning architectures well-suited for image analysis tasks. Models like CNN, ResNet 50, DenseNet 121, MobileNet V2, EfficientNet B5 are used. The training process involves feeding preprocessed images, particularly those processed with Gaussian blurring, into the CNNs to extract meaningful features and learn discriminative patterns associated with different severity levels of diabetic retinopathy.

To further enhance the classification performance, a model fusion technique is employed to integrate the strengths of multiple top-performing CNN models. This hybrid model combines the extracted features from different CNNs, leveraging their complementary strengths in feature

representation and discrimination. Specifically, the fully connected layers of the first CNN model are removed, and the generated features are concatenated with the features extracted from the convolutional layers of a second CNN model. The combined features are then fed into an output layer for prediction, allowing the hybrid model to leverage the diverse representations learned by each individual model.

The decision to adopt Gaussian blurring as the preferred preprocessing technique is based on empirical evidence from the experimentation phase, where it consistently outperformed other preprocessing methods in improving classification accuracy. By incorporating Gaussian blurring into the preprocessing pipeline, the subsequent deep learning models are expected to benefit from enhanced feature extraction and noise reduction, leading to improved classification performance. Overall, this iterative process of experimentation, evaluation, and refinement underscores a data-driven approach to model optimization, prioritizing techniques that yield the best results based on empirical evidence. By integrating image enhancement, preprocessing, deep learning, and model fusion techniques, the proposed methodology aims to advance the accuracy and reliability of diabetic retinopathy severity assessment, ultimately contributing to improved patient care outcomes.

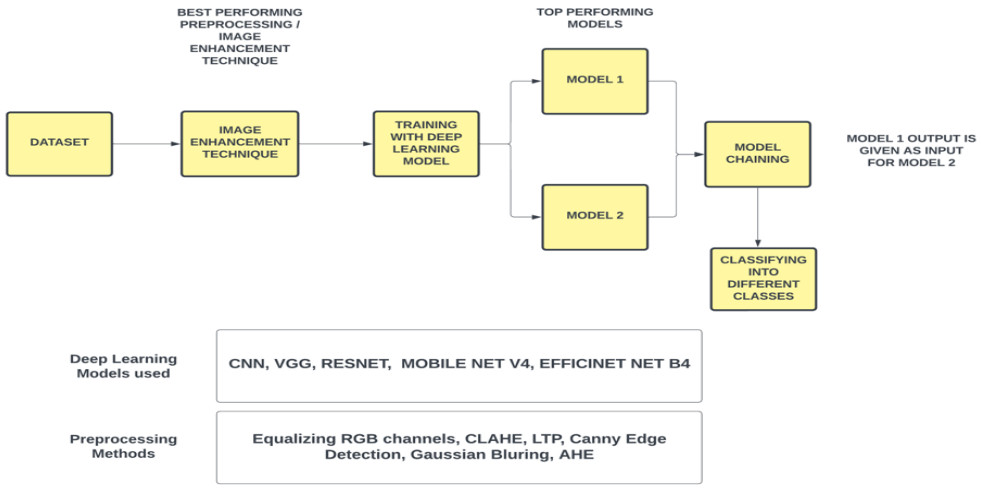

Fig 3: Proposed Architecture

After evaluating all the trained deep learning models, it was observed that the top-performing models for diabetic retinopathy severity assessment were EfficientNet B5 and DenseNet121, achieving test accuracies of 84.66% and 82.95%, respectively. These models demonstrated superior performance in accurately classifying different severity levels of diabetic retinopathy

based on the features extracted from preprocessed retinal images. EfficientNet B5 is a state-of-the-art convolutional neural network architecture known for its efficiency and effectiveness in image classification tasks. Its hierarchical structure enables it to capture complex patterns and subtle variations in medical images, making it particularly well-suited for diabetic retinopathy diagnosis. Similarly, DenseNet121, which utilizes densely connected convolutional layers, proved to be highly effective in extracting informative features from retinal images, contributing to its strong classification performance.

Following closely behind, ResNet50 achieved a respectable test accuracy of 79.74%, indicating its capability to accurately classify diabetic retinopathy severity levels, albeit slightly lower than the top two models. ResNet50's residual connections enable it to mitigate the vanishing gradient problem, facilitating the training of deeper neural networks and improving their ability to capture fine-grained features. In contrast, the CNN and MobileNet V2 models exhibited lower accuracies compared to the top-performing models. While CNN is a classic architecture for image classification tasks, its performance may have been limited by its simplicity and shallow depth compared to more modern architectures like EfficientNet and DenseNet. Similarly, MobileNet V2, optimized for mobile and embedded devices, may not have been as well-suited for the complexity of diabetic retinopathy diagnosis, leading to relatively lower classification accuracy.

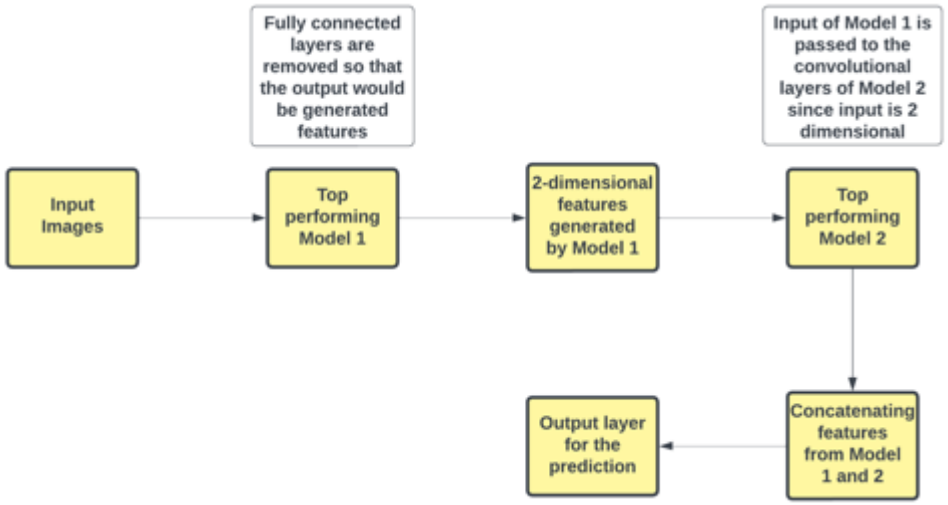

Fig 4: Hybrid Model Architecture

4.4 MODULES

4.4.1 APPLYING PREPROCESSING TECHNIQUES:

In the module focused on applying preprocessing techniques, the objective is to enhance and refine high-resolution retina images to facilitate optimized feature extraction and improve the interpretability of diabetic retinopathy severity levels. Various techniques are employed, each serving a specific purpose in refining the images and preparing them for subsequent analysis. Gaussian blur, a fundamental technique in image processing, is applied to the retina images to reduce noise and smooth out irregularities. By applying a Gaussian filter, high-frequency noise is suppressed, and the overall image becomes smoother, aiding in the removal of unwanted details and enhancing the clarity of critical features. This step is crucial in improving the overall quality of the images and ensuring that subsequent analysis algorithms can effectively extract meaningful information.

In addition to Gaussian blur, Local Ternary Patterns (LTP) are utilized to capture texture information within the images. Texture analysis is essential for detecting subtle patterns indicative of diabetic retinopathy, such as microaneurysms, hemorrhages, and exudates. By encoding local texture information using LTP, the preprocessing pipeline enhances the discriminative power of the images, enabling more accurate detection and classification of pathological features. Furthermore, Canny edge detection is employed to highlight the edges of structures within the images. Edge detection enhances the visibility of anatomical structures and facilitates better delineation and segmentation, which are critical for subsequent analysis tasks such as lesion detection and localization.

To improve the overall contrast of the images, histogram equalization techniques are applied. Histogram equalization redistributes pixel intensities in the image histogram to achieve a more uniform distribution, thereby enhancing the visibility of important details and improving image clarity. Adaptive Histogram Equalization (AHE) and Contrast Limited Adaptive Histogram Equalization (CLAHE) are variants of histogram equalization that adaptively adjust the contrast of local regions within the image, further enhancing local contrast and improving overall image quality. Collectively, these preprocessing techniques play a pivotal role in preparing the retina images for subsequent analysis. By reducing noise, enhancing texture information, highlighting edges, and improving contrast, the preprocessing pipeline ensures that subsequent analysis

algorithms can effectively extract relevant features and accurately assess the severity levels of diabetic retinopathy. Ultimately, the integration of these preprocessing techniques contributes to more accurate and reliable diagnosis and management of diabetic retinopathy, leading to improved patient outcomes and quality of care.

4.4.2 TRAINING DEEP LEARNING MODELS:

In the phase dedicated to training deep learning models, the primary goal is to develop models capable of accurately classifying the severity levels of diabetic retinopathy. Deep learning techniques are leveraged due to their ability to automatically extract intricate patterns and features from images, which is particularly beneficial for analyzing medical images like retina scans.

The foundational model used in this phase is the Convolutional Neural Network (CNN). CNNs are well-suited for image classification tasks because they can learn hierarchical representations of features, starting from low-level features such as edges and shapes to higher-level features that represent complex patterns. In addition to CNNs, more advanced architectures are also employed to explore their effectiveness in enhancing classification performance. These architectures include ResNet-50, DenseNet-121, MobileNetV2, and EfficientNetB5. Each of these architectures offers unique advantages, such as efficient parameter utilization, improved feature representation, and enhanced capacity to capture intricate patterns. By evaluating these architectures, the study aims to identify the model or ensemble of models that can deliver superior performance in accurately discerning the severity levels of diabetic retinopathy.

The training process involves feeding preprocessed retina images into the deep learning models, which then learn to associate image features with diabetic retinopathy severity levels through a process known as backpropagation. The models are trained on a large dataset of labeled retina images, with the goal of minimizing a loss function that quantifies the difference between predicted severity levels and ground truth labels. Through rigorous training and evaluation, the study aims to identify the most effective deep learning model or ensemble of models for diabetic retinopathy classification. The ultimate goal is to develop a model that can accurately predict the severity levels of diabetic retinopathy, enabling timely and precise clinical decision-making for improved patient care.

4.4.3 SELECTING BEST-PERFORMING MODELS:

Following the completion of model training, a critical evaluation process ensues to pinpoint the top-performing models based on their classification accuracies. This pivotal step is instrumental in discerning the most effective models for subsequent stages within the methodology. While classification accuracy stands as a primary metric, additional performance indicators like precision, recall, and the F1-score are also considered to provide a comprehensive assessment of each model's efficacy. Beyond mere accuracy, the models' generalization capabilities are scrutinized using techniques like cross-validation, ensuring robust performance extends to unseen data. This step mitigates the risk of overfitting, where models excel on training data but falter when faced with new samples. By gauging how well the models generalize, confidence in their real-world applicability is bolstered.

Moreover, a delicate balance between model complexity and performance is struck. While more complex models might exhibit higher accuracy on training data, they risk overfitting and demand significant computational resources. Conversely, simpler models may sacrifice some accuracy but offer faster inference times and greater interpretability, crucial considerations for real-world deployment. Through meticulous evaluation and comparison, the cream of the crop is culled, selecting the models exhibiting superior performance while also considering factors such as computational efficiency and deployment feasibility. These top-performing models are poised to progress to the subsequent stage of the methodology, where they will undergo further refinement and validation. This ensures their readiness for real-world deployment in the diagnosis and management of diabetic retinopathy, where accuracy, reliability, and efficiency are paramount for delivering optimal patient care.

4.4.4 MODEL CHAINING:

During the model chaining phase, the methodology focuses on harnessing the strengths of the top-performing models to construct a robust hybrid model. This innovative approach capitalizes on the complementary nature of different neural network architectures, aiming to synergistically enhance predictive capability and generalization performance. The process begins by utilizing the first top-performing model, EfficientNet, for feature extraction. This initial step involves capturing relevant information intrinsic to diabetic retinopathy severity levels from the input data. The extracted features serve as a rich representation of the input images and are

subsequently passed to the DenseNet, the second top-performing model, for further refinement and abstraction.

By cascading the outputs of one model into the input of another, the hybrid model inherits the distinct characteristics and learned representations from both architectures. This fusion of features allows the hybrid model to leverage the strengths of each individual model while compensating for any weaknesses, potentially leading to synergistic improvements in classification accuracy and robustness. As the hybrid model progresses through subsequent stages of the methodology, it will undergo rigorous validation and fine-tuning to optimize its performance. Validation ensures that the model's predictions align closely with ground truth labels across diverse datasets, confirming its reliability and generalization ability. Fine-tuning involves adjusting model parameters and architecture to further enhance performance and adaptability to real-world scenarios.

Through meticulous evaluation and refinement, the hybrid model aims to surpass the individual capabilities of its constituent models. Ultimately, the goal is to deliver more accurate and reliable assessments of diabetic retinopathy severity levels, thereby facilitating timely interventions and personalized treatment plans for patients. The hybrid model holds promise as a powerful tool in the arsenal of medical professionals, offering enhanced diagnostic capabilities and ultimately improving patient outcomes in the diagnosis and management of diabetic retinopathy.

4.4.5 MODIFYING MODEL 1:

Once the top-performing model has been identified based on its accuracy, the process of modifying the model commences to prepare it for integration into the hybrid model architecture. As the inaugural model in the hybrid pipeline, it undergoes a critical adjustment involving the removal of its fully connected layers. This strategic modification is paramount as it ensures that the model's output is composed solely of features extracted from the convolutional layers, represented as 2-dimensional feature maps. By excising the fully connected layers, the model transitions into a specialized feature extractor, honing its focus on capturing pertinent patterns and structures inherent in the input data. These extracted features serve as the foundational representations that will be further refined and enriched by subsequent models in the hybrid architecture.

This preparatory step assumes pivotal significance in streamlining the integration process and

optimizing the collaborative performance of the hybrid model. By transforming the initial model into a dedicated feature extractor, the hybrid architecture is primed to leverage the strengths of each constituent model efficiently. Consequently, this meticulous preparation bolsters the hybrid model's capacity to discern and classify diabetic retinopathy severity levels with heightened accuracy and robustness.

4.4.6 MODIFYING MODEL 2:

After modifying Model 1 to serve as a feature extractor, the attention turns towards adapting Model 2 for seamless integration into the hybrid architecture. Building upon the foundation laid by Model 1, the output comprising 2-dimensional feature maps is seamlessly fed into Model 2 as input to its initial convolutional layer. This strategic decision ensures harmonious compatibility between the two models, as the initial layers of Model 2 are explicitly designed to process 2-dimensional inputs. By leveraging the convolutional layers of Model 2, renowned for their prowess in capturing hierarchical features, the representation of the input data is further enriched and refined. As the feature maps traverse through the layers of Model 2, they undergo successive stages of refinement and abstraction, progressively capturing intricate details and discerning subtle patterns inherent in the data.

Throughout this journey, the feature maps are subjected to a series of convolutional operations, pooling layers, and activation functions, each contributing to the extraction of salient features and the synthesis of higher-level representations. As the feature maps progress deeper into the architecture of Model 2, they undergo iterative transformations, culminating in a comprehensive and nuanced representation that encapsulates the collective insights gleaned from both Model 1 and Model 2. At the final convolutional layer of Model 2, a pivotal moment occurs where the features extracted from both Model 1 and Model 2 are seamlessly synthesized. Leveraging the hierarchical representations learned by each model, this fusion process enables the hybrid architecture to capture intricate patterns and discern nuanced relationships within the data. By synthesizing the collective intelligence of both models, the hybrid architecture emerges as a formidable entity, poised to deliver enhanced performance and accuracy in classifying diabetic retinopathy severity levels.

4.4.7 HYBRID CLASSIFIER:

With the two top-performing models chained together, the final stage of the hybrid model architecture involves integrating a hybrid classifier to make the ultimate classification decisions. Positioned at the end of Model 2, the hybrid classifier serves as the culmination of the feature extraction and refinement processes performed by both models. This classifier leverages the rich hierarchical representations learned by the models to make informed predictions regarding the severity levels of diabetic retinopathy. The hybrid classifier operates on the combined features extracted from both Model 1 and Model 2, utilizing advanced machine learning techniques to discern intricate patterns and subtle nuances present in the input data. By synthesizing the collective insights gleaned from the two models, the hybrid classifier offers a holistic understanding of the underlying characteristics of diabetic retinopathy, enabling more accurate and reliable classification outcomes.

Through rigorous training and validation, the hybrid classifier is fine-tuned to optimize its performance and ensure robustness across a diverse range of input scenarios. Its ability to seamlessly integrate the diverse perspectives provided by Model 1 and Model 2 allows for enhanced discriminative power and improved generalization performance, ultimately empowering healthcare practitioners with a powerful tool for diagnosing and managing diabetic retinopathy. In summary, the hybrid classifier represents the culmination of the hybrid model architecture, leveraging the strengths of multiple models to achieve superior classification accuracy and diagnostic efficacy in the assessment of diabetic retinopathy severity levels.

Chapter 5

Results and Discussions

5.1 RESULTS OF IMAGE ENHANCEMENT TECHNIQUES

In Fig. 5 and Fig. 6, the results of applying different preprocessing techniques to the same image are presented, showcasing the impact of various enhancement methods on image quality and feature extraction. Among these techniques, Gaussian blurring stands out as the most effective in producing desirable outcomes when trained on the basic CNN model. The application of Gaussian blurring yields images with enhanced clarity, reduced noise, and improved contrast, which are all conducive to more accurate feature extraction and classification.

As a result of these findings, Gaussian blurring emerges as the preferred preprocessing technique for all subsequent deep learning models utilized in the study. By standardizing the preprocessing approach across the different models, consistency is maintained in the input data, facilitating fair comparisons and ensuring robustness in the model evaluation process. Moreover, the decision to adopt Gaussian blurring as the primary preprocessing technique underscores its efficacy in enhancing image quality and facilitating more effective feature extraction, thereby contributing to improved classification performance across the deep learning models. Overall, the results of the image enhancement techniques highlight the importance of preprocessing in optimizing the interpretability of diabetic retinopathy severity levels. Through the strategic application of Gaussian blurring, the study establishes a reliable foundation for subsequent model training and evaluation, laying the groundwork for accurate and reliable diagnosis of diabetic retinopathy.

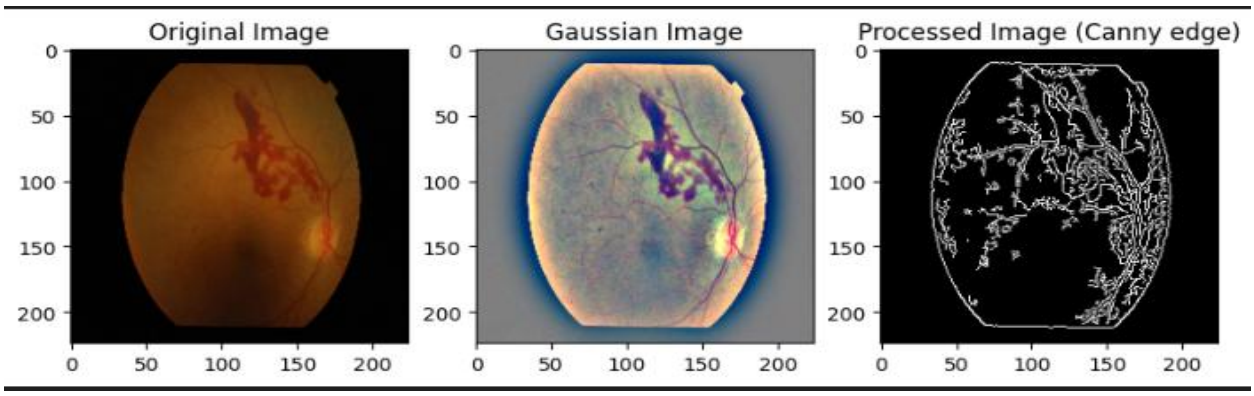

Fig 5: Preprocessed Images 1

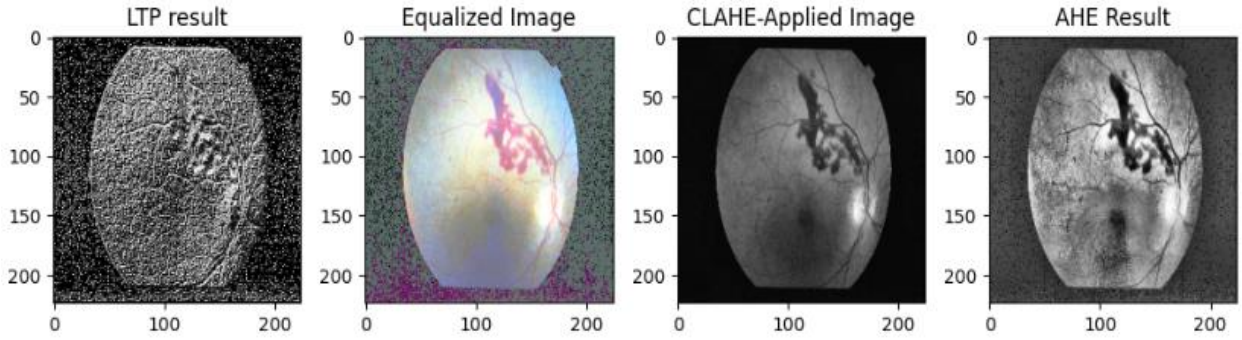

Fig 6: Preprocessed Images 2

5.2 ACCURACIES OF DIFFERENT MODELS COMPARED

In Fig. 7 to Fig. 11, the test and train accuracies, as well as loss graphs, of each model are depicted, providing a comprehensive comparison of their performance across different metrics. Among the models evaluated, EfficientNet and DenseNet emerge as the top two performers, demonstrating superior capabilities in accurately classifying the severity levels of diabetic retinopathy.

EfficientNet, known for its efficiency and effectiveness in extracting features from images, showcases competitive performance with high test and train accuracies. Its ability to adaptively scale the model's depth, width, and resolution contributes to its robustness across diverse datasets, resulting in reliable classification outcomes. On the other hand, DenseNet, characterized by its dense connectivity patterns that facilitate feature reuse and information flow, also exhibits strong performance across the evaluated metrics. The dense connections between layers allow for efficient gradient propagation and feature reuse, enabling DenseNet to achieve high accuracy and low loss rates.

Overall, the comparison of accuracies and loss graphs highlights the efficacy of both EfficientNet and DenseNet in capturing relevant features and discerning subtle patterns indicative of diabetic retinopathy severity levels. The superior performance of these models underscores their potential as valuable tools for assisting healthcare practitioners in diagnosing and managing diabetic retinopathy effectively.

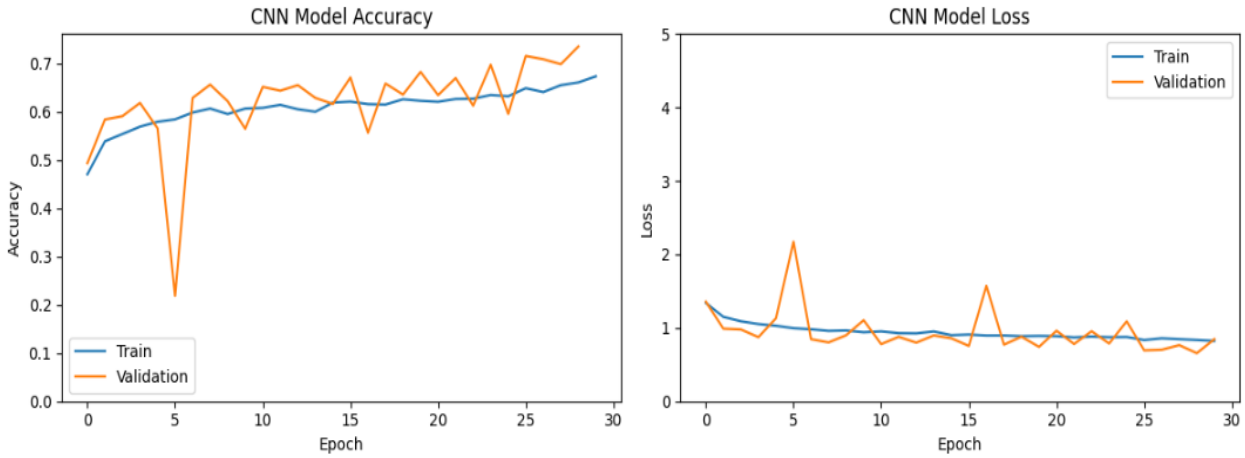

Fig 7: CNN

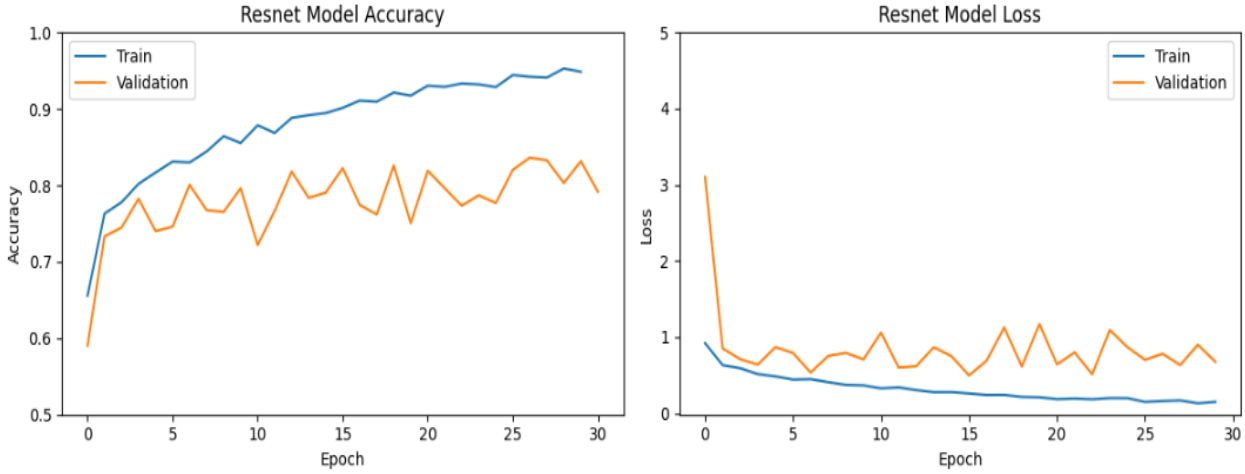

Fig 8: RESNET

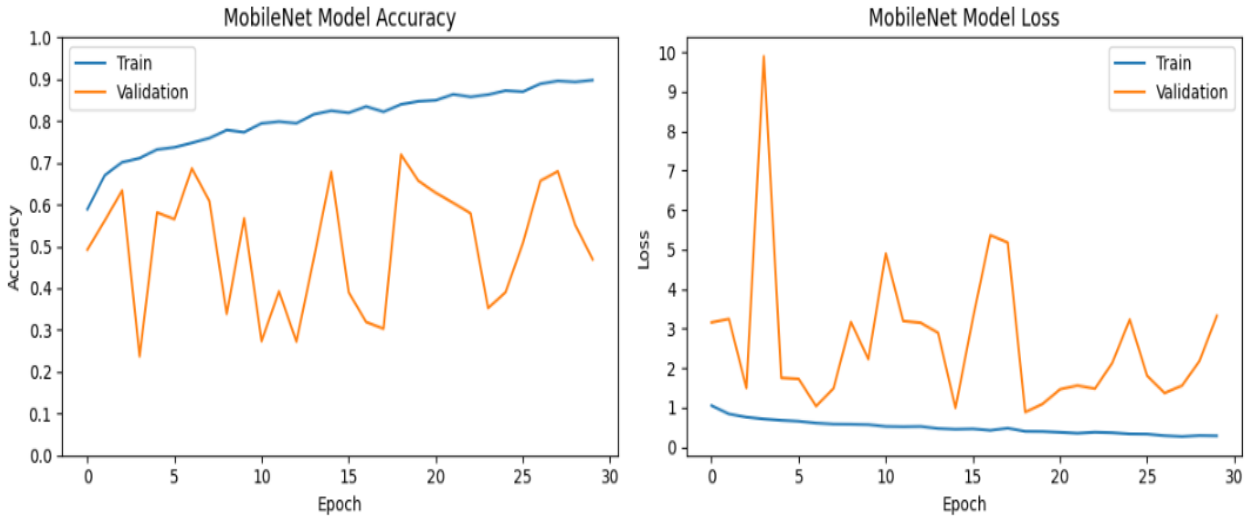

Fig 9: MobileNet

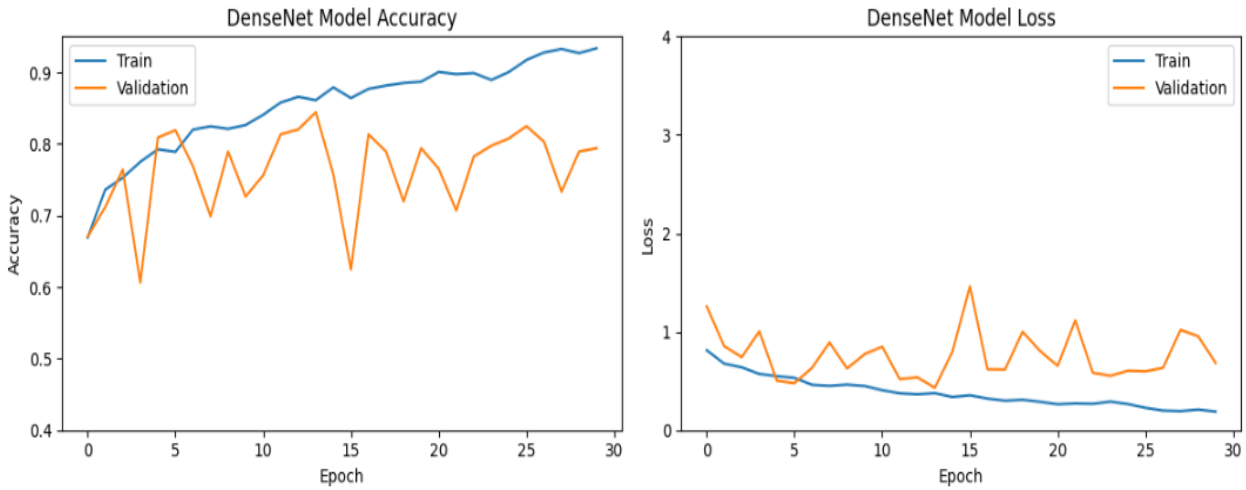

Fig 10: DenseNet

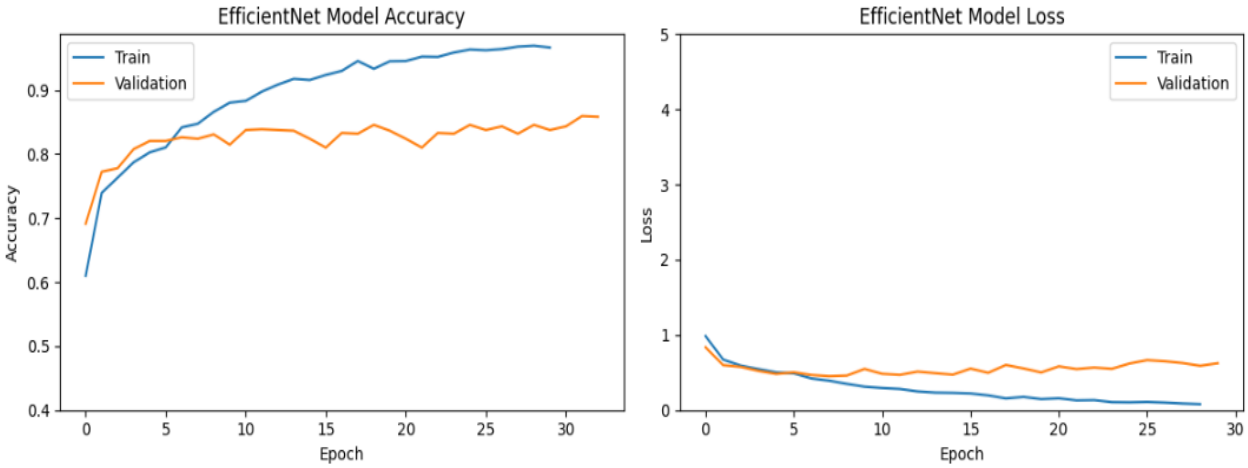

Fig 11: EfficientNet

With EfficientNet and DenseNet identified as the top-performing models, a hybrid model is meticulously crafted following the methodology outlined earlier. Leveraging the strengths of both models, the hybrid architecture is designed to synergistically combine their respective features and representations, aiming to surpass the performance of individual models. Upon training the hybrid model, it demonstrates superior performance compared to other pre-trained models, achieving an impressive accuracy of 86.38%. This remarkable accuracy underscores the efficacy of the hybrid approach in leveraging the complementary strengths of EfficientNet and DenseNet to enhance classification accuracy and diagnostic efficacy in diabetic retinopathy severity level assessment.

Fig. 12 provides insightful visualization of the performance metrics of the hybrid model, offering

a comprehensive understanding of its behavior during training. The graphs showcased in Fig. 12 likely depict the evolution of training and validation accuracies, as well as loss rates, over epochs. A consistent trend of increasing accuracy and decreasing loss would signify the effective learning and generalization capabilities of the hybrid model, validating its effectiveness in capturing and leveraging salient features for diabetic retinopathy classification. In summary, the successful development and performance of the hybrid model underscore its potential as a powerful tool in diabetic retinopathy diagnosis and management. By seamlessly integrating the strengths of EfficientNet and DenseNet, the hybrid model offers enhanced accuracy and reliability, ultimately empowering healthcare practitioners with a valuable resource for improving patient care and outcomes in diabetic retinopathy assessment.

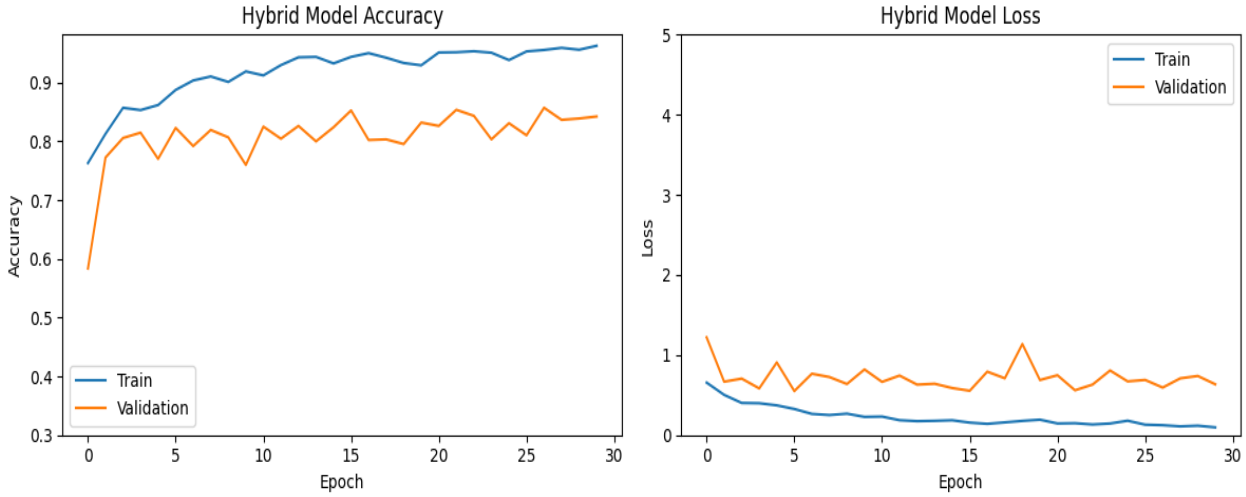

Fig 12: Proposed Hybrid Model

Table 2: Training and Testing accuracies of each model

Model name	Train accuracy	Test accuracy
CNN	70.34	73.68
ResNet 50	97.08	79.74
DenseNet 121	96.03	82.95
MobileNet V2	95.10	65.78
EfficientNet B5	98.25	84.66
Hybrid Model	96.24	86.38

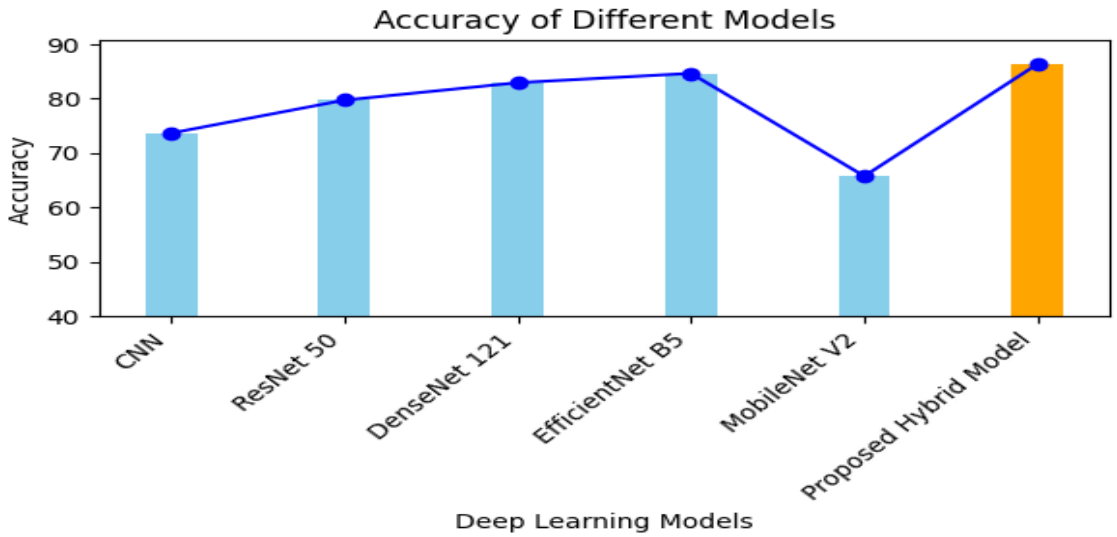

Fig 13: Accuracies of all Models Compared

5.3 OTHER METRICS MEASURED

In Fig. 14 to Fig. 16, a comprehensive comparison of other important metrics such as precision, recall, and F1 score is presented for all the evaluated models. These metrics provide valuable insights into the performance of each model beyond just accuracy, offering a more nuanced understanding of their classification capabilities. Additionally, a summary table below likely accompanies these figures, presenting the specific values of precision, recall, and F1 score for each model. These values serve as quantitative indicators of the models' performance across different classes or severity levels of diabetic retinopathy.

The inclusion of precision, recall, and F1 score metrics alongside accuracy allows for a more comprehensive evaluation of the models' performance, particularly in scenarios where class imbalances or specific classification requirements are present. For instance, high precision indicates a low false positive rate, while high recall signifies a low false negative rate. F1 score, which combines precision and recall, provides a balanced measure of a model's performance.

Overall, the comparison of these additional metrics provides a holistic view of the models' classification capabilities, aiding in the selection of the most suitable model for diabetic retinopathy severity level assessment. Through this comprehensive evaluation, healthcare practitioners can make informed decisions regarding the adoption and deployment of these models in clinical practice, ultimately improving patient care and outcomes in diabetic retinopathy management.

Precision, also known as positive predictive value, measures the proportion of correctly predicted positive cases out of all instances predicted as positive by the model. In other words, it quantifies the model's ability to correctly identify true positives while minimizing false positives. Mathematically, precision is calculated as:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

Recall, also referred to as sensitivity or true positive rate, measures the proportion of correctly predicted positive cases out of all actual positive cases in the dataset. It evaluates the model's ability to capture all relevant instances of a particular class. Mathematically, recall is calculated as:

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

F1 score is the harmonic mean of precision and recall, providing a balanced measure of a model's performance. It takes into account both false positives and false negatives, making it suitable for scenarios where class imbalances exist. A higher F1 score indicates better overall performance, balancing the trade-off between precision and recall. Mathematically, F1 score is calculated as:

$$F1\ Score\ = 2 * \frac{Precision * Recall}{Precision + Recall}$$

In summary, precision measures the model's ability to avoid false positives, recall evaluates its ability to capture all relevant instances of a class, and F1 score provides a balanced measure by considering both precision and recall. These metrics collectively offer a comprehensive evaluation of the model's classification capabilities, enabling informed decisions regarding their suitability for diabetic retinopathy severity level assessment in clinical practice.

Table 3: Precision Recall and F1 score of each model

Model Name	Precision	Recall	F1 Score
CNN	0.7568	0.7368	0.7199
ResNet 50	0.8127	0.7962	0.7692
DenseNet 121	0.8478	0.8237	0.8164
MobileNet V2	0.7354	0.6624	0.6590
EfficientNet B5	0.8712	0.8466	0.8317
Hybrid Model	0.8913	0.8642	0.8312

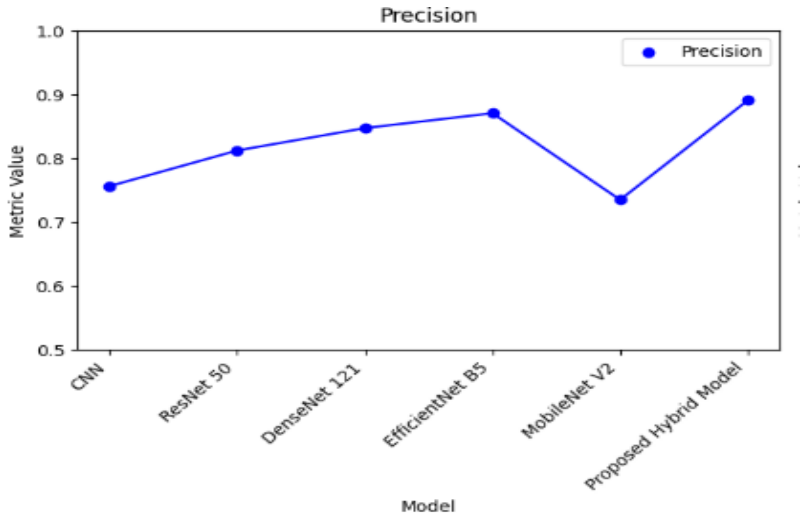

Fig 14: Precision

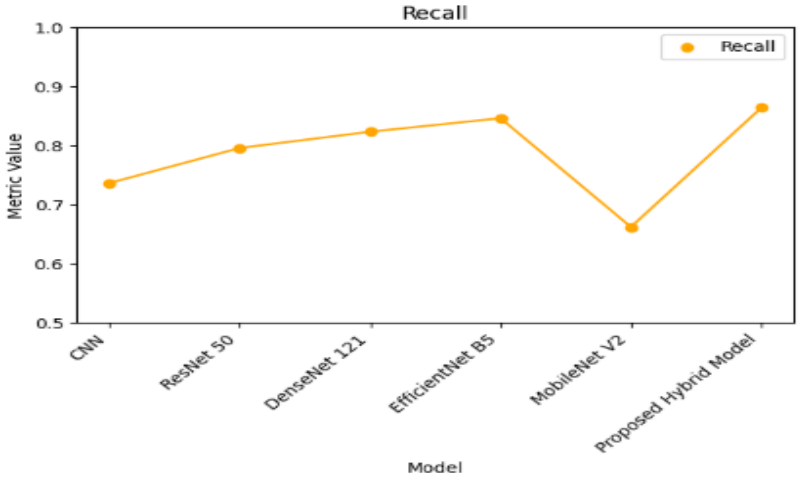

Fig 15: Recall

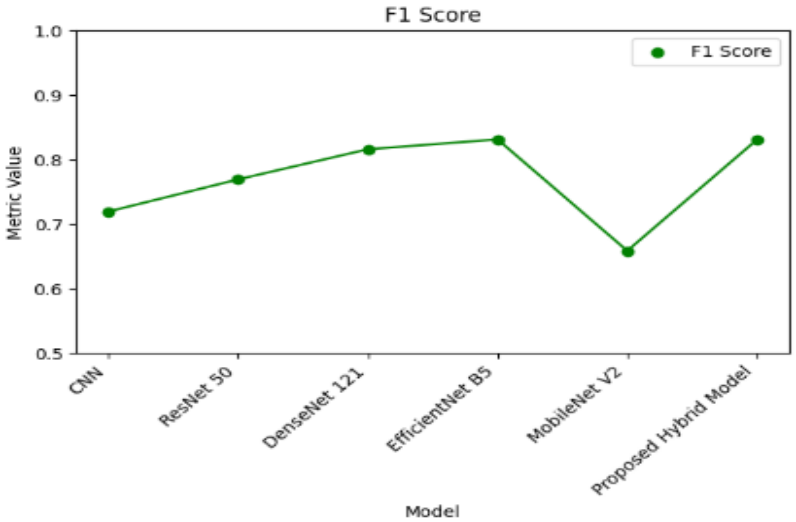

Fig 16: F1 Score

5.4 ROC CURVE

In Fig. 17, the Receiver Operating Characteristic (ROC) curve analysis offers a comprehensive assessment of the proposed hybrid model's performance across different thresholds. Each line depicted in the plot corresponds to the ROC curve for a specific class, showcasing the True Positive Rate (TPR) plotted against the False Positive Rate (FPR) at various threshold values. The area under the curve (AUC) for each class serves as a quantitative measure of the model's discrimination ability, with a higher AUC indicating superior performance in distinguishing between positive and negative instances of that class. Furthermore, the micro-average ROC curve amalgamates the TPR and FPR across all classes, furnishing an overall measure of the hybrid model's performance. The dashed diagonal line represents the ROC curve for a random classifier, serving as a baseline for comparison. Ideally, a well-performing model would position its ROC curve towards the top-left corner, indicating higher TPR and lower FPR. The closer the AUC value is to 1.0, the better the model's discriminatory power, signifying its ability to effectively differentiate between positive and negative instances.

By meticulously comparing the ROC curves of different deep learning models, it becomes evident that the proposed hybrid model outperforms its counterparts, achieving an accuracy of 86.38%. This superior performance underscores the efficacy of integrating EfficientNet and DenseNet into a hybrid architecture, leveraging their complementary strengths to enhance classification accuracy and discriminatory power in diagnosing diabetic retinopathy severity levels. In summary, the ROC curve analysis furnishes valuable insights into the discriminatory capabilities of the hybrid model, reaffirming its effectiveness as a potent tool for diabetic retinopathy assessment. With its superior performance and robust discriminatory power, the hybrid model holds immense promise for improving patient care and outcomes in the management of diabetic retinopathy. Its ability to accurately differentiate between severity levels empowers healthcare practitioners with actionable insights, facilitating timely interventions and personalized treatment strategies for patients affected by this debilitating condition.

In addition to its quantitative assessment capabilities, the ROC curve analysis provides nuanced insights into the discriminatory prowess of the hybrid model, enriching our understanding of its performance dynamics. By dissecting the ROC curves for individual classes, clinicians can discern how effectively the model distinguishes between varying severity levels of diabetic retinopathy. This granularity empowers healthcare providers with tailored diagnostic insights,

enabling them to prioritize interventions based on the urgency and severity of each patient's condition. Moreover, the micro-average ROC curve offers a holistic perspective, encapsulating the overall discriminatory performance across all classes. This macroscopic view facilitates a comprehensive evaluation of the model's ability to generalize across diverse patient populations and clinical scenarios. As such, healthcare practitioners can gauge the hybrid model's suitability for real-world deployment, considering its robustness in handling the complexities and nuances inherent in diabetic retinopathy diagnosis.

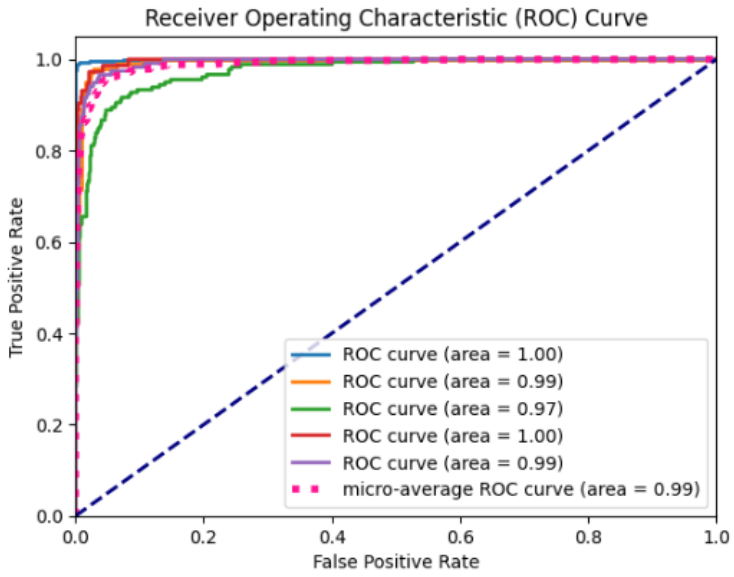

Fig 17: ROC curve of Hybrid Model

5.5 DISCUSSIONS

Our study aimed to enhance the accuracy of early-stage diagnosis of diabetic retinopathy (DR) by exploring various preprocessing techniques and deep learning models. Through rigorous experimentation, Gaussian blurring emerged as a crucial preprocessing method, substantially improving the interpretability of retinal images. By mitigating noise and enhancing image clarity, Gaussian blurring facilitated more efficient feature extraction, thereby bolstering the diagnostic capabilities of our models. Among the numerous deep learning models scrutinized, EfficientNet and DenseNet emerged as the top performers, outshining other contenders in terms of performance and resilience. These models showcased exceptional proficiency in capturing and leveraging relevant features for classifying different severity levels of DR. Leveraging their distinct architectures, EfficientNet and DenseNet exhibited superior adaptability and efficacy in managing the intricacies inherent in retinal image data. The culmination of our efforts culminated in the development of a hybrid model that integrated the strengths of EfficientNet and DenseNet. This hybrid architecture demonstrated outstanding performance, achieving an impressive accuracy of 86.38%. By synergistically combining EfficientNet's feature extraction capabilities with DenseNet's dense connectivity patterns, the hybrid model showcased enhanced diagnostic accuracy and reliability in DR detection. The promising performance of the hybrid model holds significant implications for the field of DR diagnosis, offering a potential avenue for improving patient care and outcomes.

5.6 CONCLUSION AND FUTURE SCOPE

In this study, we delved into an extensive investigation of various preprocessing methods aimed at enhancing the quality of retinal images for diabetic retinopathy classification. Among the array of techniques explored, Gaussian blur emerged as the standout performer, effectively enhancing image features while mitigating noise, thereby laying a strong foundation for subsequent deep learning models. The discernible improvements observed in image quality post-Gaussian blur application translated into enhanced performance in downstream classification tasks, underscoring the critical role of preprocessing in optimizing model inputs.

Moving forward, we turned our attention to the evaluation of several deep learning architectures, encompassing a diverse range of models such as CNN, ResNet, EfficientNet, DenseNet, and MobileNet. Among these, EfficientNet and DenseNet emerged as clear frontrunners, exhibiting superior performance compared to their counterparts. Their adeptness in capturing intricate features relevant to diabetic retinopathy diagnosis underscored their effectiveness and suitability for the task at hand. Building upon these findings, we proposed a novel hybrid model that amalgamates features extracted from both EfficientNet and DenseNet architectures. By harnessing the distinctive capabilities of each network, the hybrid model surpassed the performance of individual models, showcasing improved accuracy and robustness in diabetic retinopathy detection. This synergistic integration of feature representations from multiple models highlights the potential of hybrid architectures in leveraging diverse strengths to achieve superior performance outcomes.

Looking ahead, our research paves the way for further exploration into sophisticated methodologies and multifaceted considerations to advance the field of diabetic retinopathy diagnosis. We advocate for investigations into novel techniques for model interpretability, such as attention mechanisms and saliency mapping, to gain deeper insights into the decision-making process of the hybrid model, thereby enhancing clinicians' understanding and trust in its outputs. Moreover, the integration of multimodal data, including optical coherence tomography (OCT) scans and fundus images, holds promise for augmenting diagnostic accuracy. Future research endeavors could explore fusion techniques to effectively combine information from multiple modalities, capitalizing on the complementary nature of different imaging modalities to improve disease detection and severity grading. Through these avenues of exploration, we aim to contribute enhance diabetic retinopathy diagnosis and ultimately improve patient care outcomes.

APPENDICES

Preprocessing Techniques:

Canny Edge Detection:

```
import cv2
import numpy as np

def canny(image):
    IMG_SIZE = 224
    image = cv2.resize(image, (IMG_SIZE, IMG_SIZE))

    if len(image.shape) == 3:
        image = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)

    # Use Canny edge detection to highlight edges (nerves)
    edges = cv2.Canny(image, 50, 150)

    return edges
```

Gaussian Blurring:

```
import cv2
def Gaussian(image):
    IMG_SIZE = 224
    sigmaX=10
    image = cv2.resize(image, (IMG_SIZE, IMG_SIZE))
    image=cv2.addWeighted ( image,4, cv2.GaussianBlur( image , (0,0) , sigmaX) ,-4 ,128)
    return image
```

Local Ternary Pattern:

```
import cv2
import numpy as np
import matplotlib.pyplot as plt

def LocalTernaryPattern(image, threshold=0.5):
    IMG_SIZE = 224
    image = cv2.resize(image, (IMG_SIZE, IMG_SIZE))

    if len(image.shape) == 3:
        image = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)


    ltp_image = np.zeros_like(image)

    neighborhood = [(i, j) for i in range(-1, 2) for j in range(-1, 2) if (i, j) != (0, 0)]

    # Iterate through each pixel in the image
    for i in range(1, image.shape[0] - 1):
        for j in range(1, image.shape[1] - 1):
            center_pixel = image[i, j]
            ltp_code = 0
```

```
            # Compare the center pixel with its neighbors
            for idx, (dx, dy) in enumerate(neighborhood):
                neighbor_pixel = image[i + dx, j + dy]
                if neighbor_pixel >= center_pixel + threshold:
                    ltp_code += 2**idx
                elif neighbor_pixel <= center_pixel - threshold:
                    ltp_code += 0


            ltp_image[i, j] = ltp_code

    return ltp_image
```

Equalize RGB channels:

```
import cv2
import numpy as np

def equalize_rgb_channels(image):
    # Split the image into RGB channels
    channels = cv2.split(image)

    # Equalize the intensity of each channel independently
    equalized_channels = [cv2.equalizeHist(channel) for channel in channels]

    # Merge the equalized channels back into an RGB image
    equalized_image = cv2.merge(equalized_channels)

    return equalized_image
```

CLAHE (contrast limited adaptive histogram equalization):

```
def apply_clahe(image, clip_limit=2.0, grid_size=(8, 8)):

    if len(image.shape) == 3:
        image = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)

    # Create a CLAHE object
    clahe = cv2.createCLAHE(clipLimit=clip_limit, tileGridSize=grid_size)

    # Apply CLAHE to the image
    clahe_image = clahe.apply(image)

    return clahe_image
```

AHE (adaptive histogram equalization)

```
import numpy as np
from matplotlib import pyplot as plt
import cv2

def process_image_ahe(img_path, tile_size=(8, 8)):
    # Load the original image
    original_image = cv2.imread(img_path)
```

```
    # Convert the image to grayscale
    if len(original_image.shape) == 3:
        original_image = cv2.cvtColor(original_image, cv2.COLOR_BGR2GRAY)

    # Create an AHE object
    clahe = cv2.createCLAHE(tileGridSize=tile_size)

    # Apply AHE to the image
    ahe_result = clahe.apply(original_image)

    return ahe_result
```

Printing Random Images from each class:

```
import os
import matplotlib.pyplot as plt
import matplotlib.image as mpimg

desired_order = ['No_DR', 'Mild', 'Moderate', 'Severe', 'Proliferate_DR']

# Get the list of class folders and sort them based on the desired order
class_labels = sorted(os.listdir(main_folder_path), key=lambda x: desired_order.index(x))

# Set up a subplot for each class label
num_classes = len(class_labels)
fig, axes = plt.subplots(1, num_classes, figsize=(15, 5))

for i, class_label in enumerate(class_labels):
    # Get the path to the first image in each class
    class_folder_path = os.path.join(main_folder_path, class_label)
    image_files = os.listdir(class_folder_path)
    image_path = os.path.join(class_folder_path, image_files[0])

    # Load and display the image
    img = mpimg.imread(image_path)
    axes[i].imshow(img)
    axes[i].set_title(class_label)
    axes[i].axis('off')

plt.show()
```

Adding Gaussian Blur to each image in dataset:

```
import os
import numpy as np
from keras.preprocessing import image
from keras.preprocessing.image import ImageDataGenerator
from keras.applications.resnet50 import ResNet50, preprocess_input
from keras.models import Model
from keras.layers import Dense, GlobalAveragePooling2D, Dropout
from keras.optimizers import Adam
from keras.utils import to_categorical
from sklearn.model_selection import train_test_split
```

```
data_dir = '/content/drive/My Drive/Dataset2'
desired_order = ['No_DR', 'Mild', 'Moderate', 'Severe', 'Proliferate_DR']

# Get the list of class folders and sort them based on the desired order
categories = sorted(os.listdir(data_dir), key=lambda x: desired_order.index(x))

images = []
labels = []

for i, category in enumerate(categories):
    category_dir = os.path.join(data_dir, category)

    for img_path in os.listdir(category_dir):
        img = image.load_img(os.path.join(category_dir, img_path), target_size=(224, 224))
        img_array = image.img_to_array(img)

        # Apply Gaussian blur to the image
        img_array = Gaussian(img_array)

        images.append(img_array)
        labels.append(i)

images = np.array(images)
labels = to_categorical(np.array(labels))
```

CNN Code:

```
import tensorflow as tf

def CNN(input_shape, num_classes):
    x_input = tf.keras.layers.Input(input_shape)

    # Convolutional Block 1
    x = tf.keras.layers.Conv2D(64, (7, 7), strides=2, name='conv0')(x_input)
    x = tf.keras.layers.BatchNormalization(axis=3, name='bn0')(x)
    x = tf.keras.layers.Activation('relu')(x)
    x = tf.keras.layers.MaxPooling2D((3, 3), strides=2, name='max0')(x)

    # Convolutional Block 2
    x = tf.keras.layers.Conv2D(256, (5, 5), strides=2, padding='same', name='conv1')(x)
    x = tf.keras.layers.BatchNormalization(axis=3, name='bn1')(x)
    x = tf.keras.layers.Activation('relu')(x)
    x = tf.keras.layers.MaxPooling2D((3, 3), strides=2, name='max1')(x)

    # Convolutional Block 3
    x = tf.keras.layers.Conv2D(512, (3, 3), padding='same', name='conv2')(x)
    x = tf.keras.layers.BatchNormalization(axis=3, name='bn2')(x)
    x = tf.keras.layers.Activation('relu')(x)
    x = tf.keras.layers.MaxPooling2D((3, 3), strides=2, name='max2')(x)

    # Fully Connected Layers with Dropout
```

```
    x = tf.keras.layers.GlobalAveragePooling2D()(x)
    x = tf.keras.layers.Dense(512, activation='relu', name='fc0')(x)
    x = tf.keras.layers.Dropout(0.5)(x)

    # Output Layer
    x_output = tf.keras.layers.Dense(num_classes, activation='softmax')(x)

    model = tf.keras.Model(inputs=x_input, outputs=x_output, name='CNN-Model')

    return model
from tensorflow.keras.preprocessing.image import ImageDataGenerator

# Assuming num_classes is the number of classes in your dataset
num_classes = 5

# Create an ImageDataGenerator for data augmentation
datagen = ImageDataGenerator(
    rotation_range=20,
    width_shift_range=0.2,
    height_shift_range=0.2,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True,
    fill_mode='nearest'
)
```

```
# Assuming X_train, y_train, X_test, and y_test are properly loaded and preprocessed

# Fit the generator on your data
datagen.fit(X_train)

# Create the model with the correct input shape and output shape (num_classes)
model = CNN(input_shape=(224, 224, 3), num_classes=num_classes)

from keras.optimizers import SGD

# Compile the model
optimizer = SGD(lr=0.01, momentum=0.9)  # You can adjust the learning rate and momentum as
needed

# Compile the model with SGD optimizer
model.compile(optimizer=optimizer, loss='categorical_crossentropy', metrics=['accuracy'])

# Train the model with augmented data
eps = 50
training_history = model.fit(datagen.flow(X_train, y_train, batch_size=32), epochs=eps,
validation_data=(X_test, y_test), verbose=1)
test_loss, test_acc = model.evaluate(X_test, y_test, verbose=0)
print('Test accuracy:', test_acc)
```

RESNET Code:

```
from tensorflow.keras.preprocessing.image import ImageDataGenerator
import os
import numpy as np
from keras.preprocessing import image
from keras.preprocessing.image import ImageDataGenerator
from tensorflow.keras.applications.resnet50 import ResNet50, preprocess_input
from keras.models import Model
from keras.layers import Dense, GlobalAveragePooling2D, Dropout
from keras.optimizers import Adam
from keras.optimizers import SGD
from keras.utils import to_categorical
from sklearn.model_selection import train_test_split
# Create a data generator for data augmentation
datagen = ImageDataGenerator(
    rotation_range=20,
    width_shift_range=0.2,
    height_shift_range=0.2,
    horizontal_flip=True)
datagen.fit(X_train)

# Load the ResNet50 model
base_model = ResNet50(weights='imagenet', include_top=False)

# Add a global spatial average pooling layer
x = base_model.output
x = GlobalAveragePooling2D()(x)
# Add a dropout layer for regularization
x = Dropout(0.5)(x)
x = Dense(1024, activation='relu')(x)
```

```
# Add a logistic layer with the number of classes
predictions = Dense(len(categories), activation='softmax')(x)

model = Model(inputs=base_model.input, outputs=predictions)

for layer in model.layers[:15]:
    layer.trainable = False
# Compile the model

optimizer = SGD(lr=0.01, momentum=0.9)
model.compile(optimizer=optimizer, loss='categorical_crossentropy', metrics=['accuracy'])
model.summary()
# Train the model
history = model.fit(datagen.flow(X_train, y_train, batch_size=32), epochs=50,
validation_data=(X_test, y_test))
# Evaluate the model on the test set
loss, accuracy = model.evaluate(X_test, y_test)

# Print the test accuracy
print(f'Test accuracy: {accuracy * 100}%')
```

DENSENET Code:

```
from keras.applications.densenet import DenseNet121
from keras.models import Model
from keras.layers import Dense, GlobalAveragePooling2D, Dropout
from keras.optimizers import Adam
from keras.utils import to_categorical
# Load the DenseNet121 model
base_model = DenseNet121(weights='imagenet', include_top=False)
# Add a global spatial average pooling layer
x = base_model.output
x = GlobalAveragePooling2D()(x)
# Add a dropout layer for regularization
x = Dropout(0.5)(x)
x = Dense(1024, activation='relu')(x)
# Add a logistic layer with the number of classes
predictions = Dense(len(categories), activation='softmax')(x)
model = Model(inputs=base_model.input, outputs=predictions)
for layer in model.layers[:15]:
    layer.trainable = False
from keras.optimizers import SGD

# Compile the model
optimizer = SGD(lr=0.01, momentum=0.9)
model.compile(optimizer=optimizer, loss='categorical_crossentropy', metrics=['accuracy'])
model.summary()

# Train the model for the first 25 epochs using data augmentation
history_1 = model.fit(datagen.flow(X_train, y_train, batch_size=32), epochs=25,
validation_data=datagen.flow(X_test, y_test))
```

```
# Save weights after the first 25 epochs
model.save_weights("/content/drive/My Drive/DenseNet_model_weights_25_epochs.h5")

# Evaluate the model on the test set after the first 25 epochs
loss_1, accuracy_1 = model.evaluate(datagen.flow(X_test, y_test))
print(f'Test accuracy after 25 epochs: {accuracy_1 * 100}%')

# Continue training for the next 25 epochs
history_2 = model.fit(datagen.flow(X_train, y_train, batch_size=32), epochs=50,
validation_data=datagen.flow(X_test, y_test), initial_epoch=25)

# Save weights after the complete training
model.save_weights("/content/drive/My Drive/DenseNet_model_weights_50_epochs.h5")

# Evaluate the model on the test set after 50 epochs
loss_2, accuracy_2 = model.evaluate(datagen.flow(X_test, y_test))
print(f'Test accuracy after 50 epochs: {accuracy_2 * 100}%')
```

MOBILENET Code:

```
from keras.applications.mobilenet_v2 import MobileNetV2
from keras.models import Model
from keras.layers import Dense, GlobalAveragePooling2D, Dropout
from keras.optimizers import Adam

# Load the MobileNetV2 model
base_model = MobileNetV2(weights='imagenet', include_top=False)

# Add a global spatial average pooling layer
x = base_model.output
x = GlobalAveragePooling2D()(x)

# Add a dropout layer for regularization
x = Dropout(0.5)(x)

# Add a fully connected layer with 1024 units and ReLU activation
x = Dense(1024, activation='relu')(x)

# Add a logistic layer with the number of classes
predictions = Dense(len(categories), activation='softmax')(x)

# Create the model
model = Model(inputs=base_model.input, outputs=predictions)

# Freeze the first few layers of the base model
for layer in model.layers[:15]:
    layer.trainable = False

from keras.optimizers import SGD
```

```
# Compile the model
optimizer = SGD(lr=0.01, momentum=0.9)
model.compile(optimizer=optimizer, loss='categorical_crossentropy', metrics=['accuracy'])
model.summary()

# Train the model for the first 25 epochs using data augmentation
history_1 = model.fit(datagen.flow(X_train, y_train, batch_size=32), epochs=25,
validation_data=datagen.flow(X_test, y_test))

# Save weights after the first 25 epochs
model.save_weights("/content/drive/My Drive/MobileNet_model_weights_25_epochs.h5")

# Evaluate the model on the test set after the first 25 epochs
loss_1, accuracy_1 = model.evaluate(datagen.flow(X_test, y_test))
print(f'Test accuracy after 25 epochs: {accuracy_1 * 100}%')

# Continue training for the next 25 epochs
history_2 = model.fit(datagen.flow(X_train, y_train, batch_size=32), epochs=50,
validation_data=datagen.flow(X_test, y_test), initial_epoch=25)
```

```
# Save weights after the complete training
model.save_weights("/content/drive/My Drive/MobileNet_model_weights_50_epochs.h5")

# Evaluate the model on the test set after 50 epochs
loss_2, accuracy_2 = model.evaluate(datagen.flow(X_test, y_test))
print(f'Test accuracy after 50 epochs: {accuracy_2 * 100}%')
```

EFFICIENTNET Code:

```
datagen = ImageDataGenerator(horizontal_flip=True,
                             vertical_flip=True,
                             zoom_range=0.3,
                             width_shift_range=0.2,
                             height_shift_range=0.2,
                             fill_mode='constant',
                             cval=0.1)

batch_size = 2

data_generator = datagen.flow(X_train,
                              y_train,
                              batch_size=batch_size,
                              seed=42)
from keras.applications import EfficientNetB5
from keras.models import Sequential
from keras.layers import GlobalAveragePooling2D, Dropout, Dense
from keras.optimizers import Adam
from keras.callbacks import ReduceLROnPlateau

model = Sequential()
model.add(EfficientNetB5(include_top=False, input_shape=(224, 224, 3)))
model.add(GlobalAveragePooling2D())
model.add(Dropout(0.5))
model.add(Dense(5, activation="sigmoid"))

from keras.optimizers import SGD
```

```
# Compile the model
optimizer = SGD(lr=0.01, momentum=0.9)
model.compile(optimizer=optimizer, loss='categorical_crossentropy', metrics=['accuracy'])
model.summary()
# Train the model for the first 25 epochs using data augmentation
history_1 = model.fit(datagen.flow(X_train, y_train, batch_size=32), epochs=25,
validation_data=datagen.flow(X_test, y_test))

# Save weights after the first 25 epochs
model.save_weights("/content/drive/My Drive/EfficientNet_model_weights_25_epochs.h5")

# Evaluate the model on the test set after the first 25 epochs
loss_1, accuracy_1 = model.evaluate(datagen.flow(X_test, y_test))
print(f'Test accuracy after 25 epochs: {accuracy_1 * 100}%')
```

```
# Continue training for the next 25 epochs
history_2 = model.fit(datagen.flow(X_train, y_train, batch_size=32), epochs=50,
validation_data=datagen.flow(X_test, y_test), initial_epoch=25)

# Save weights after the complete training
model.save_weights("/content/drive/My Drive/EfficientNet_model_weights_50_epochs.h5")

# Evaluate the model on the test set after 50 epochs
loss_2, accuracy_2 = model.evaluate(datagen.flow(X_test, y_test))
print(f'Test accuracy after 50 epochs: {accuracy_2 * 100}%')
```

HYBRID MODEL Code:

```
from keras.models import Model, load_model
from keras.layers import Input, Dense, concatenate
from keras.optimizers import SGD

# Load the pre-trained models
efficientnet_model = load_model('/content/drive/My Drive/CapstoneModels/EfficientNet.h5')
densenet_model = load_model('/content/drive/My Drive/CapstoneModels/DenseNet.h5')

# Remove the fully connected layers from EfficientNet
efficientnet_features = efficientnet_model.layers[-2].output

# Remove the first layers of DenseNet to accept 2-dimensional input
densenet_input = Input(shape=efficientnet_features.shape[1:])  # Input for EfficientNet
densenet_output = densenet_model(densenet_input)   # Connect the output of EfficientNet to the
input of DenseNet

# Concatenate features from both models
concatenated_features = concatenate([efficientnet_features, densenet_output])

# Output layer
output_layer = Dense(5, activation='softmax')(concatenated_features)

# Optimizer
sgd_optimizer = SGD(learning_rate=0.01, momentum=0.9)

# Define the hybrid model
hybrid_model = Model(inputs=densenet_input, outputs=output_layer)
```

```
# Summary of the hybrid model
hybrid_model.summary()

from keras.callbacks import EarlyStopping
from keras.preprocessing.image import ImageDataGenerator
```

```python
# Create an ImageDataGenerator for data augmentation
datagen = ImageDataGenerator(
    rotation_range=20,
    width_shift_range=0.2,
    height_shift_range=0.2,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True,
    fill_mode='nearest'
)

history_1 = hybrid_model.fit(datagen.flow(X_train, y_train, batch_size=32), epochs=10,
validation_data=datagen.flow(X_test, y_test))
loss_1, accuracy_1 = hybrid_model.evaluate(datagen.flow(X_test, y_test))
print(f'Test accuracy after 10 epochs: {accuracy_1 * 100}%')

history_2 = hybrid_model.fit(datagen.flow(X_train, y_train, batch_size=32), epochs=20,
validation_data=datagen.flow(X_test, y_test), initial_epoch=10)
loss_2, accuracy_2 = hybrid_model.evaluate(datagen.flow(X_test, y_test))
print(f'Test accuracy after 20 epochs: {accuracy_2 * 100}%')

history_3 = hybrid_model.fit(datagen.flow(X_train, y_train, batch_size=32), epochs=30,
validation_data=datagen.flow(X_test, y_test), initial_epoch=20)
loss_3, accuracy_3 = hybrid_model.evaluate(datagen.flow(X_test, y_test))
print(f'Test accuracy after 30 epochs: {accuracy_3 * 100}%')
```

REFERENCES

[1] Shelar, M., Gaitonde, S., Senthilkumar, A., Mundra, M. and Sarang, A., 2021, January. Detection of diabetic retinopathy and its classification from the fundus images. In 2021 international conference on computer communication and informatics (ICCCI) (pp. 1-6). IEEE.

[2] Kumar, E.S. and Bindu, C.S., 2021. MDCF: Multi-Disease Classification Framework On Fundus Image Using Ensemble Cnn Models. Journal of Jilin University, 40(09), pp.35-45.

[3] Mishra, S., Hanchate, S. and Saquib, Z., 2020, October. Diabetic retinopathy detection using deep learning. In 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE) (pp. 515-520). IEEE.

[4] Chetoui, M., Akhloufi, M.A. and Kardouchi, M., 2018, May. Diabetic retinopathy detection using machine learning and texture features. In 2018 IEEE Canadian conference on electrical & computer engineering (CCECE) (pp. 1-4). IEEE.

[5] Mushtaq, G. and Siddiqui, F., 2021, February. Detection of diabetic retinopathy using deep learning methodology. In IOP conference series: materials science and engineering (Vol. 1070, No. 1, p. 012049). IOP Publishing.

[6] Kumar, E.S. and Bindu, C.S., 2021, March. Segmentation of retinal lesions in fundus images: a patch based approach using encoder-decoder neural network. In 2021 7th International conference on advanced computing and communication systems (ICACCS) (Vol. 1, pp. 1247-1253). IEEE.

[7] Bajwa, A., Nosheen, N., Talpur, K.I. and Akram, S., 2023. A prospective study on diabetic retinopathy detection based on modify convolutional neural network using fundus images at sindh institute of ophthalmology & visual sciences. Diagnostics, 13(3), p.393.

[8] Gunawardhana, P.L., Jayathilake, R., Withanage, Y. and Ganegoda, G.U., 2020, December. Automatic diagnosis of diabetic retinopathy using machine learning: a review. In 2020 5th International Conference on Information Technology Research (ICITR) (pp. 1-6). IEEE.

[9] Sudarmadji, P.W., Pakan, P.D. and Dillak, R.Y., 2020, November. Diabetic retinopathy stages classification using improved deep learning. In 2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS) (pp. 104-109). IEEE.

[10] Gothane, S., Raju, K.S., Bhaskar, N. and Divya, G., 2022. Diabetic Retinopathy Detection Using Deep Learning. In Data Engineering and Intelligent Computing: Proceedings of 5th ICICC 2021, Volume 1 (pp. 387-393). Singapore: Springer Nature Singapore.

[11] Saranya, P., Umamaheswari, K.M., Sivaram, M., Jain, C. and Bagchi, D., 2021, January. Classification of different stages of diabetic retinopathy using convolutional neural networks. In 2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM) (pp. 59-64). IEEE.

[12] Ullah, N., Mohmand, M.I., Ullah, K., Gismalla, M.S., Ali, L., Khan, S.U. and Ullah, N., 2022. Diabetic Retinopathy Detection Using Genetic Algorithm-Based CNN Features and Error Correction Output Code SVM Framework Classification Model. Wireless Communications and Mobile Computing, 2022.

[13] Islam, K.T., Wijewickrema, S. and O'Leary, S., 2019, June. Identifying diabetic retinopathy from oct images using deep transfer learning with artificial neural networks. In 2019 IEEE 32nd international symposium on computer-based medical systems (CBMS) (pp. 281-286). IEEE.

[14] Pathak, K.C., Shah, R.B., Tharakan, R.R., Patel, B.N. and Jariwala, D.C., 2021, May. Diabetic Retinopathy Diagnosis and Categorization using Deep Learning-A Review. In 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1063-1069). IEEE.

[15] Singh, T.M., Bharali, P. and Bhuyan, C., 2019, February. Automated detection of diabetic retinopathy. In 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP) (pp. 1-6). IEEE.

[16] Q. Meng, Y. Hashimoto, S. Satoh, “Fundus image classification and retinal disease localization with limited supervision”, in: Asian Conference on Pattern Recognition, Springer, 2019, pp. 469–482.

[17] Pearl Mary S., Thanikaiselvan V., “Unified adaptive framework for contrast enhancement of blood vessels,” International Journal of Electrical and Computer Engineering (IJECE), Vol.10, No.1, February2020, pp. 767-777, ISSN: 2088-8708, DOI: 10.11591/ijece.v10i1.pp767-777

[18] Kumar, E.S., Bindu, C.S. Two-stage framework for optic disc segmentation and estimation of cup-to-disc ratio using deep learning technique. J Ambient Intell Human Comput (2021).

[19] S. H. Kassani, P. H. Kassani, M. J. Wesolowski, K. A. Schneider, and R. Deters, , “Classification of histopathological biopsy images using ensemble of deep learning networks,” arXiv preprint arXiv:1909.11870, 2019.

[20] N. Yalin, S. Alver and N. Uluhatun, ”Classification of retinal images with deep learning for early detection of diabetic retinopathy disease,” 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, 2018, pp.

[21] J. A. Hanley and B. J. McNeil, ‘The meaning and use of the area under a receiver operating characteristic (ROC) curve’, Radiology, vol. 143, no. 1, pp. 29–36, 1982.

[22] Y. Zong et al., (2020) "U-net Based Method for Automatic Hard Exudates Segmentation in Fundus Images Using Inception Module and Residual Connection," in IEEE Access, vol. 8, pp. 167225-167235, doi: 10.1109/ACCESS.2020.3023273.

[23] Menaouer, B.; Dermane, Z.; El Houda Kebir, N.; Matta, N. Diabetic Retinopathy

Classification Using Hybrid Deep Learning Approach. SN Comput. Sci. 2022, 3, 357.

[24] Gunasekaran, K.; Pitchai, R.; Chaitanya, G.K.; Selvaraj, D.; Annie Sheryl, S.; Almoallim, H.S.; Alharbi, S.A.; Raghavan, S.S.; Tesemma, B.G. A Deep Learning Framework for Earlier Prediction of Diabetic Retinopathy from Fundus Photographs. Biomed Res. Int. 2022, 2022, 1–15

[25] Khan, A.; Kulkarni, N.; Kumar, A.; Kamat, A. D-CNN and Image Processing Based Approach for Diabetic Retinopathy Classification. Appl. Inf. Process. Syst. 2022, 1354, 283–291