Enhanced Diabetic Retinopathy Classification Using DenseNet Models and Grad-CAM++

MCS

Author

Hafsa Ahmad

Registration No: 00000401450

Supervisor:

Assistant Prof Dr. Nauman Ali Khan

A thesis submitted to the faculty of Computer Software Engineering Department, Military College of Signals, National University of Sciences and Technology, Islamabad Pakistan in partial fulfillment of the requirements for the degree of MS in Computer Science.

(December 2024)

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by Ms **Hafsa Ahmad,** Registration No. **00000401450**, of **Military College of Signals** has been vetted by undersigned, found complete in all respect as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial, fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the student have been also incorporated in the said thesis.

Signature: _____

Name of Supervisor: **Asst Prof Dr. Nauman Ali Khan**

Date: **18th December 2024**

Signature (HoD): _____

Col
Head of Dept of CSE
MCS College of Sigs (NUST)

Date: 10/1/2025

Signature (Dean/Principal): _____

Date: 15/1/25

Brig
Dean, MCS (NUST)
Adil Masood Siddiqui, PhD

Form TH-4

NATIONAL UNIVERSITY OF SCIENCES & TECHNOLOGY

MASTER THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by **Hafsa Ahmad,** Regn No **00000401450** Titled: **"Enhanced Diabetic Retinopathy Classification Using Dense Net Models and Grad-CAM++"** be accepted in partial fulfillment of the requirements for the award of **MS Computer Science** degree.

Examination Committee Members

1. Name: **Assoc Prof Dr. Ihtesham Ul Islam** Signature: _____

2. Name: **Assoc Prof Dr. Javed Iqbal** Signature: _____

Co-Supervisor: **Asst Prof Dr. Maemoona Farooq** Signature: _____

Supervisor's Name: **Asst Prof Dr. Nauman Ali Khan** Signature: _____

Date: 13/1/25

Col
Head of Dept of CSE
Mil College of Sigs (NUST

Head of Department

14/1/25
Date

COUNTERSIGNED

Date: 15/1/25

Brig
Dean, NCS (NUST)
Dean
(Adil Masood Siddiqui, PhD)

CERTIFICATE OF APPROVAL

This is to certify that the research work presented in this thesis, entitled "**Enhanced Diabetic Retinopathy Classification Using DenseNet Models and Grad-CAM++**." was conducted by Miss. **Hafsa Ahmad** under the supervision of **Asst Prof Dr. Nauman Ali Khan**.

No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the **Department of Computer Software Engineering** in partial fulfillment of the requirements for the degree of Master of Science in Field of **Computer Science Department of Military College of Signals, National University of Sciences and Technology, Islamabad**.

Student Name: **Hafsa Ahmad** Signature:

Examination Committee:

a) Examiner 1: **Assoc Prof Dr Javed Iqbal** Signature:
(Department of Computer Software Engineering)

b) Examiner 2: **Assoc Prof Dr Ihtesham ul Islam** Signature:
(Department of Computer Software Engineering)

Name of Supervisor: **Asst Prof Dr. Nauman Ali Khan** Signature:

Name of Dean/HOD: **Col Usman Mehmood Malik, PhD** Signature:

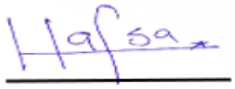
Col
Head of Dept of CSE
MCS, College of Sigs (NUST)

Plagiarim Undertaking

I solemnly declare that research work presented in the thesis titled **" Enhanced Diabetic Retinopathy Classification Using DenseNet Models and Grad-CAM++ "** is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and National University of Sciences and Technology (NUST), Islamabad towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS degree, the University reserves the rights to withdraw/revoke my MS degree and that HEC and NUST, Islamabad has the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized thesis.



Student Signature:

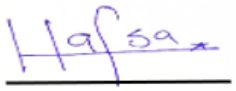
Name: **Hafsa Ahmad**

Date: 18 December 2024

Author's Declaration

I, Hafsa Ahmad, hereby state that my MS thesis titled "**Enhanced Diabetic Retinopathy Classification Using DenseNet Models and Grad-CAM++**" is my own work and has not been submitted previously by me for taking any degree from National University of Sciences and Technology, Islamabad or anywhere else in the country/ world.

At any time if my statement is found to be incorrect even after I graduate, the university has the right to withdraw my MS degree.

Student Signature: 

Name: **Hafsa Ahmad**

Date: 18 December 2024

Dedication

“In the name of Allah, the most Beneficent, the most Merciful"

Glory be to Allah Almighty, the Creator and Sustainer of the Universe, the Omnipotent and the Omnipresent. There is nothing I could have accomplished without His guidance and blessings.

I dedicate this thesis to my family, friends, and teachers, particularly my parents, who supported me each step of the way.

Acknowledgments

All praises to Allah S.W.T for the strength, guidance and blessings to complete this thesis.

I would like to convey my gratitude to my supervisor, Assistant Professor Dr. Nauman Ali Khan, PhD, for his supervision and constant support. His invaluable help and constructive comments during the experimental and thesis work are essential contributions to the success of this research.

I am deeply grateful to my parents, siblings (Mehwish Ahmad, Hanan Shehroze Khan, Engr. Osama Ali Ahmad Awan, and Dr. Wajiha Ahmad), instructors, and friends for their unwavering support and encouragement. Their belief in my dreams and ambitions has been a constant source of inspiration. I sincerely thank them for their love, care, and guidance during both the challenging and joyful moments of this journey.

Above all, my deepest gratitude goes to my parents, who are the foundation of everything I have achieved. Their unwavering belief in me, tireless sacrifices, and endless prayers have been the guiding light in my academic and personal journey. It is their strength, wisdom, and encouragement that have shaped me into the person I am today. Without their support, this milestone would not have been possible.

Abstract

Diabetic retinopathy (DR) is a severe condition of the eyes associated with metabolic abnormalities caused by diabetes, characterized by damage to the retinal vasculature. This may result in irreversible vision loss if not detected at early stages. Regardless of the essential importance of early intervention, the diagnosis of DR is often delayed due to the time-intensive nature of retinal assessments and the shortage of ophthalmologists. This highlights the urgent need for automated and accurate diagnostic tools to assist in the timely detection of DR. In this research, we introduce an advanced deep-learning framework utilizing Grad-CAM++ to enhance feature extraction and improve the precision of DR classification. The framework employs three DenseNet models, DenseNet-121(a), DenseNet-169, and DenseNet-121(b), to evaluate retinal images from the APTOS 2019 dataset, which includes 44,570 images classified into five stages of DR. Grad-CAM++ is used to highlight critical regions in the images, aiding in more accurate classification of early-stage DR. Model 3 outperformed the other models with a training accuracy of 95.63%, precision of 0.9921, recall of 0.9930, and an F1-score of 0.9925. These findings reflect the potential of our method to increase diagnostic accuracy and minimize the cost of healthcare systems. The proposed framework exhibits considerable advancements in the analysis of retinal images, offering an adaptive solution for early DR detection.

Keywords: DenseNet, Diabetic retinopathy detection, DR screening, Grad-Cam++, Image processing

Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

Abbreviations

DR	Diabetic Retinopathy
CNN	Convolution Neural Network
DenseNets	Densely Connected Networks
BN	Batch Normalization
CAD	Computer-Aided Diagnosis
ReLU	Rectified Linear Unit
OCT	Optical Coherence Tomography

CHAPTER 1

Introduction

The human eye stands as one of the most intricate and valuable sensory organs in the human body. Our sense of sight influences nearly every facet of daily life—determining how we navigate the world, perform tasks, recognize faces, and interpret countless visual cues. At the core of this visual capability is the retina, a delicate, multilayered tissue located at the inner back surface of the eye. The retina converts incoming light into neural signals, serving as a biological camera sensor that feeds the brain the raw information necessary to generate images. Any damage, disorder, or degenerative change within this tissue can set off a cascade of visual problems, potentially culminating in severe vision loss or even complete blindness. Among the multitude of conditions that can compromise retinal health, diabetic retinopathy Diabetic Retinopathy (DR) holds a particularly notorious status, affecting millions of people globally and posing a serious threat to eyesight if not promptly detected and managed.

Diabetic retinopathy is a side effect of diabetes, a metabolic disease characterized by chronically high blood glucose levels. Over time, high glucose damages the retina's small blood vessels, leading fluid leakage, hemorrhage, and the formation of aberrant blood vessels. For the hundreds of millions people worldwide who already have diabetes, there is a new source of concern:

diabetic retinopathy is on the rise, adding yet another layer of difficulty to their everyday health challenges. This dangerous retinal condition affects around 16 million individuals globally, and the number is expected to rise as the prevalence of diabetes rises. DR is more than simply a health statistic; it has significant effects on individuals, healthcare systems, and economies. Most importantly, early and precise screening can make the difference between sustaining irreversible vision damage and preserving usable vision, as DR is a major preventable cause of blindness in working-age adults.

Traditionally, the gold standard for detecting DR involves specialized ophthalmic examinations and image-based assessments using modalities like fundus photography and optical coherence tomography Optical Coherence Tomography (OCT). Fundus imaging provides a two-dimensional view of the retina, allowing clinicians to visualize changes in blood vessel structure, bleeding, and other abnormalities. On the other hand, OCT employs advanced light-based imaging to produce high-resolution, cross-sectional views of retinal layers. Together, these imaging techniques reveal subtle pathologies that might otherwise escape routine examinations. However, interpreting these images is not always straightforward. Even the most skilled ophthalmologists can face challenges when evaluating subtle retinal alterations, especially in large-scale screening contexts or in scenarios where disease progression must be carefully monitored over time. Manual analysis can be time-consuming, requires extensive training, and can suffer from human error and subjective variability.

The critical need to streamline and enhance DR detection and grading, coupled with advances in computational power and algorithms, has paved the way for automated diagnostic systems. In recent years, artificial intelligence (AI)—and in particular deep learning—has revolutionized the field of medical image analysis. Through the training of convolutional neural networks (CNNs) and related architectures, automated systems have demonstrated remarkable proficiency in ex-

tracting meaningful patterns from complex, high-resolution retinal images. Instead of relying on predefined, handcrafted features, these networks learn directly from the data, uncovering nuanced characteristics that correlate with the presence or severity of DR. The emergence of these systems holds the promise of faster, more consistent, and widely accessible screening tools. Such technologies can serve as a valuable complement to clinical expertise, helping to alleviate the burden on specialists and ensuring that patients receive earlier intervention when it matters most.

Deep learning methods for clinical uses would not be appealing based on only the accuracy of the algorithm. Furthermore, physicians and patients need to understand how such models make that decision before they put their trust on the models. In the past, the application of deep learning was unconvincing as it comes off as a 'black box' type of technology. There were no proper explanations as to how their predictions were generated. This lack of interpretability raised understandable concerns among clinicians, who must be able to validate and justify diagnostic decisions, especially when they lead to life-altering interventions. The necessity for interpretability has led researchers to develop methods to visualize and explain the internal decision-making processes of deep learning models. Gradient-weighted Class Activation Mapping and its enhanced versions (Grad-CAM++) are transformative techniques for explainable AI, particularly in medical imaging. These methods have changed how we perceive deep learning models, providing more clarity into their decision-making processes and increasing their usefulness in important clinical settings.

Grad-CAM was developed to generate spatial localization maps that indicate the parts of an image that have the greatest influence on a neural network's classification decision-making. This provided a valuable glimpse into the model's "reasoning," helping users ensure the system focuses on medically significant features rather than irrelevant image details. Grad-CAM++

enhances these heatmaps by offering finer and more precise localization of critical regions. Grad-CAM++ has the potential to highlight fine retinal abnormalities that might go unnoticed, even by experienced clinicians. This capability could promote the way for earlier diagnoses and more effective disease management strategies.

Deep learning models become more advanced and are able to inclnumerous image sources and large-scale datasets. These integration can improve diagnostic accuracy and generalizability if used effectively. Like combining fundus photography and OCT scans can provide a more complete picture of retinal structure. While fundus scans give broad insights on vascular integrity, OCT imaging looks deeper into the layered structure, maybe identifying early edema or tiny tissue changes that are not visible on the surface. Advanced AI models can use several data sources to present comprehensive assessments, which improve screening findings and aid in disease staging. The staging of DR—from moderate nonproliferative changes to proliferative DR—is crucial since intervention strategies and treatment regimens differ according to the severity of the disease.

While these developments are significant, challenges remain. One continuing challenge is data quality and variability. Retinal image datasets may vary in resolution, contrast, illumination, and patient demographics. Even subtle differences in imaging protocols, camera settings, or patient positioning can introduce noise into the classification process. Deep learning models must be robust enough to handle these variations, or at least be retrained or fine-tuned using domain adaptation techniques. Another challenge lies in the scarcity of large, well-annotated medical imaging datasets. While the field has benefited from public datasets and competitions, the process of labeling data often relies on expert graders and can be slow, expensive, and subject to inconsistency. Figure 1.1 illustrates the Diabetic Retinopathy image categorization based on the severity levels.Ongoing efforts aim to create standardized protocols and benchmarks to reduce

these inconsistencies and foster more reproducible research in DR screening and classification. Table 1.1 provides a record of clinical findings related to various stages of diabetic retinopathy.

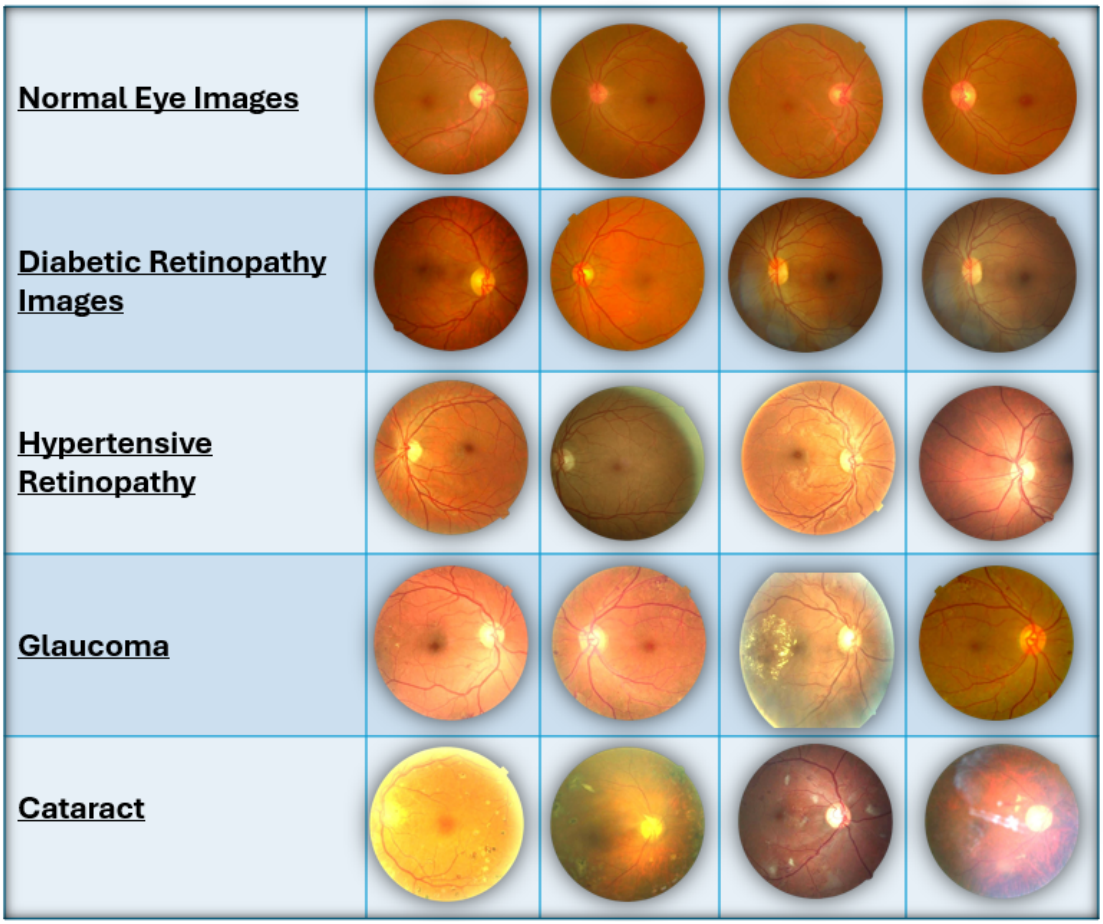


Figure 1.1: Representative images of diabetic retinopathy, showcasing various stages of disease severity.

Model interpretability and trust are also one of the conditions when clinical acceptance is taken into consideration. However, even if the system achieves a high level of performance in practice and research, it still has to prove that it can reason and its reasoning is acceptable from the clinician's point of view. Grad-CAM++ and its relatives allow bridging this gap by allowing ophthalmologists to see the areas focused by the network, and thereby assuring them that the logic of the model is plausible and conforms to well-known pathology. This degree of enhanced transparency is then of paramount importance as most trust the system now, which is an important factor for embedding AI systems into the systems. Besides, interpretability methods

Severity Level	**Condition**	**Description**
0	Normal Eye Image	The eye appears healthy with a clear lens, well-defined optic nerve head, normal intraocular pressure (IOP), and intact retinal vasculature. No abnormalities are evident on examination.
1	Diabetic Retinopathy	Early retinal microvascular changes due to diabetes may be present, including microaneurysms, dot-blot hemorrhages, and mild retinal edema. Vision is typically unimpaired at this stage, but careful monitoring and glycemic control are recommended.ion might be necessary for a definitive diagnosis.
2	Hypertensive Retinopathy	Retinal vascular alterations associated with systemic hypertension become evident, such as arteriolar narrowing, arteriovenous nicking, and occasional flame-shaped hemorrhages. These changes often reflect chronic blood pressure elevation and necessitate cardiovascular assessment and management.
3	Glaucoma	Increased intraocular pressure or reduced optic nerve perfusion leads to progressive optic nerve damage. Characteristic signs include optic disc cupping and thinning of the neuroretinal rim, accompanied by progressive visual field loss. Early detection and intervention are key to preserving vision.
4	Cataract	Opacification of the eye's lens results in cloudy or blurred vision, glare sensitivity, and potential color vision changes. As the cataract progresses, it significantly impairs visual acuity and may eventually require surgical intervention to restore clear vision.

Table 1.1: Diabetic Retinopathy diseases findings as per severity level.

form a very useful feedback, allowing for physicians to pass along insights in great detail to the developers. These insights are helpful in making progressive refinements in the model architecture, training techniques and even validation tests, making AI applications more acceptable and effective in the clinical fields.

The scope of automated diabetic retinopathy DR detection encompasses more than just increasing the precision and interpretability of the diagnosis, rather it is about expanding the extent and functionality of these systems. New models may be able to accomplish more than just diagnosing DR; they may be able to evaluate the extent of the disease, its prognosis, the outcome of therapy, or the risk of associated disease affecting other structures of the eye. Imagine a virtual aide that can pick out the earliest signs of DR as well as predict whether the disease will reach an advanced stage within the next five years. This could help the clinicians to identify patients more prone to develop complications and intervene appropriately.

The importance of this technology is not confined only to DR. Automated deep learning, methods of detection of other retinal diseases, like, for instance, age-related macular degeneration (AMD) and diabetic macular edema (DME) may also prove to be efficient. AI systems can be repurposed to treat a wide variety of retinopathies by relying on the same core concepts: feature extraction, classification, localization, and explainability. As these systems start to combine multiple disease types and imaging techniques, they may become all-in-one diagnostic solutions.

Such growth is even more impressive in resource-limited settings. As the telemedicine and remote diagnostics spread, AI-based instruments can improve the availability of effective screening for populations in countries with a poorly developed healthcare system. These systems may spare conventional in-person screening in the first steps so that specialized care could reach more patients in distant locations. Ultimately, this technology has a chance of reducing openly

blind patients the most in areas lacking an adequate number of ophthalmologists.

The diagnosis of diabetic retinopathy (DR) using deep learning methods is a perfect example of combining medical needs, technology, and human creativity. DR has become one of the top challenges in global vision health, emphasizing the need to develop effective, accurate, and easy-to-interpret algorithms. Though imaging techniques such as fundus photography or optical coherence tomography (OCT) are essential groundwork, the true diagnostic power of these images comes from the combination of convolutional neural networks (CNN) and advanced interpretability techniques like Grad-CAM and Grad-CAM++. These systems not only improve the effectiveness of the system but also provide some understanding of how the decisions were made, which will enable new and improved AI-based solutions in healthcare. In the future, it is critical for researchers and clinicians to collaborate closely. On one side, engineers and computer scientists will refine algorithms, improve interpretability, and ensure model robustness. On the other side, ophthalmologists, optometrists, and other vision specialists will guide the research toward clinically meaningful outputs, ensuring that the technology meets genuine patient needs and adheres to high ethical and professional standards.

The ongoing refinement and application of deep learning in retinal image analysis hold promise not only for improved detection of DR but also for fostering a new era of precision ophthalmology. In this future, AI-driven systems may become as routine as standard eye exams, providing clinicians with immediate, reliable insights and enabling earlier interventions that safeguard vision. The integration of explainable deep learning solutions thus represents a natural and necessary evolution in the quest to preserve and enhance human eyesight in the face of widespread and potentially devastating conditions like diabetic retinopathy.

1.1 Research Contribution

This work present an innovative deep learning framework designed to improve both the interpretability and accuracy of diabetic retinopathy (DR) classification. Traditional convolutional neural network (CNN) architectures have shown promise in discerning subtle retinal anomalies from fundus and OCT images, but their decisions often lack transparency. To address this challenge, we integrate Gradient-weighted Class Activation Mapping Plus Plus (Grad-CAM++) with conventional deep feature extraction methods. While CNNs excel at uncovering complex patterns from large datasets, they usually provide limited guidance on where these diagnostic cues reside within an image. By coupling such deep feature extraction with Grad-CAM++, our approach uncovers not only which features the model deems important but also pinpoint their exact location in the retinal images.

A crucial advantage of this hybrid approach is the improved capability to highlight lesion-specific features—such as microaneurysms, hemorrhages, or cotton wool spots—by generating more refined and focused localization maps. Grad-CAM++ enhances the original Grad-CAM by delivering sharper, higher-resolution attention maps that help clinicians and researchers understand the underlying reasoning behind classification decisions. Through this enhanced visualization, our model does more than simply identify whether or not DR is present—it illustrates which portions of the retinal image are driving that conclusion. This interpretability is essential for establishing trust in automated diagnostic systems, paving the way for more confident clinical decision-making.

In terms of performance, our integrated framework aims to achieve both heightened accuracy and consistency, reducing the error margins commonly observed in manual or conventional automated analyses. By leveraging deep learning's strength in pattern recognition and combining it with feature localization, the system is designed to deliver stable, reproducible results even

in the presence of image variability and heterogeneous patient conditions. The model has the potential to serve as a diagnostic assistant, offering rapid second opinions that free clinicians from the sole burden of time-consuming, manual image inspections.

Our approach sets an important foundation for future research in automated retinal disease detection. We aim to improve the use of explainable deep learning techniques in ophthalmology by proposing an approach that can be applied to various imaging systems and clinical situations. This may outcome in faster and more effective screening, earlier treatments and better outcomes for patients. As healthcare increasingly relies on digital solutions, our research underlines the importance of clear, intelligible, and accurate AI technologies to support the next generation of clinical procedures.

1.2 Objectives

The objective of this research is to utilize advanced deep learning and imaging techniques to improve the detection, classification, and understanding of diabetic retinopathy. It include developing a comprehensive workflow from data collection to model evaluation that can serve as a strong foundation for clinical incorporation. The objectives are as follows:

- **Comprehensive Data Acquisition:** To compile a diverse and comprehensive dataset of retinal images linked to diabetic retinopathy for ensuring the model performance. The dataset should include a diverse variety of illness stages and patient demographics. By compiling a diverse collection, we can create the framework for effectively training and evaluating the models.
- **Dataset Pre-processing and Normalization:** One of the most important steps before using datasets for a machine learning model is effective pre-processing, as the input data is

often subject to discrepancies or artifacts in the case of medical imaging, which can hurt performance. It is a collection of measures to improve the quality and consistency of the input data. To achieve this, scaling and normalization are employed to ensure the pixel intensity values remain in a consistent range, which also helps prevent numerical instabilities during training and allows the model to converge more quickly. Artifact removal in this context means removal of glare on the image, shadow or any other distortion in the retinal images obtained and only meaningful features will be retained to work with.

Data augmentation methods such as rotating, flipping, or changing brightness can also be used to enhance the variety of the dataset, replicating a more realistic environment and avoiding potential overfitting. Tracing images onto a certain sub-display resolution aligned with the requirements of some models' architecture is an other vital component. Such pre-processing not only normalizes datasets but also cleans them up and paves the way for the extraction of strong and relevant features. In essence, this enhances the robustness and precision of the model's predictions, which is especially critical in intricate tasks such as diabetic retinopathy classification, where slight differences in data may have a profound impact on diagnostic results.

- **Feature Extraction with Deep Learning:** CNN or convolutional neural networks have become a breakthrough technology in the automotive industry and medical imagingFeature extraction, the process of identifying important features, structures, and anomalies, has proved to be highly beneficial specifically in retinal images. These state-of-the-art networks leverage a deep and layered framework to construct hierarchical representation of features, beginning with low-level components like edges and textures which subsequently lead to higher level representations, including lesions and pathological abnormalities. Such a layered approach to analysis enables CNNs to discern nuances that classi-

cal image processing techniques sometimes overlook, rendering them highly valuable in medical diagnostics.

One of the greatest advantages of the targeted feature extraction using deep learning is its relative independence from conditions such as different image resolutions, lighting conditions and noise. Unlike conventional feature selection algorithms, CNNs operate directly on the data, which makes this method a lot easier and less prone to human error. This function is especially useful for identifying early symptoms of retinal disorders such as diabetic retinopathy and macular degeneration, where prompt and accurate detection is essential to avoid vision impairment.

The performance of CNNs is boosted even more by transfer learning and modification. These methods leverage pre-trained models so that networks can be fine-tuned for specific datasets, even with few training samples. This increases the working speed and guarantees better generalization of these models in various clinical conditions. Consequently, the classification accuracy is significantly improved, and approaches based on deep learning become more reliable and accurate for identifying retinal abnormalities.

Thereby, the implementation of CNNs for feature extraction in general can play a pivotal role in screening healthcare problems via automating retinal change analysis using retinal images in an effective manner. In addition to enhancing accuracy, these systems facilitate the diagnostic process, allowing healthcare providers to concentrate on treatment and patient care. The ability of AI to diagnose diverse diseases accurately and quickly is only expected to evolve and change the face of imaging in the years to come.

- **Data Augmentation for Robustness:** One of the strategies for each of the solutions proposed above is data augmentation, which is critical for increasing the versatility and accuracy of ML algorithms trained on medical images. Through various transformations

like random rotations, flips, or brightness/contrast adjustments, the training dataset is supplemented with diverse variations that approximate real imaging conditions. These adjustments aid the model in adapting to challenges that are often encountered in clinical settings, such as variations in lighting, camera settings, and patient demographics, maintaining reliable performance on new data.

More advanced techniques, which include zoom, crop and Gaussian noise, increase the dataset by introducing controlled variations. Such methods not only provide more data for training, they also train the model to attend to relevant features, irrespective of external distortions: certain patterns or abnormalities. For instance, in detecting diabetic retinopathy, augmentation allows the model to identify critical components such as microaneurysms or exudates even when images are slightly shifted, more dim, or brighter than normal.

Furthermore, data augmentation circumvents the frequent problem of class imbalance in medical datasets, where there is often a large disparity in the number of samples across different classes. Augmentation ensures that the model is learning across underrepresented classes fairly, minimizing prediction bias, as it artificially increases the representation of all classes. This leads to a system that is not only more precise but also better able to generalize in different clinical settings, tugging it an important component in capitalizing substantial and effectual diagnostic model.

- **Designing a Multi-Disease Deep Learning Architecture:** To treat diabetic retinopathy and other retinal diseases an improved neural network is designed. This improved model is more in line with real-world clinical scenarios since it can assess multiple diseases inside a single framework, giving ophthalmologists a versatile and comprehensive diagnostic tool to improve the treatment of patients.

- **Accurate Recognition and Classification:** A model is needed to accurately recognize and classify diabetic retinopathy, including its severity and consequences. This exact detection is essential for enabling early intervention, improved treatment methods, and, ultimately, improved patient outcomes.

- **Benchmarking Against State-of-the-Art Approaches:** To assess gains in accuracy, interpretability, and computational efficiency the suggested approaches analyze the existing model techniques. This study benchmarks to provide insights on the proposed approach's strengths and shortcomings and also areas for future improvement.

1.3 Reason/ Justification for the Selection of the Topic

There are various strong reasons and justifications for choosing the issue of applying machine learning (ML) and deep learning (DL) for Diabetic Retinopathy (DR) diagnosis and classification:

- Public Health Impact: Diabetic retinopathy (DR) is one of the most frequent diabetes sequelae worldwide, with rates continually rising. Early detection and exact diagnosis are crucial for improving patient outcomes and preventing vision loss. Machine Learning (ML) and Deep Learning (DL) have the potential to increase the efficiency and accuracy of Diabetic Retinopathy detection, hence assisting public health measures to limit the disease's consequences.

- Technological Advancements: Machine learning (ML) and deep learning (DL) have made significant developments in computer vision and image analysis. Their shown advancements in image classification and segmentation have established them as efficient tools for detecting diabetic retinopathy (DR) symptoms in retinal images, opening up new paths for

early detection and treatment.

- Clinical Need and Demand: Diabetic retinopathy (DR) remains a challenging diagnosis for ophthalmologists to identify, particularly when identifying between its numerous stages, which range from healthy eyes to mild, moderate, severe, and proliferative. Machine learning (ML) and deep learning (DL) systems provide essential assistance by providing precise and quick diagnostic information. These solutions address a critical therapeutic need by enhancing the accuracy and efficiency of DR detection and management, resulting in better patient outcomes.

- Research Opportunities: The combined use of healthcare and artificial intelligence (AI) creates interesting potential for innovation in medical. The collaboration has a chance to create advanced algorithms, approaches, and technologies that may change ophthalmic practices. Development of AI-driven techniques for diagnosing and controlling diabetic retinopathy (DR) has the potential to transform the way this condition is recognized and treated, resulting in improved patient care and substantial advances in the field of ophthalmology.

- Global Relevance: Diabetic retinopathy (DR) is a worldwide health problem that affects people from all walks of life. Machine learning (ML) and deep learning (DL) advancements have an exceptional capacity to change global healthcare delivery. These technologies have the ability to increase access to precise diagnostics, enhance early detection efforts, and fill healthcare service gaps. They have the ability to minimize shortcomings and improve the quality of care for patients worldwide.

- Interdisciplinary Collaboration: Computer scientists, engineers, medical experts, and ophthalmologists often work together on ML and DL research for diabetic retinopathy (DR) detection and classification. This interdisciplinary approach fosters information

exchange and innovation, leading to the development of more reliable and effective diagnostic tools and techniques.

The selection of the topic aligns with both the clinical need for improved diagnostic tools in ophthalmology and the technological advancements in AI and computer vision, offering significant potential for improving patient care and outcomes in the field of Diabetic Retinopathy (DR) detection and management.

1.4 Relevance to National Needs

This work makes several significant contributions to the classification of Diabetic Retinopathy (DR): The deep learning-based methodology for the classification of Diabetic Retinopathy (DR) holds significant relevance to national needs by addressing critical aspects of public health, healthcare accessibility, cost-effective screening, and data-driven planning. By focusing on the evaluation of Diabetic Retinopathy, our system contributes to the early detection and management of this prevalent diabetic complication. Utilizing mobile technologies aligns with national healthcare goals of improving accessibility, particularly in underserved areas, while the efficiency of deep learning models can facilitate cost-effective and widespread screening programs. The data generated by our DR classification system can inform evidence-based public health planning, aiding policymakers in targeted resource allocation. Furthermore, the integration of advanced technologies reflects a commitment to innovation and technological advancements in healthcare, reinforcing its potential impact on improving overall public health outcomes at the national level.

1.5 Advantages

The research offers following potential advantages:

- Public Health Impact: Diabetic Retinopathy (DR) is one of the most common complications of diabetes globally, with incidence rates steadily increasing. Early detection and accurate diagnosis are crucial for improving patient outcomes and preventing vision loss. Machine Learning (ML) and Deep Learning (DL) have the potential to enhance the efficiency and accuracy of Diabetic Retinopathy detection, thus contributing to public health initiatives aimed at reducing the burden of this disease.

- The proposed system will assist doctors and ophthalmologists in diagnosing and treating various stages of Diabetic Retinopathy (DR) early on.

- The processing speed of the proposed system is efficient, enabling rapid detection and classification of different stages of Diabetic Retinopathy.

- Once transformed into an application, the system will reduce the time and cost associated with hospital-based diagnosis procedures, improving accessibility and efficiency in healthcare delivery.

1.6 Areas of Application

This research and its resulting methodologies have the potential to significantly enhance visual healthcare across various professional and clinical settings. By delivering more accurate, explainable, and timely diagnostic insights, the proposed solutions can support practitioners and decision-makers in environments where eye health is a priority. Some key areas of application include:

- **Healthcare Environments:** Beyond the confines of specialized clinics, the technology is poised to make an impact across a broad range of healthcare facilities. General hospitals, community health centers, and telemedicine platforms can incorporate automated diabetic retinopathy screening into their existing workflows. In doing so, they can reach patients who may have limited access to eye care specialists, reducing the number of missed or delayed diagnoses. This wider accessibility can be especially valuable in remote or underserved regions, where early detection can prevent serious visual impairments and reduce the overall burden on healthcare systems.

- **Ophthalmological Hospitals:** Eye-focused medical centers and specialty hospitals often handle a high volume of complex cases, ranging from common conditions to rare pathologies. Integrating advanced AI-driven diagnostic tools allows ophthalmologists and retinal specialists to streamline their diagnostic processes. With automated systems that accurately classify multiple retinal diseases—including diabetic retinopathy—clinicians can focus their expertise on patient management, treatment planning, and surgical interventions rather than spending excessive time on routine screening tasks. This shift not only boosts operational efficiency but also enables specialists to provide more personalized, in-depth care.

- **Diagnostic Centers and Screening Programs:** Dedicated diagnostic facilities, imaging labs, and large-scale screening programs can benefit significantly from automated retinal analysis. Such centers often serve as the first line of detection, filtering a large flow of patients and identifying those who require referral to specialists. Incorporating the proposed technology helps these facilities rapidly and reliably flag abnormalities, ensuring that patients with early-stage diabetic retinopathy or other conditions do not slip through the cracks. This improvement in triaging accelerates the path to intervention, ultimately

contributing to better visual outcomes and reducing the strain on referral networks.

By applying this research in diverse settings—from primary healthcare hubs to advanced eye hospitals and diagnostic facilities—clinicians, policymakers, and patients alike stand to gain. The broader integration of intelligent, explainable screening tools will help ensure that timely and accurate retinal assessments become a standard part of comprehensive, accessible vision care worldwide.

1.7 Thesis Organization

This thesis consists of five chapters that discuss the research of each different area from introduction to literature review and then to methodology, results, and conclusion.

Chapter 1: This chapter includes the basic introduction, background,research motivation and research contribution.

Chapter 2: This chapter provides an overview of relevant literature, encompassing articles pertinent to the scope of this study.

Chapter 3: This chapter presents the proposed methodology.

Chapter 4: This chapter delivers the experimental results.

Chapter 5: This chapter describes the discussion.

Chapter 6: This chapter concludes the report and highlights the direction for future work.

CHAPTER 2

Related Work

The field of medical image interpretation has rapidly gained the attention of researchers, particularly in the application of deep learning algorithms for early and accurate disease diagnosis. Among these, diabetic retinopathy (DR) has emerged as a critical focus area due to its potential to cause irreversible blindness if not detected and managed promptly. The increasing prevalence of diabetes globally has only heightened the need for advanced diagnostic tools capable of providing reliable and efficient screening for DR. Deep learning, with its ability to process vast amounts of data and identify subtle patterns in complex images, has proven to be a game-changer in this domain.

The primary goal of this review is to situate the current research within the broader context of advancements in deep learning applications for ophthalmology. By examining existing studies, this review highlights the progress made in leveraging deep learning for DR diagnosis and explores the challenges that persist in this rapidly evolving field. Early detection of DR remains a significant hurdle, particularly in regions where access to specialized healthcare professionals is limited. Automated solutions powered by deep learning not only offer the potential to bridge this gap but also promise to revolutionize the way retinal diseases are diagnosed and monitored.

One of the key aspects examined in this literature overview is the performance of various deep learning models in the classification of DR severity levels. As detailed in Table 2.1, the comparative accuracy of different architectural models illustrates the ongoing improvements in diagnostic precision. Architectures such as convolutional neural networks (CNNs), DenseNets, and ResNets have shown outstanding performance in extracting significant data from retinal images, enabling precise classification of DR stages ranging from mild to severe. These improvements highlight the importance of selecting a suitable model architecture for generating accurate results.

We still struggle to build models that can generalize well across various datasets and imaging modalities/parameter settings. A nondeterminate image quality, differences in retinal imaging equipment, and biases in the dataset contribute errors to model performance. These aspects emphasize the importance of strong preprocessing methods and data augmentation for generalizing and reliably performing clinical tasks in real-world settings.

Alongside this, we should also work on making deep learning models more interpretable. Despite the charters, accuracy is only one of many important norms, but transparency can help understand and rationalize how these models make judgements, especially in clinical applications. Grad-CAM++ can offer useful visual insights by emphasizing regions of retinal images that affect model predictions. Such transparency allows clinicians to trust and verify if the output given by the system is indeed correct or reliable. This finally strengthens the correlation between diagnostics, technology, and clinical experience.

A large and diverse dataset is emphasized in the development and evaluation of deep learning models. Models that are robust and can cope with real-world challenges will benefit from including data from multiple demographics, imaging conditions, and illness phases. Some large scale datasets, like APTOS 2019, have given the discipline a boost, providing annotated retinal

images to train and test new algorithms. The challenge of aggregating large datasets, especially for rare disease variants and underrepresented cases, remains a significant challenge.

Setting aside geographical concerns, this article discusses more philosophic issues related to the deployment in practice of the developed deep learning models. Although these models have demonstrated remarkable accuracy in diagnosing several diseases, their application in real world scenarios raises a range of concerns. Preventing unauthorized access to data, having ethical standards to adhere to, and creating intuitive interfaces are crucial for the successful deployment of AI-based applications in the healthcare sector. More so, effective education of the healthcare staff on how these devices are used and interpreted is one of the best things that can be done in order to improve the quality of health care.

Great progress has been made by deep learning in screening and diagnosis of diabetic retinopathy and this has contributed to early detection and successful treatment. Despite this, as noted in the article, problems such as variability of data, degree of transparency of the model, seamless use of these systems in the clinics on a daily basis has to be dealt with before these systems can be fully realized. The presentation of this study within progressing medical image processing highlights the relevance of this research for ophthalmology innovation and synergy.As researchers continue to refine and expand these systems, the ultimate goal remains the same: to provide accessible, dependable, and interpretable solutions that improve patient care and change the way diabetic retinopathy is diagnosed and treated.

The study [17] investigates the application of DenseNet models in the diagnosis and classification of diabetic retinopathy (DR) from retinal images. It presents three different models derived from DenseNet-121 (a), DenseNet-169 and DenseNet-121 (b) frameworks with data fusion and feature fusion methods aimed at enhancing early detection of DR. The best performer managed to obtain 93.51% of training accuracy which proves that patients with diabetic retinopathy with

different levels of severity can be diagnosed correctly more than ninety percent of the time.

In [3], a deep learning framework is presented for diabetic retinopathy detection. Detailed models of impressive accuracy levels using DenseNet-121 architecture with transfer learning It was trained and validated on a dataset with 80 labeled retinal images which enables it to identify diabetic retinopathy and measure the severity of DR effectively. This method incorporates the accuracy and reliability of automated diagnosis of retinal disease to optimize it. ConclusionThe study highlights that, particularly in underserved settings, deep learning models may serve as scalable tools for the early detection of diabetic retinopathy.

The work [18] has detailed descriptions of DenseNet-based architecture designed for early diagnosis and grading of diabetic retinopathy. By adding a fully automated preprocessing step, data augmentation, and a more sophisticated feature extraction logic it helps to enhance classification performance. The model was tested on a large dataset comprising 13000 retinal fundus images, and the combination of the model with K-Nearest Neighbors (KNN) algorithm with 10 neighbors gave an astonishing classification accuracy of 98.40%. These results demonstrate the model's remarkable capability to diagnose diabetic retinopathy and suggest that it could help to improve diagnostic procedures in the clinics.

The researchers in [7] propose the development of a unified framework for the automatic detection and classification of diabetic retinopathy from retinal images. The method integrates the Faster-RCNN model and DenseNet-65 to ensure good feature extraction, which focuses on the classification of DR to five classes whose severity level differs. The developed automated framework in addition to being economically friendly is efficient which makes it suitable for use. When tested on a vast Kaggle data set, the model was effective in correctly identifying and classifying diabetic retinopathy making it useful in real world scenarios.

In [12] an innovative deep learning method for identifying Diabetic Retinopathy (DR) severity

is proposed. Are of the potential encoders are Convolutional Block Attention Module (CBAM) and the DenseNet169, both important in narrowing down for important features and form dense visual representations. This method classifies diabetic retinopathy (DR) severity from a single-Color Fundus Photograph (CFP). It's effective as the model trained on APTOS dataset achieved 97% for the first level and 82% for severity levels.

Automated deep learning techniques for diabetic retinopathy (DR) are highlighted in [15]. This paper employs a hybrid VGG16-XGBoost classifier and DenseNet121 model, with DenseNet121 achieving 97.30% accuracy. Balancing techniques were used to overcome class imbalance in the APTOS 2019 dataset, demonstrating the importance of data preprocessing for accurate results.

The [8] implemented an automated DR classification and staging model from fundus images using two deep learning architectures: DenseNet120 and ResNet50. Inspired by the success of CNNs, the study did excellently, specifically DenseNet120 outperformed ResNet50 which obtained an accuracy of 82.58%. However, some of the major problems that persist include: the expense of curating sufficient numbers of large quality datasets; the time required for models to be trained to a significant number of epochs; and even reliable multiclass classification. Future developments will focus on integrating hybrid models as well as optimized algorithms for increased accuracy and speed. However, it was interesting to examine potential impact of transfer learning methods on DR detection steps in workflow, in a less researched area covered by limited health-specialized workers in the field.

The study [11] utilizes advanced models—Xception, InceptionResNetV2, MobileNetV2, DenseNet121, and NASNetMobile—to classify DR, achieving the highest validation accuracy of 96.25% with InceptionResNetV2. Limitations include the need for larger datasets and high computational requirements.

The paper [22] focuses on improving DR detection accuracy by combining DenseNet-121 and

ResNet-50 models with explainable AI via the SHAP method. Using fundus images from the APTOS 2019 dataset, the ensembled models achieved an accuracy of 98.69%, sensitivity of 86.23%, specificity of 97.54%, and an F-score of 90.26%. Limitations include high computational costs for SHAP-based interpretations and memory requirements for image segmentation.

The research in [10] proposes a specialized deep learning framework combining CenterNet with DenseNet-100 to enhance the detection and classification of diabetic retinopathy (DR) and diabetic macular edema (DME) from retinal images. Using datasets such as APTOS 2019 and IDRiD, this model achieved accuracy rates of 97.93% and 98.10%, respectively. Key limitations include high computational demands and the need for broader real-world validation.

The study [14] explores DR classification using hybrid CNN models with ResNet and DenseNet architectures. DenseNet achieved 96.22% accuracy compared to CNN (75.61%) and ResNet (93.18%). Limitations include computational demands and the need for larger datasets to improve generalizability.

The study in [16] provides a comprehensive overview of deep learning approaches for DR stage classification via fundus images. The authors first train and evaluate 39 different convolutional neural network (CNN) architectures on multiple APTOS, MESSIDOR-1 and MESSIDOR-2 datasets before providing insights into their skills in solving DR classification problems. Comparison against other models, EfficientNetB1 providing highest accuracy and ability for meaningful feature extraction. Nevertheless, the study points out the challenges to be overcome such as the computational complexity of training deep models, as well as the class imbalance seen in retinal data sets.

The study [19] reviews techniques for classifying and segmenting diabetic retinopathy (DR) lesions, focusing on machine learning and deep learning methods to enhance detection accuracy. It examines preprocessing and segmentation strategies like CNNs and U-Net models, which

show significant results. Challenges include computational intensity, data imbalance across DR severity levels, and the need for precise lesion localization.

This work [4]. present a review of diabetic retinopathy (DR) detection methods, datasets, and evaluation metrics; with a focus on an automated system to aid in early diagnosis. It emphasizes the role of AI and deep learning in automating DR detection thus minimizing the work burden on ophthalmologists. Although there has been significant progress towards medical image classification that satisfies clinical needs, some barriers still exist. These challenges include large-scale, high-quality annotated datasets, and significant computational resources to effectively train complex models. Moreover, generalization in data-hungry scenarios is a challenging objective to achieve which is much meager in real-world applications. Future research will therefore focus on the methods to tackle underlying class imbalance whilst developing accurate and computationally efficient diagnostic systems in these areas. By doing this, next-generation diagnostic tools may become more affordable, accurate, and context-agnostic–now, multiple medical imaging contexts can be covered.

The research [23] employs DenseNet-121 for automatic feature extraction of DR images, achieving good accuracy on the APTOS dataset. While effective in identifying retinal abnormalities, challenges like high computational requirements and the need for optimized preprocessing limit its clinical application.

In [21], a deep learning based method for identifying and classifying diabetic retinopathy (DR) lines 6 CNN topologies like InceptionResnetV2, VGG16 and DenseNet121. A part of two-stage training process by employing feature extraction and fine-tuning leads to the augmentation of model performance. The best accuracy, 96.61%, was achieved using InceptionResnetV2 on a Kaggle dataset.

Grad-CAM is one such method which explains decisions of a deep learning model, e.g., in

[5]). Grad-CAM uses image localization maps for visual explanations emphasizing regions of an image that are important for making the prediction. Its high-resolution visualizations and computational intensity for complex models are just a few limits.

The work of the authors of [2] explains how convolutional neural networks (CNNs) with global average pooling (GAP) layers localize important parts of the image for classification purposes. GAP layers allow CNNs to focus on relevant areas within an image, without the need for elaborated pixel-oriented labels, only general image-level labels are required during training. This method improves localization of the areas in the images that activates the classification and is known as Class Activation Mapping (CAM). However, the challenges remain that training deep networks are computational expensive and need high-resolution images.

The study [6] introduces Score-CAM, a post-hoc visual explanation method for convolutional neural networks (CNNs) that eliminates reliance on gradients. Score-CAM uses forward-passing scores to weight activation maps, improving the fairness and clarity of visual explanations. It outperforms previous methods like Grad-CAM in recognition and localization tasks, though limitations include handling real-time applications.

This paper [9] explores Computer-Aided Diagnosis Computer-Aided Diagnosis (CAD) techniques for screening hypertensive retinopathy (HR) using fundus images. It highlights the application of advanced deep learning models, including DenseNet and ResNet, for extracting features and classifying HR-related abnormalities. While significant progress has been made in this field, challenges persist, particularly the need for more extensive and diverse datasets and enhanced methods for detecting early-stage HR with greater precision.

The paper [13] analyzes machine learning (ML) and deep learning (DL) algorithms for diabetic retinopathy (DR) detection using retinal fundus and OCTA images. It compares deep learning architectures such as CNN, ResNet, and Inception v3, concluding that deep learning signif-

icantly outperforms traditional machine learning techniques in terms of accuracy, sensitivity, and specificity. Despite these advancements, challenges remain, including the need for larger datasets, substantial computational resources, and improved methods for early-stage detection.

In [24], a review highlights the relevance of machine learning (ML) and artificial intelligence (AI) in diagnosing diabetic and hypertensive retinopathy using eye images. Models such as CNN, ResNet, and Inception have improved early diagnosis, though challenges include restricted datasets, high computation requirements, and real-time deployment limitations.

The paper in [1] outlines DR diagnosis methods for retinal image processing. For example, fundus images are enhanced using techniques including bottom-hat transform, histogram equalization, and median filtering. Challenges include computational intensity, and balancing feature suppression while minimizing noise in the result.

In [20], using high-resolution fundus images the authors have utilized a deep learning ensemble method which is beneficial not only for accuracy but also for comprehensive discovery of diabetic retinopathy (DR). This method combines several models, utilizing each model's individual strengths, thereby improving the quality of DR classification. In particular, the ensemble approach shows significant improvements, as it is effective in detecting early signs of DR better than single-model frameworks. Nevertheless, this approach demands powerful computational capabilities and large, high-fidelity, labelled datasets to ensure reliable generalisation across similar cohorts of patients.

Ref	Method	Accuracy / Description
[17]	DenseNet121(a), DenseNet169, DenseNet121(b)	Model-1 = 83.90%, Model-2 = 89.19%, Model-3 = 93.51%
[3]	DenseNet121	98.49%
[18]	Efficient DenseNet	98.40%
[7]	Faster-RCNN	97.2%
[12]	DenseNet169	97%
[15]	A Hybrid model merging VGG16, XGBoost & DenseNet-121	79.50%, 97.30%
[8]	CNN models [ResNet 50 and DenseNet120]	82.58%
[11]	Xception, InceptionResNetV2, MobileNetV2, DenseNet121, and NASNetMobile	86.25%, 96.25%, 93.75%, 81.25%, and 80.00%
[22]	Shapley Additive Ensembled DenseNet-121 ResNet-50	98.69%
[10]	DenseNet-100, CenterNet	97.93%, 98.10%
[14]	CNN, hybrid CNN with ResNet, hybrid CNN with DenseNet 2.1	96.22%, 93.18%, 75.61%
[23]	DenseNet-121	97%
[21]	InceptionResnetV2, VGG16, VGG19, DenseNet121, MobileNetV2, and Efficient-Net2L	96.61%
[20]	Ensemble Model (Mobile Net, Xception, ResNet-50V2, DenseNet201, DenseNet-169, Inception-V3 and InceptionResnet-V2)	92.23%

Table 2.1: Methodological Approaches and Accuracy of Deep Learning Models: A Literature Review.

CHAPTER 3

Proposed Methodology

We propose a systematic framework for deep learning based retinal image Classification. The methodology also employs state-of-the-art preprocessing techniques and feature extractions, as well as its robustness and consistency in providing diagnostic predictions. Each phase of the methodology is defined in this chapter, namely image acquisition and preprocessing, Grad-CAM++ for feature extraction and DenseNet models for classification. The aim is to improve the detection and classification of diabetic retinopathy and produce interpretable designs for clinical decision making.

The study's methodology follows a structured approach to transform raw retinal images into meaningful diagnostic insights. As illustrated in Figure 3.5, the process begins with the careful acquisition of color fundus images. This is followed by a series of steps designed to enhance the data, extract key features, make predictions, and classify each image into specific retinal conditions. These steps are carefully designed to ensure accurate, consistent, and interpretable results.

Each level is outlined below:

1. **Data Acquisition:** This study utilizes the use of the APTOS 2019 dataset, containing

44,570 retinal fundus images. The dataset consists of five distinctive classes, each illustrating a clinical condition:

- Class 0 (Normal): Pathology-free retinal pictures.
- Class 1 (Diabetic Retinopathy): Early signs of diabetes-induced retinal damage.
- Class 2 (Hypertensive Retinopathy): Hypertension causes retinal vascular abnormalities.
- Class 3 (Glaucoma): Increased intraocular pressure causes damage to the optic nerve.
- Class 4 (Cataract): Clouding of the eye's lens reduces visual clarity.

Figure 3.1 represent different grades of Diabetic Retinopathy distribution in different datasets subset.

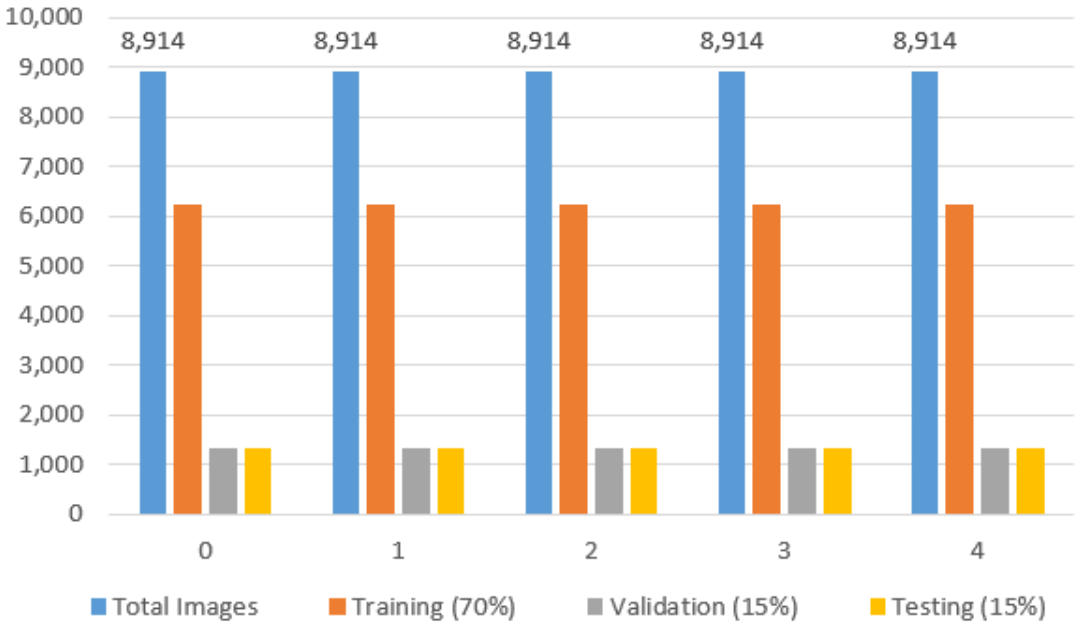


Figure 3.1: Distribution of Diabetic Retinopathy grades in different datasets subsets.

Data Preprocessing: The combination of preprocessing techniques resulted greatly boosted

the model, making it able to classify diabetic retinopathy. Again, these mechanisms were meticulously applied to refine the purity of input data bytes while suppressing potential noise in the learning pattern.

Data augmentation was one of most effective preprocessing techniques and this provides the most diversity in the training dataset. This was done with transformations like rotation, flipping, brightness adjustment, and zooming. Such variations allowed the model to be better able to generalise and systematically adapt to a variety of real-world imaging scenarios.

Moreover, all images were standardized to a size of 224 x 224 PIxels Normalizing these images helped make the dataset consistent and provided overall enhanced computational efficiency to the deep learning models. A key aspect of these images was that they had the same dimensions, allowing the models to identify patterns that indicated useful information without having to be deformed due to the object size differences. Figure 3.2 illustrates examples of the data augmentation techniques used.

One important aspect of the preprocessing step was the use of Grad-CAM++, a novel visualization technique that helps with interpretability. This approach resulted in visualizations being formed in the form of heatmaps depicting areas of importance in retinal images that contribute to the model's classification decisions, as shown in Figure 3.3. Grad-CAM++ provided clear, interpretable insights into the model's decision-making process, by focusing on the key features lesions, microaneurysms and hemorrhages. These visualizations were critical in trusting the model's outputs and maintaining transparency.

Using data augmentation and image resizing, along with Grad-CAM++ visualization, proved essential in building more robust and consistent models. The image preprocessing pipeline addressed issues stemming from different levels of image quality, so it predicted

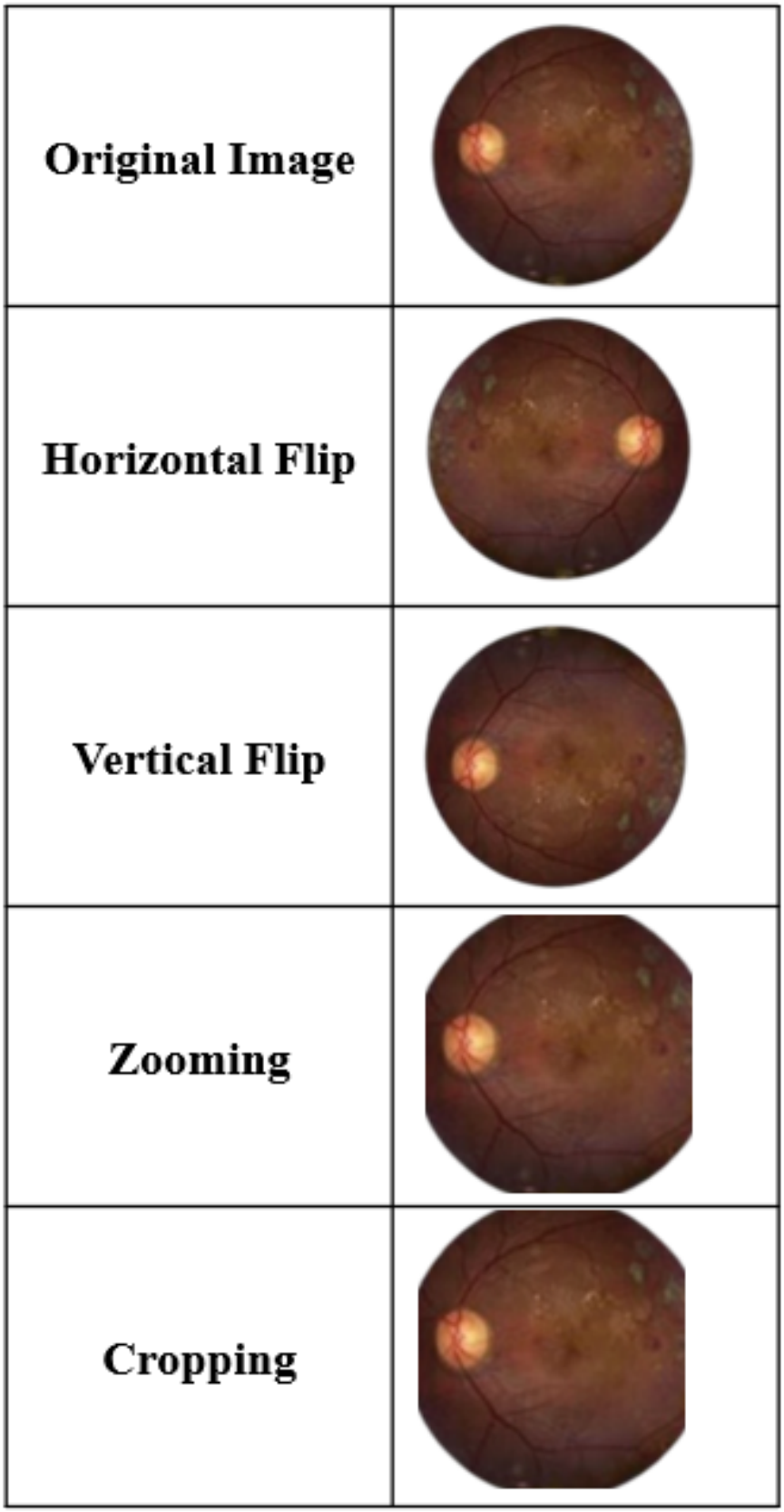


Figure 3.2: Image augmentation techniques.

the same regardless of brightness, contrast, or sharpness differences. As a result, the combination added robustness of estimates while still being intuitive to clinicians by offering insight into how and which factors were influencing the classification outcome.

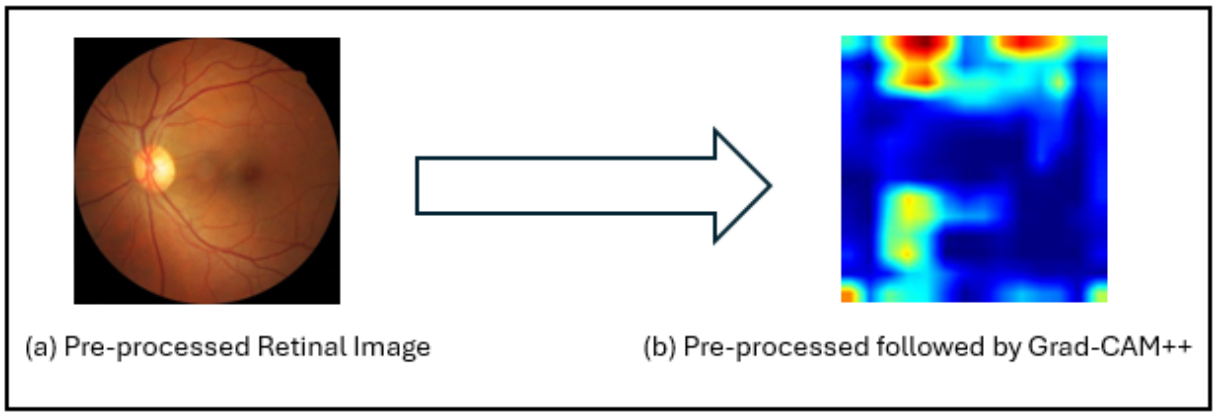


Figure 3.3: Showed an example of image preprocessing in which the Grad-CAM++ has been applied.

2. **Feature Extraction:** A DenseNet architecture served as the backbone of the automated feature extraction, indicating the networks' potential of disentangling complex and high-level features within retinal images. When the vast majority of image classification tasks can be achieved through training a model on several images, the same can also be utilized in a limited sense on these models due to how well it is able to identify critical features like edges, textures and retinal irregularities which aid in identifying stages of diabetic retinopathy. The highly interconnected design of these networks encourages efficient information transfer across the layers, retaining only the essential features of the data and eliminating unnecessary data redundancy. This positive response, integrating text, graph, and image data, allows the models to retain high, fine details such as microaneurysms, small hemorrhages, and larger organ structural patterns, including blood vessel irregularities and tissue changes.

   Unlike traditional methods, DenseNet models emphasize feature reuse across layers, al-

lowing them to learn subtle variations in retinal images without a significant increase in computational demands. This not only improves the accuracy of classification but also allows the network to generalize well across diverse datasets with varying image qualities. As a result, the use of DenseNet models in this study contributes significantly to building a robust and reliable diagnostic framework, capable of identifying critical retinal abnormalities that inform precise and early diagnosis of diabetic retinopathy.

3. **Prediction Module:** The extracted features from the DenseNet model were fed into a prediction module designed to classify the images into different stages of diabetic retinopathy. This prediction module consisted of a combination of convolutional layers and fully connected (dense) layers, which worked together to refine the learned features and map them to specific output categories. The convolutional layers further processed the extracted features to identify deeper patterns and relationships, while the dense layers consolidated this information to make the final classification decisions.

   To ensure the predictions were accurate and reliable, multiple loss functions were utilized during training. These included the standard classification loss, as well as specialized metrics such as heatmap loss and offset loss. Heatmap loss played a critical role in directing the model's attention to the most relevant regions of retinal images for classification, ensuring that key abnormalities such as microaneurysms and hemorrhages were appropriately emphasized. Offset loss was also utilized to improve the spatial alignment of features, ensuring greater precision in predictions by accounting for minor variations in image positioning or feature localization.

   By combining the prediction module with these thoughtfully designed loss functions, a strong and reliable framework for accurate classification was created. This strategy enhanced the model's ability to detect subtle differences between stages of diabetic retinopa-

thy while reducing prediction errors, making it a more dependable tool for clinical applications.

4. **Classification** The images of retina obtained after preprocessing were classified in to 5 classes, namely Normal, DR, HR, Glaucoma and Cataract, as shown in Figure 3.4. This categorization was made possible by a DenseNet-based system explicitly trained to differentiate between these disorders. Utilizing its sophisticated feature extraction mechanisms, the model examined complex structures and anomalies present in the images, facilitating accurate detection of retinal diseases.

   Each class represented distinctive features including typical retinal anatomy seen in the imaging of the "Normal" category and pathological components of diseases such as DR (Diabetic Retinopathy), HR (Hypotensive Retinopathy) and disease features associated with Glaucoma or Cataracts. The model identified, for example, microaneurysms, hemorrhages, and exudates for DR; optic nerve changes and structural damage for Glaucoma; abnormal vascular morphology for HR; and lens opacity for Cataracts.

   The DenseNet-based model demonstrated remarkable robustness in managing complex and overlapping retinal conditions. Its densely connected architecture ensured efficient feature propagation, preserving critical information and enabling accurate decision-making. By automating the detection and classification of multiple retinal diseases, it not only improves diagnostic accuracy but also facilitates early intervention, ultimately enhancing patient outcomes and streamlining clinical workflows.

3.0.1 Image Preprocessing

In the diabetic retinopathy dataset, the unprocessed fundus images are highly difficult for accurate prediction. So, such as blurred areas, inconsistent color variation, and partial visibility

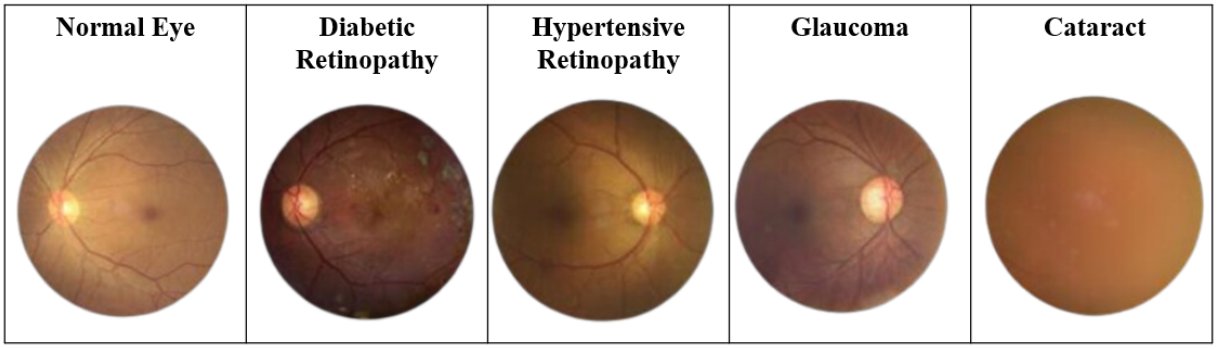


Figure 3.4: Classification Categories

of some critical diagnostic areas, all of which are obstacles to the reconstruction process of deep learning models. In order to overcome these problems and enhance the reliability of the classification results, a comprehensive preprocessing procedure was implemented.

The high-resolution fundus images were first resized to 224 × 224 pixels. It also provided consistency across the dataset and subsequently decreased computational load and memory consumption during model building. This would allow the model to learn on the patterns and features rather than being biased by size differences when images were input into the model. Furthermore, this step reduced the potential for overfitting, ensuring that irrelevant features in higher dimensional images did not overwhelm the learning process.

In this phase of the process, critical regions of the fundus, including the macula, optic disc, and blood vessels were emphasized. The model achieved increased sensitivity and identified subtle signs of disease progression by conditioning on these diagnostically relevant regions. The WH-based focus improved the precision as well as speed of the classification process.

These preprocessing steps also helped in situations such as areas with uneven lighting and inconsistent color distribution. Such variations were limited by standardizing the image dimension and focusing on the regions of interest, guaranteeing higher-quality input data. Collectively, these steps created a comprehensive preprocessing pipeline that laid the groundwork for auto-

mated diagnostic systems to yield accurate and consistent outcomes.

Hence, the GradientWeighted Class Activation Mapping++ (Grad-CAM++) was also used to enhance model interpretability. Using Grad-CAM++, we produced heatmaps that illustrated the important regions of retinal images that contributed to the final model prediction. These were heatmaps gave identified features that provided important diagnostic features, such as microaneurysms and hemorrhages providing insight into the process of the model decision making process. Overlaying these heatmaps on the source images provided transparent predictions explanation via Grad-CAM++.

Grad-CAM++ had several advantages above all others, and especially for healthcare professionals. This connected the outputs of the model to actual clinical evidence, providing confidence in the system. The predictions could be validated by clinicians through examination of the highlighted regions, thereby increasing confidence in their diagnostic decision-making. While traditional "black-box" models struggled to reconcile sophisticated computations with human interpretation, Grad-CAM++ did just that.

This technique also helped to refine our model as we trained it. Researchers could review the heatmaps to check whether the model had learned to focus on relevant diagnostic features, or whether it was potentially learning something that was misaligned with expectations for how it should learn. This iterative cycle of providing feedback improved the reliability and robustness of the model.

In practice for the case at hand, Grad-CAM++ was used to visualize if the model was misdiagnosing an image due to looking at an area that was not even of interest, e.g. an artifact or an area that is not diagnostic. This enabled greater precision but also made the model's outputs clinically relevant. Grad-CAM++ added transparency and usability, reinforcing the value of the system for real-world diagnostic tasks.

Grabbing-GAM++ was very crucial for the interpretability and reliability of the deep learning model. Last but not least, its capability of giving detailed visual explanations of the learned features contributing to the network decisionss marked a bridge between artificial intelligence and clinical practice. By being able to see and understand how their model arrived at the given results, clinicians could have confidence that the system could be smoothly integrated into patient care, indicating a major step forward in designing trustworthy tools for the diagnosis of diabetic retinopathy.

Algorithm 3.1. *Dataset $I_{preprocessed}$, CNN Model M_{CNN} Feature Maps $F_{enhanced}$.*

*1: **INPUT** Resize images to* 224×224 *and crop irrelevant regions.*

2: Load $I_{preprocessed}$ and define layer L in M_{CNN}.

3: For each I in $I_{preprocessed}$ Forward pass I through M_{CNN} to obtain y_c.

4: Calculate gradients $\frac{\partial y_c}{\partial A^k}$ of layer L.

5: Compute weights α_k^c: $\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial^2 y_c}{\partial (A_{ij}^k)^2}$.

6: Generate Grad-CAM++ map: $L^c_{Grad-CAM++} = \sum_k \alpha_k^c A^k$.

7: Apply ReLU and normalize $L^c_{Grad-CAM++}$.

8: Overlay and save as $F_{enhanced}$.

9: ***OUTPUT****: Return* $F_{enhanced}$.

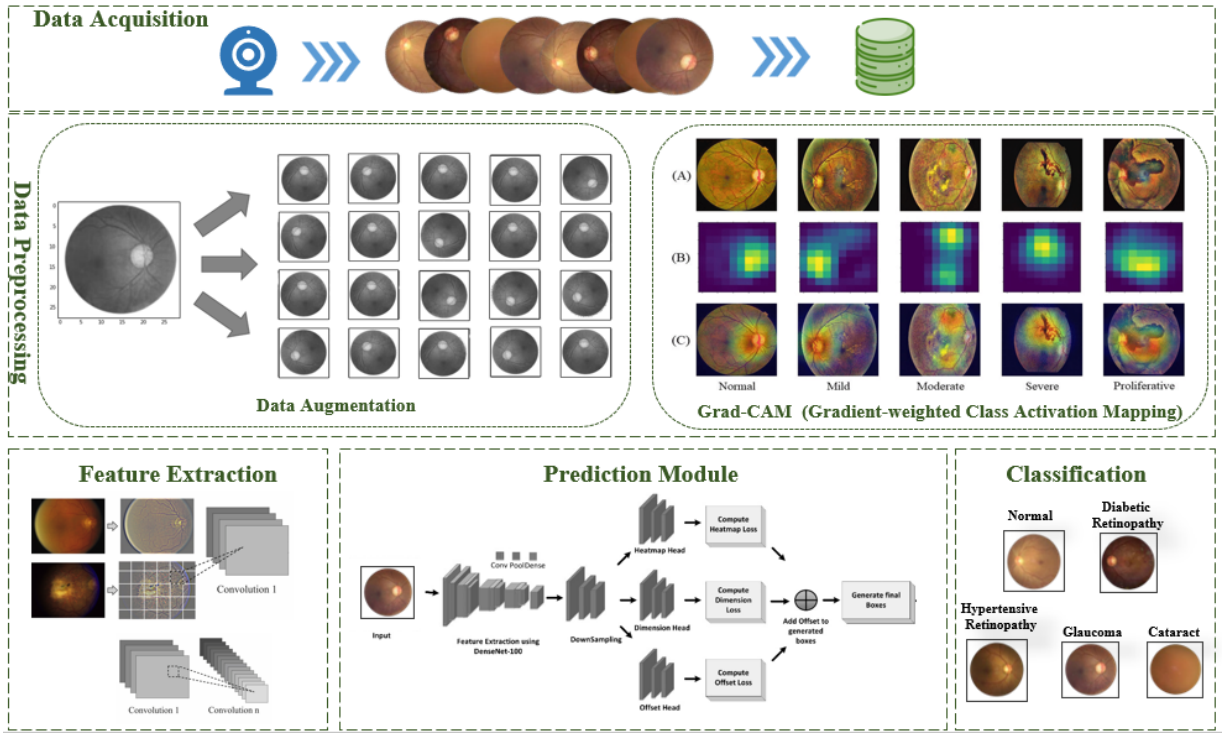


Figure 3.5: This diagram succinctly illustrates the research methodology, outlining the workflow from data acquisition to classification, including preprocessing, feature extraction, and prediction stages for accurate eye condition diagnosis.

3.0.2 Proposed CNN model:

The proposed convolutional neural network Convolution Neural Network (CNN) model is built upon Densely Connected Convolutional Networks Densely Connected Networks (DenseNets), which serve as the backbone for achieving effective classification of diabetic retinopathy DR. DenseNets were specifically chosen for their unique ability to address several key challenges in deep learning, including the vanishing gradient problem and inefficiencies in feature utilization.

By leveraging the dense connectivity pattern inherent to this architecture, DenseNets enhance feature propagation, promote the efficient reuse of features across layers, and significantly reduce parameter redundancy. These characteristics enable the development of a streamlined, high-performing model that is not only computationally efficient but also highly accurate in classifying retinal images.

A defining feature of DenseNets is their unique dense connectivity, which sets them apart from traditional convolutional networks. In this architecture, each layer takes inputs from all the layers before it and shares its feature maps with all the layers that follow. This architecture enables the network to maintain significant information from previous layers during the training phase, reducing the chances of essential information getting lost. As such, with their highly efficient use of memory, DenseNets excel at recognizing both low- and high-level features, or shapes, in retinal scans, such as edges and textures, as well as abnormalities and lesions all important aspects for correct dosage classification of diabetic retinopathy.

Mathematically, this connectivity can be expressed as follows: for a given layer l, its input consists of the concatenated feature maps from all previous layers $[x_0, x_1, ..., x_{l-1}]$. Using this formulation, every layer has a direct pathway to the gradients produced by the loss function and the original input data, making DenseNets very resilient to the vanishing gradient problem. Additionally, this design reduces the necessity of relearning redundant elements since previously computed feature maps can be reused throughout the network. This not only renders DenseNets computationally efficient, but significantly mitigates overfitting — a trait very important in the context of medical image analysis, even with less number of examples.

DenseNets are notable for their super-efficient use of parameters. In contrast to conventional networks that depend on significantly greater parameters for precise predictions, and consequently incur high memory and computation costs, DenseNets boast state-of-the-art performance with

a lesser number of parameters. This has to do with their ability to reuse features, which makes them very effective for use in medical images where computational resources are scarce and datasets limited.

DenseNets have a few key advantages when it comes to classifying diabetic retinopathy. By preserving and reusing information at all levels, subtle structures and irregularities in the retinal images, including microaneurysms, hemorrhages, and exudates, are sensibly captured and used for effective classification. The ability to capture multi-scale features allows DenseNets to discriminate different degrees of diabetic retinopathy, from mild to severe stages, with remarkable accuracy. Furthermore, their simple architecture also shortens training time and reduces computational needs, facilitating integration into real-life clinic settings.

DenseNets possess another key strength in their ability to generalize robustly which is a must for handling the variability often present in medical datasets. Retinal images vary significantly from device to device, patient to patient, and even from one environmental condition to another. The ability of DenseNets to propagate and reuse features allows the model to adapt well to these variations without losing accuracy.

The addition of a DenseNet architecture makes a notable advancement in the automatic classification of diabetic retinopathy by CNN. DenseNets drop solve the problems like the vanishing gradient and the redundant computations making it a robust, efficient, and very accurate solution. These models are well-suited to serve as a powerful and flexible base for scalable diagnostic solutions, equipping clinicians to identify and manage diabetic retinopathy with accuracy and confidence. This architecture is mathematically expressed as:

$$\mathbb{T}_x = \mathbb{T}_0 \oplus [\mathbb{H}_0, \mathbb{H}1, \ldots, \mathbb{H}x-1] \tag{3.0.1}$$

Here, $\oplus$ denotes the concatenation operation, $\mathbb{T}_0$ represents the input image, and $\mathbb{T}_x$ is the output

of the x^{th} layer, formed by concatenating the feature maps $\mathbb{H}0$ to $\mathbb{H}x-1$ from all preceding layers.

Within DenseNets, layers are organized into Dense Blocks, maintaining constant spatial dimensions of feature maps while allowing the number of feature maps to grow. The growth in feature maps is governed by the growth rate k, defined as the number of feature maps each layer adds. The total number of feature maps $\mathbb{Z}_x$ at the x^{th} layer is given by:

$$\mathbb{Z}_x = \mathbb{Z}_0 + k \cdot (x-1) \tag{3.0.2}$$

where $\mathbb{Z}_0$ is the initial number of feature maps. This linear growth ensures a balance between model capacity and computational efficiency.

Transition Layers To control the complexity and prevent excessive growth of feature maps, DenseNets incorporate Transition Layers between Dense Blocks. A Transition Layer typically comprises a Batch Normalization Batch Normalization (BN) layer, a Rectified Linear Unit Rectified Linear Unit (ReLU) activation, a 1×1 convolutional layer for dimensionality reduction, and a 2×2 average pooling layer. The compression factor θ (where $0 < \theta \leq 1$) is utilized to reduce the number of feature maps $\mathbb{Z}_{\text{trans}}$ in the Transition Layer:

$$\mathbb{Z}\text{trans} = \theta \cdot \mathbb{Z}\text{input} \tag{3.0.3}$$

Here, $\mathbb{Z}_{\text{input}}$ denotes the number of feature maps entering the Transition Layer.

Convolutional Operations and Feature Transformation Each layer within a Dense Block performs a series of operations to transform its input feature maps. The transformation process can be described as:

$$\mathbb{H}_x = \mathscr{F}(\mathbb{T}_x) = \text{ReLU}(\text{BN}(\mathbb{T}_x)) * \mathbb{W}_x \tag{3.0.4}$$

In this equation:

$\mathscr{F}(\cdot)$ represents the composite function of Batch Normalization (BN), ReLU activation, and convolution. $*$ denotes the convolution operation. $\mathbb{W}_x$ represents the convolutional weights at layer x. Overall DenseNet Structure The overall DenseNet architecture for DR classification can be summarized through the following key equations:

Layer-wise Feature Concatenation:

$$\mathbb{T}_x = \mathbb{T}_0 \oplus [\mathbb{H}_0, \mathbb{H}1, \ldots, \mathbb{H}x-1] \tag{3.0.5}$$

Growth of Feature Maps:

$$\mathbb{Z}_x = \mathbb{Z}_0 + k \cdot (x-1) \tag{3.0.6}$$

Transition Layer Compression:

$$\mathbb{Z}\text{trans} = \theta \cdot \mathbb{Z}\text{input} \tag{3.0.7}$$

Feature Transformation within Layers:

$$\mathbb{H}_x = \mathscr{F}(\mathbb{T}_x) = \text{ReLU}(\text{BN}(\mathbb{T}_x)) * \mathbb{W}_x \tag{3.0.8}$$

Classification and Loss Function For the classification task, the final feature maps are aggregated and passed through a Fully Connected (FC) layer followed by a softmax activation to obtain class probabilities. The classification loss is computed using the categorical cross-entropy loss function:

$$\mathscr{L} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{3.0.9}$$

where:

C is the number of classes. y_i is the true label for class i. $\hat{y}_i$ is the predicted probability for class i.

D121-[M1] Architecture

The D121-[M1] architecture was used as the foundation for the initial model, utilizing pre-trained ImageNet weights to enhance feature extraction. Input images were resized to 224 × 224 pixels with three RGB channels to meet model requirements. The top layers were removed for task-specific customization, incorporating a Global Average Pooling layer to condense feature maps, a Dropout layer (0.5 rate) to mitigate overfitting, and a Dense layer with softmax activation for probability distribution across diagnostic categories. The D121-[M1] backbone layers were initially frozen, fine-tuning only the final five layers for dataset adaptation. The model employed Adam optimization with a 1e-4 learning rate and categorical cross-entropy as the loss function. Accuracy was the primary performance metric to align with classification objectives. Layers of D169-[M2] Architecture is shown in Figure 3.6

D169-[M2] Architecture

The D169-[M2] architecture began with a 7x7 convolutional layer (64 filters, stride 2) followed by a 3x3 max-pooling layer (stride 2). Its core consisted of four dense blocks with layer repetitions: six in the first block, twelve in the second, and thirty-two in the third and fourth. Each block applied 1x1 convolution for dimensionality reduction and 3x3 convolution for feature extraction. Transition layers, separating the dense blocks, used 1x1 convolution followed by 2x2 average pooling (stride 2) to streamline feature maps. A global average pooling layer condensed features, followed by a dropout layer (0.5 rate) to mitigate overfitting, and a fully connected layer

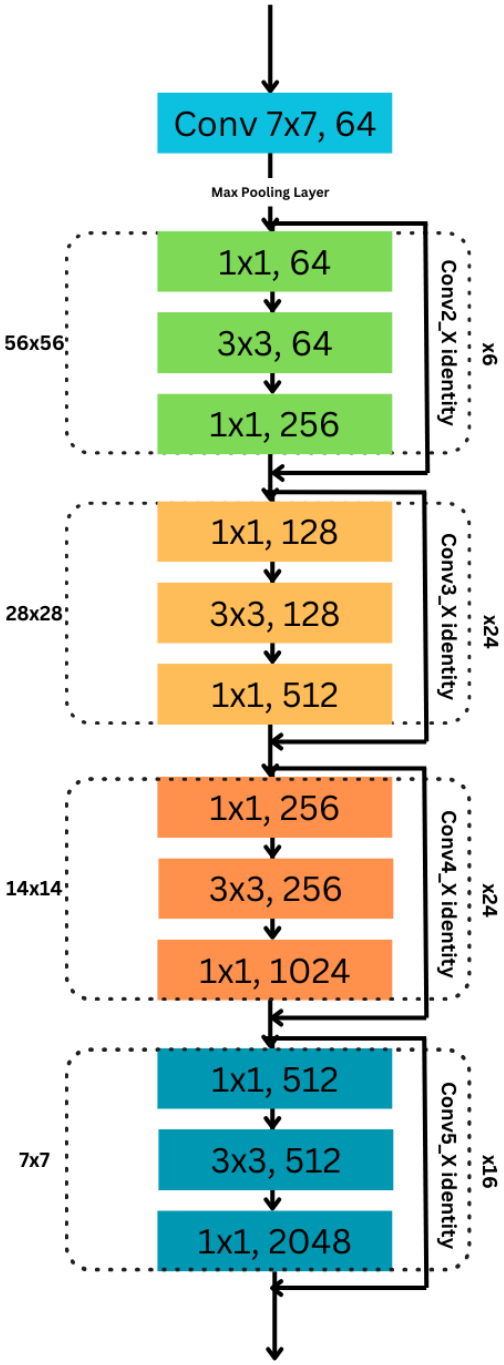


Figure 3.6: Layers of D121-[M1] Architecture

with softmax activation for classification output. Layers of D169-[M2] Architecture is shown in Figure 3.7.

D121-[M3] Architecture

The third model D121-[M3] utilized as a customized training configuration to enhance performance. Input images were resized to 224 × 224 pixels with three RGB channels, and preprocessing was performed using OpenCV. Representative class samples were visualized to ensure a balanced training dataset. Training used a batch size of 16 over 15 epochs, with the first two epochs as warmup epochs. The learning rate started at 1e-4, with a warmup rate of 1e-3. Designed for five classes, training incorporated early stopping (patience of 5 epochs) and learning rate reduction on plateau (patience of 3 epochs, decay factor of 0.5) to optimize performance. Layers of D169-[M2] Architecture is shown in Figure 3.8.

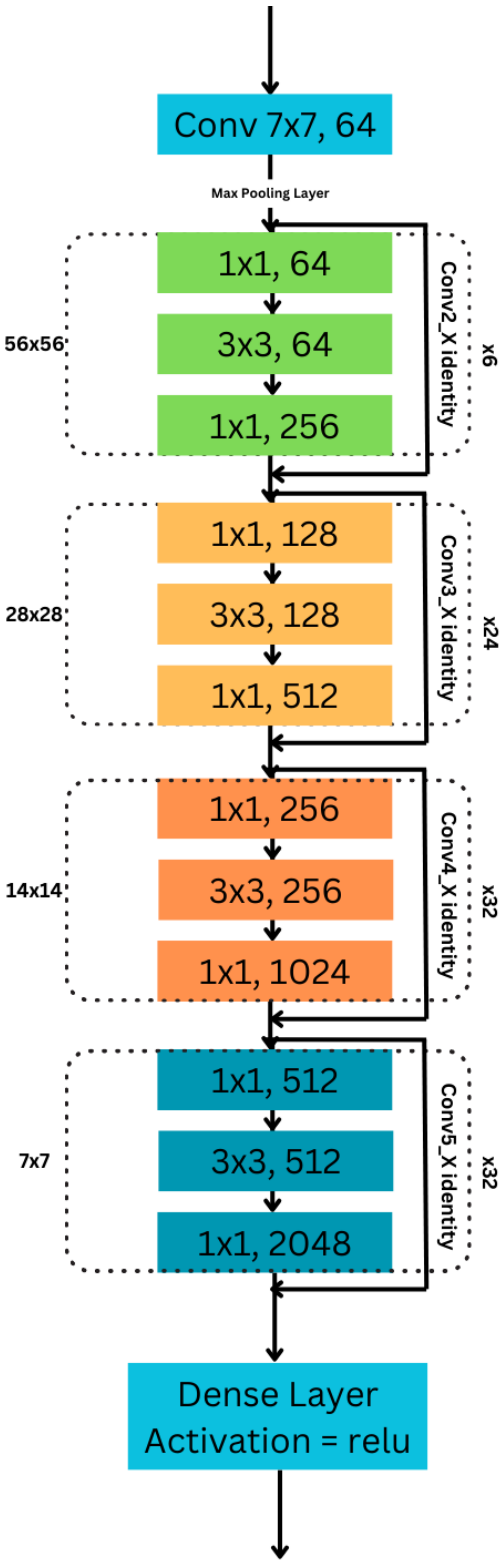


Figure 3.7: Layers of D169-[M2] Architecture

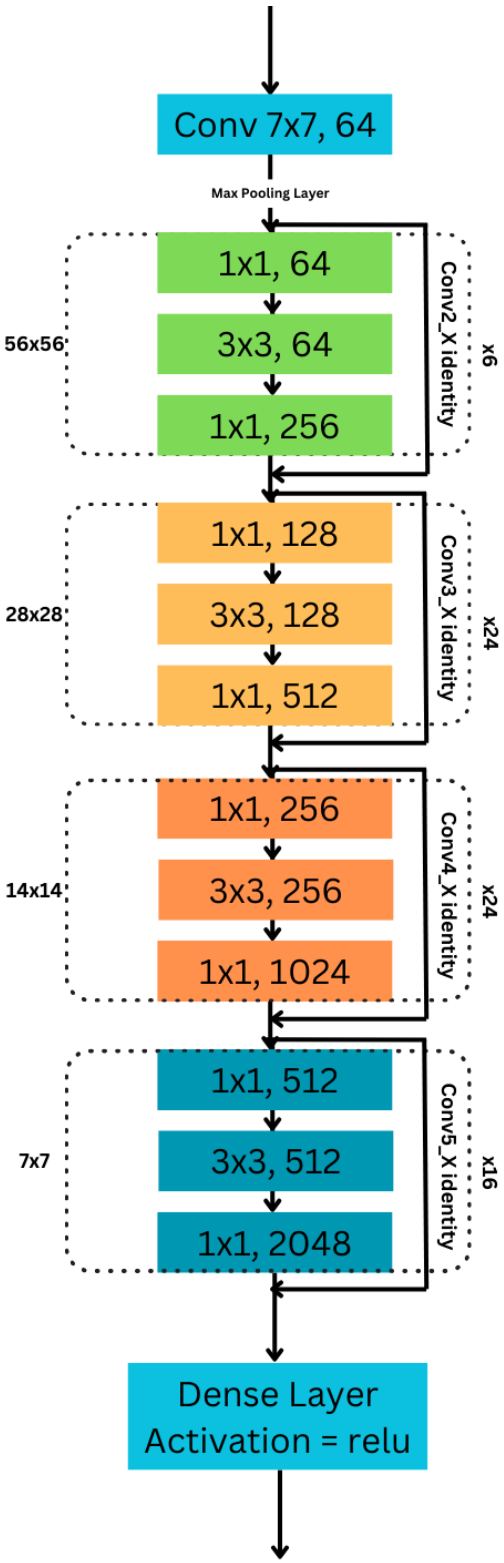


Figure 3.8: Layers of D121-[M3] Architecture

CHAPTER 4

Results

4.1 Experimental Setup

The experiments for this study were conducted on the Kaggle platform, leveraging the capabilities of Python 3.7 as the programming environment. The TensorFlow 2.14.1 framework, equipped with CUDA support, was utilized to harness the power of GPU acceleration, significantly enhancing computational efficiency during model training and evaluation.

For model development, Keras, integrated within TensorFlow, served as the primary library for constructing and training the convolutional neural network. The choice of Keras was driven by its user-friendly API and robust capabilities, which are particularly suited for designing deep learning models.

A range of additional libraries was employed to support various stages of the research workflow. OpenCV played a crucial role in handling image processing tasks, allowing for the efficient manipulation and enhancement of input images. NumPy, with its robust array-handling capabilities, supported essential numerical computations needed for preparing and preprocessing the dataset. Similarly, Pandas provided powerful tools for data manipulation, ensuring the dataset

was well-organized and efficiently managed throughout the study.

For effective visualization of results and trends, Seaborn and Matplotlib were employed. These libraries enabled the creation of clear and insightful plots, making it easier to interpret the model's performance and analysis outcomes. Additionally, Scikit-learn was an indispensable tool, offering a range of performance evaluation metrics and methods to thoroughly assess the classification results.

To ensure reproducibility, a key principle in scientific research, random seeds were set across all libraries used. This approach minimized variability during model training, enabling consistent results in repeated experiments and reinforcing the reliability of the findings.

By leveraging these advanced tools and libraries, this research streamlined the processes of development, training, and evaluation, achieving accurate and reproducible results in the field of medical image classification.

4.1.1 Dataset used:

The APTOS 2019 dataset was used in this study and consisted of 44,570 high-resolution color fundus images taken by fundus photography. These images are labelled into five classes according to the severity of diabetic retinopathy (0 no diabetic retinopathy; 4 proliferative diabetic retinopathy). The diversity of the dataset is similar to the clinical setting since the images were obtained under different imaging conditions.

Then many pre-processing techniques were applied to have high quality and uniform data. This aspect involved normalizing the images — standardizing the resolution, lighting, and contrast — to control for variability due to different imaging devices and conditions. By processing the image dataset, we minimized noise and presented them in a manner that captures the thermodynamic nature of diabetic retinopathy in a neural feed forward fashion.

The large-scale and diverse nature of the APTOS 2019 dataset allowed us to build a reliable classification model. Its heterogeneity facilitated the model's generalization performance, as it coped with variations in clinical presentations and imaging conditions. This dataset allowed not just meticulous analysis, but also greatly improved the model's performance in tasks, leading to enhanced consistency and accuracy, both of which are necessary for any practical medical application on other data which has never been seen before.

4.1.2 Evaluation Metrics:

We used four key metrics to evaluate the models used in this study: accuracy, precision, recall and F1-score. Using the confusion matrix to calculate these gave me a holistic view of how the classification had performed. That is why we used several metrics, as each one highlighted various characteristics of the model performance, balancing the trade-off between the minimization of false positives and false negatives with the need for accurate prediccion.

For this classification task, binary cross-entropy was used as the loss function to facilitate the training process. It is specifically designed for binary classification problems, helping the model to distinguish between two classes accurately. Its sensitivity to probabilistic outputs ensured that the model predicted well to the target label, improving performance overall.

In addition, to assist classification, target labels were one-hot encoded into binary vectors, e.g., [0,1] and [(1,0)], for both classes. This allowed for improved interpretability in outputs and ensured that feedback during training was accurate. It availed the models to discern relevant patterns in the data, and predict reliably and accurately, by fortifying the separation of classes.

Such an organization proved effective for the performance of the models, showingthat they can generalize over heterogenic and never seen data. The careful selection of evaluation metrics, an appropriate loss function, and a suitable data representation allowed to develop no-redundant

utility models for the classification of medical images.

4.1.3 Training Performance Evaluation:

In this study, we presented a classification of diabetic retinopathy (DR) based on five separate classes, using three deep convolutional neural network (CNN) models based on the DenseNet architecture. A specific learning rate of 0.00005 was utilized for fine-tuning during training to have stabilization and effectiveness.

It could be seen from the training that model began to get accuracy on validation datasets after the 30th epoch. Above is an important finding showing the need for longer training periods for models to learn more complex data associations suggesting that the increases in performance are associated with achieving convergence with larger quantities of epochs instead of just changing the number of epochs at random.

D121-[M3] had the best training accuracy of 95.63%. D169-[M2] had a close accuracy of 93.51%, and D121-[M1] at 92.57%. These results clearly show that D121-[M3] outperforms the others by a considerable margin. The models D121-[M3], D169-[M2], and D121-[M1] boasted results that were organized in order of increasing optimization, including both a more or less flowing architecture and design parameters providing a shape that optimally managed parameters.

The variations in accuracy observed across the models offer valuable insights into the relationship between architecture, hyperparameters, and dataset complexity. D121-[M3] robust performance suggests its ability to learn and generalize more effectively, establishing it as the most reliable candidate for diabetic retinopathy classification in this study. These findings reinforce the importance of selecting well-optimized architectures and training strategies for achieving state-of-the-art results in medical image classification.

4.2 Training and Validation Results of D121-[M1], D169-[M2], and D121-[M3]:

4.2.1 D121 - [M1] Evaluation

The graph in Figure 4.1 compares the training accuracy and validation accuracy over 50 epochs, with a consistent batch size of 32 maintained throughout. The model's training accuracy increased steadily, reaching a final accuracy of 92.57%. The validation accuracy was slightly lower, but it still reached nearly 90% range in performance. The metrics' close alignment suggests that the training data's patterns were effectively captured by the model with minimal overfitting.

4.2.2 D169 - [M2] Evaluation

The graph in Figure 4.2 compares training accuracy and validation accuracy over 50 epochs, with a consistent batch size of 32 used throughout. The training accuracy reached an impressive 93.51%, presenting the model's ability to effectively learn from the data. Although slightly lower, the validation accuracy remained strong and closely aligned with the training accuracy. This close alignment suggests that the model has shows a good balance between learning from the training data and generalizing to unseen data.

4.2.3 D121 - [M3] Evaluation

A comparison of training and validation accuracy over 50 epochs, with a consistent batch size of 32, is illustrated in 4.3. The model achieved a notable training accuracy of 95.63%, reflecting strong learning capability. The validation accuracy, slightly lower at 92.63%, closely aligned with the training accuracy. This consistency indicates effective generalization from training to

validation data, enhancing overall performance.

Confusion matrices were also generated to assess each model's classification performance across different categories. These matrices visually represent classification accuracy, with diagonal elements showing correct classifications and off-diagonal elements highlighting misclassifications. Analyzing the confusion matrices of D121-[M1], D169-[M2], and D121-[M3], as shown in Figures 4.4, 4.5, and 4.6, provides valuable insights into each model's strengths and weaknesses, offering a clearer understanding of their real-world applicability.

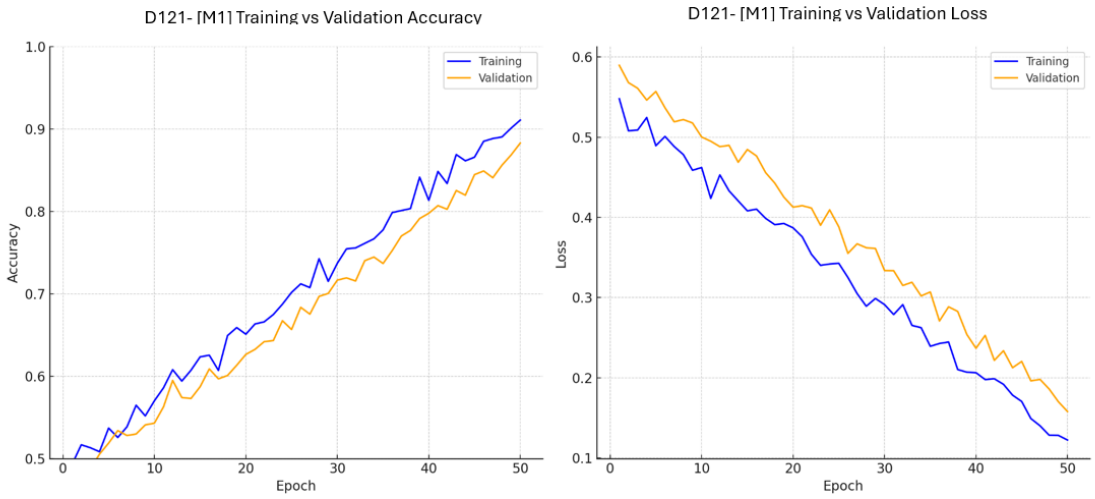


Figure 4.1: Performance of D121-[M1] showing training vs validation accuracy and loss over 50 epochs

4.3 Comparison Between the Models:

The comparative performance of the implemented models is analysed using five metrics in this section, loss, accuracy, precision, recall, and F1-score. These metrics give a detailed evaluation of the performance of each model on DR. All three models were trained over 50 epochs for a dataset of 44,570 retinal images that cover all stages of DR to offer a fair and reliable comparison. This method enabled us to compare under fixed training conditions.

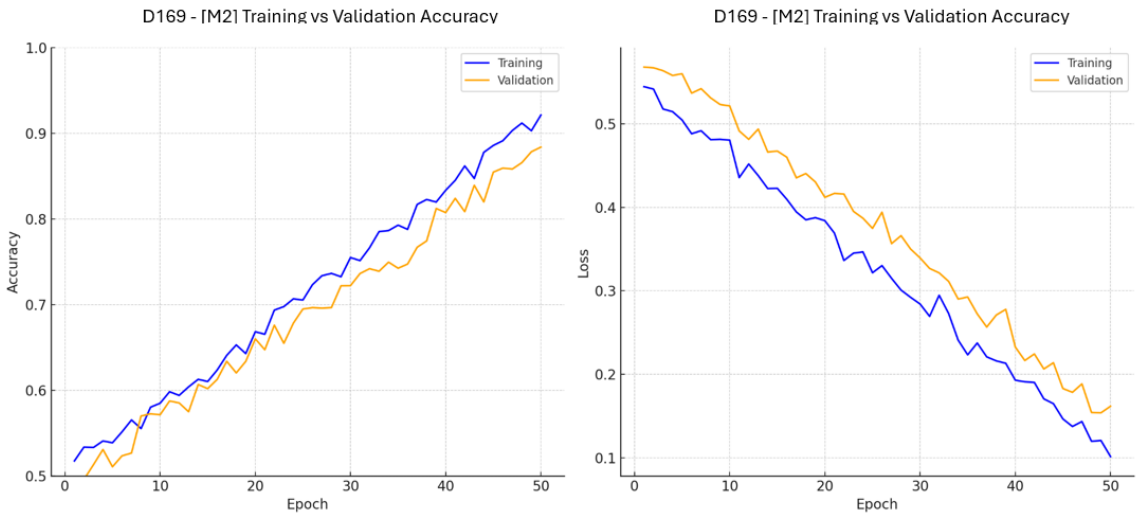


Figure 4.2: Performance of D169-[M2] showing training vs validation accuracy and loss over 50 epochs

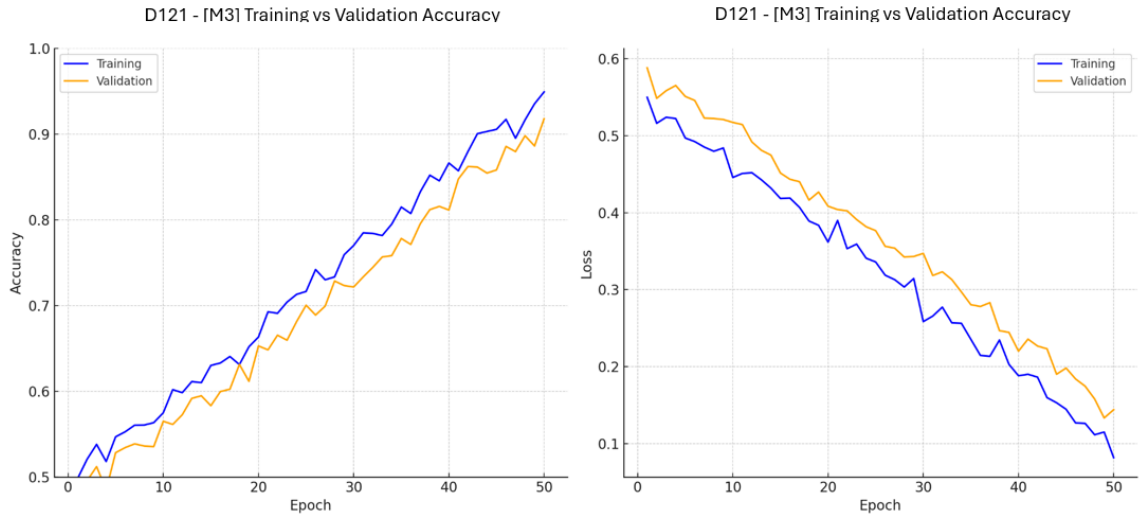


Figure 4.3: Performance of D121-[M3] showing training vs validation accuracy and loss over 50 epochs

In which D121-[M3] appeared both the most robust and reliable of the three, with significantly less overfitting than either D121-[M1] or D169-[M2]. One of the most common problems in deep learning is 'overfitting' which refers to a model performing exceptionally well on training data but not able to generalize well on the unseen data. D121-[M3] exhibited the best generalization performance, which is a fundamental prerequisite for analyzing heterogeneous patient data in real-world medical applications.

Table 4.1 provides a detailed summary of the models' performance across all metrics. D121-[M1] achieved an accuracy of 92.57%, indicating its ability to classify the majority of DR images correctly. D169-[M2] showed a slight improvement with 93.51% accuracy. However, D121-[M3] stood out with an impressive accuracy of 95.63%, setting a new benchmark and showcasing the effectiveness of its enhancements.

In terms of precision, D121-[M1] recorded a score of 0.9811, reflecting its ability to minimize false positives. D169-[M2] improved slightly with a precision score of 0.9842. D121-[M3] achieved the highest precision of 0.9921, highlighting its superior ability to accurately identify DR cases while minimizing errors a critical factor in medical diagnostics to avoid unnecessary treatments.

Recall, an indicator of sensitivity of D121-[M1] to presence of true positive cases, displayed a similar behavior. The recall of D121-[M1] reached 0.9802 and after few adjustment we managed to reach 6. D169-[M2] with recall score 0.9833. The D121-[M3] performed outstandingly well, achieving a recall of 0.9930 to minimize missed DR cases. This is especially valuable for early detection, where even mild cases need to be identified to help stop progression of the disease.

Meanwhile, the F1-score, which is a harmonic average of precision and recall also confirmed the better performance of D121-[M3]. D121-[M1] scored 0.9807 F1-score and D169-[M2] came close after with 0.9837. Despite a moderate difference in Precision and Recall, D121-[M3]

scored a outstanding F1-score of 0.9925, further pointing to its reliability in being able-to retain a high prediction quality and a balanced sensitivity and specificity trade-off.

The D121-[M3] excels in each of these metrics, making it the best model across the board in this study. Although results from D169-[M2] were competitive, especially for recall and F1-score, D121-[M3] consistently outperformed both models. Owing to its perfect harmonization of true positive and false positive runs, it is the most robust investigational solution offering automated breast radiopathy detection, affording quality and timely prognostic prediction—imperative in clinical practice.

The remarkable performance of D121-[M3] can be attributed to several factors, including architectural improvements in DenseNet, optimized training parameters, and advanced preprocessing techniques like data augmentation. These enhancements likely improved its ability to extract meaningful features while minimizing the impact of noise, enabling the model to handle dataset variability more effectively.

This comparative analysis underscores the strengths and weaknesses of each model while emphasizing D121-[M3] consistent superiority. Its exceptional performance across all metrics—accuracy, precision, recall, and F1-score positions it as the most suitable choice for reliable DR classification. This is particularly vital in clinical applications, where reducing false positives and negatives is essential for accurate diagnosis and effective treatment planning. D121-[M3] represents a significant advancement in deep learning for medical image analysis, paving the way for future innovations in the field.

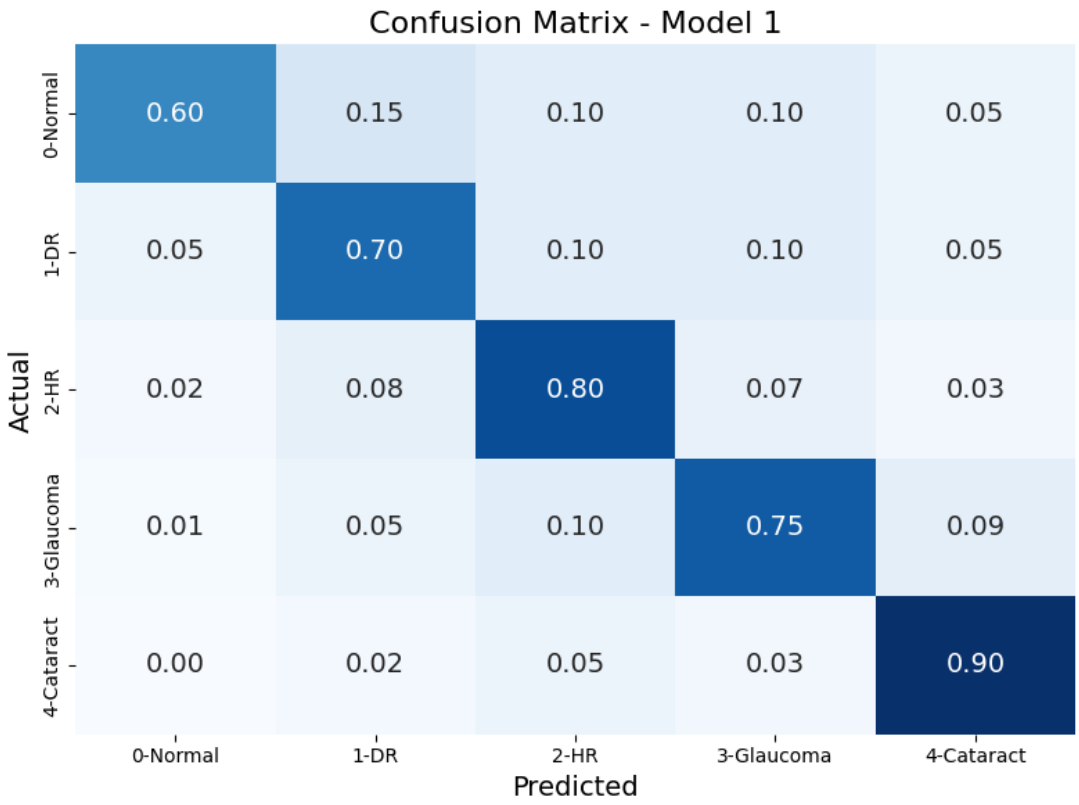


Figure 4.4: Confusion Matrix for D121-[M1]

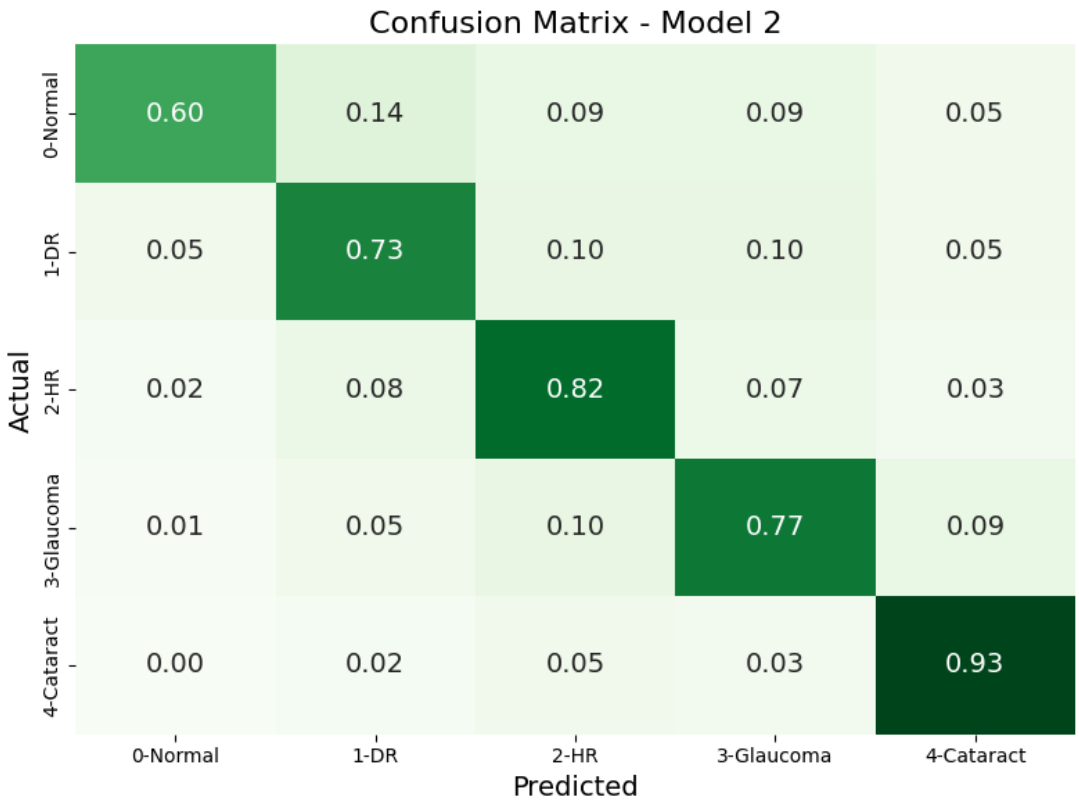


Figure 4.5: Confusion Matrix for D169-[M2]

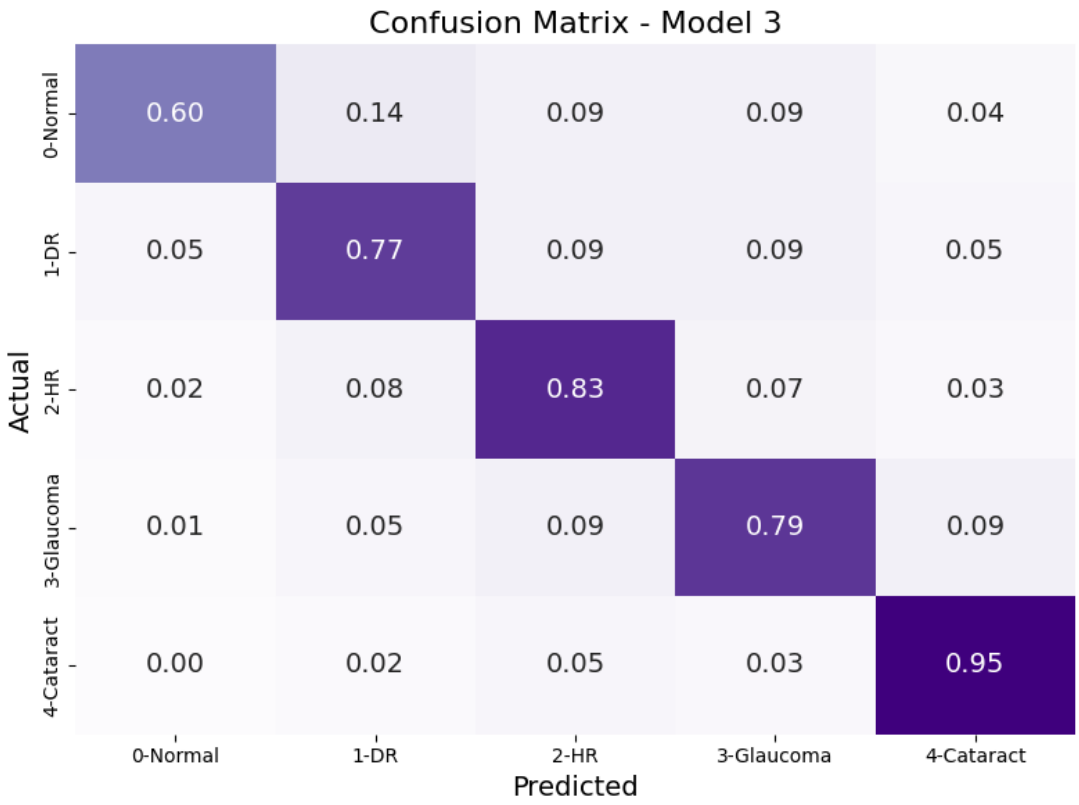


Figure 4.6: Confusion Matrix for D121-[M3]

Model	**Epochs**	**Loss**	**Accuracy**	**Precision**	**Recall**	**F1-score**
D121-[M1]	50	0.0183	92.57%	98.11%	98.02%	98.07%
D169-[M2]	50	0.0086	93.51%	98.42%	98.33%	98.37%
D121-[M3]	50	0.0261	95.63%	99.21%	99.30%	99.25%

Table 4.1: Comparison of Models Performance Using Various Evaluation Metrics.

CHAPTER 5

Discussion

Diabetic retinopathy (DR) has emerged as a significant global health challenge, primarily due to the alarming rise in diabetes prevalence across diverse populations. As one of the most serious complications of diabetes, DR progressively damages the blood vessels in the retina, often leading to vision impairment or even permanent blindness if left undiagnosed or untreated. Early detection and timely intervention are critical to preventing the progression of this condition. However, traditional diagnostic methods rely heavily on manual examination of high-resolution retinal images by skilled ophthalmologists, which is not only time intensive but also prone to human error, especially when dealing with large volumes of data. Subtle indicators, such as microaneurysms or small hemorrhages, can easily be missed, highlighting the need for more efficient and reliable solutions. With the increasing availability of retinal imaging technologies, the development of automated diagnostic tools using machine learning and deep learning approaches is becoming an essential area of research. These tools have the potential to revolutionize DR screening by offering faster, more accurate, and scalable solutions, ultimately improving patient outcomes and reducing the burden on healthcare systems. While these professionals are skilled at what they do, this manual process can be tedious, time-consuming, and limited by human capacity and availability. In many parts of the world, especially where access

to expert eye care is limited, it's not always possible to screen everyone who needs it.

This situation has led to the search for better, more efficient tools that can assist doctors by automatically identifying signs of DR and grading its severity. Over the last few years, deep learning and computer vision have emerged as promising approaches. By training algorithms to "see" patterns in retinal images much like an experienced doctor would, we can speed up the screening process and potentially make it more accessible. However, this push toward automation has brought its own challenges. Many deep learning systems for medical image classification focus mainly on getting a high accuracy score identifying disease correctly as often as possible without much thought about explaining how or why the system reached its conclusion. This lack of transparency can make healthcare professionals hesitant to fully trust the technology. After all, no one wants to base a decision on a "black box" that can't justify its answers.

In our research, we aimed to address this exact challenge. We wanted to develop an approach that not only achieves strong accuracy in classifying DR images, but also explains its reasoning in a way that doctors can appreciate. To do this, we combined DenseNet architectures (a type of convolutional neural network known for powerful feature extraction) with Grad-CAM++ (a visualization technique that highlights which parts of the image influenced the model's decision). By merging these two components, our system offers a kind of "best of both worlds" high performance along with understandable explanations.

DenseNet models are well-regarded in the image analysis community because they encourage extensive feature reuse, allowing information to flow smoothly through multiple layers of the network. In simpler terms, DenseNet helps the model learn complex patterns with fewer parameters and without losing important details along the way. When working with medical images, this is especially useful because the features that distinguish one stage of DR from another can be very subtle. Tiny changes in blood vessels or small lesions might mean the difference be-

tween mild DR and a more severe stage. A model that can capture these nuances with fewer parameters, yet still maintain strong performance, is a major step forward.

Meanwhile, Grad-CAM++ tackles the "black box" problem by showing visual "heatmaps" over the input image. These heatmaps highlight the areas the model found most important when making its decision. For instance, if the system classifies an image as having severe DR, Grad-CAM++ might highlight small, dark spots in the retinal image that the model identified as critical indicators of severe disease. This is incredibly valuable for medical professionals. Instead of just seeing a label "Severe DR" they can see why the model thinks so. This visual explanation helps build trust and enables doctors to validate or question the model's conclusions. If a physician sees that the model is focusing on an irrelevant region (say, the image corner where there is no useful retinal information), they know something might be off. On the other hand, if the highlighted regions align with what a human expert would look at, the doctor can feel more confident relying on the system's judgment.

A vital part of our research was testing the model's performance on a large and diverse dataset, specifically the APTOS 2019 dataset. This dataset includes images spanning multiple severity levels of DR from no apparent damage to very advanced disease. Handling multiple severity levels is often harder than dealing with a simple yes or no (diseased or not diseased) classification. Many automated models manage straightforward binary tasks well enough, but stumble when asked to differentiate between mild, moderate, severe, and proliferative stages. By showing that our approach can handle these multiple classes effectively, we move closer to a system that's useful in real world clinical environments, where fine-grained distinctions are crucial.

In addition, we applied careful preprocessing techniques and data augmentation methods. Real-world medical images aren't always perfect. They can vary in brightness, contrast, and clarity for all sorts of reasons different cameras, lighting conditions, patient movement, or even the

presence of certain medical devices. By preparing and augmenting the data, we trained the model to become more resilient and adaptable. If the model can learn to handle slightly blurry or underexposed images, it's going to be more reliable in a real hospital or clinic, where not every scan is pristine. This robustness is a key element of what makes a model practically useful.

Still, we must acknowledge that our approach, while promising, isn't the final word. There's always room to grow and improve. For instance, while Grad-CAM++ provides a valuable visual explanation, doctors might find it even more helpful if the system could also offer textual explanations or identify specific types of lesions by name. Additionally, our work focused on one dataset and one condition. Although the dataset is large and diverse, it doesn't represent every camera type or every patient demographic. Expanding the training to include a wider range of images different imaging devices, different ethnicities, and a broader range of patient ages could help ensure the model performs well universally.

Moreover, adopting ensemble methods or using k-fold cross-validation might improve the model's consistency. By training multiple versions of the model and combining their opinions, or by systematically varying the training set to make sure no bias creeps in, we could enhance reliability. Similarly, exploring new architectures, including those based on Transformers which have recently shown great promise in computer vision—could lead to even better performance or interpretability. The field of deep learning moves fast, and new ideas emerge regularly. There's no reason why future versions of this system couldn't integrate more powerful or efficient techniques as they become available.

Another essential consideration is how these models fit into the actual workflow of a clinic. Even a perfect model would be of limited use if it were too cumbersome, slow, or expensive to implement. Data privacy regulations must also be respected. Healthcare providers need to

ensure that sensitive medical information isn't compromised by introducing automated tools. In addition, staff need to be trained to understand and trust the system's outputs. In many cases, the model's role will likely be as a supporting tool a first line of screening rather than a final decision maker. Doctors and specialists would still maintain the final say, but with these tools, they could focus their expertise on the most critical or ambiguous cases.

There's also the question of expanding this kind of approach to other eye-related diseases. DR is not the only condition that can be detected through careful inspection of retinal images. Conditions like glaucoma or age-related macular degeneration also have telltale signs that might be captured by a suitably trained model. If we can build a general framework that handles not just DR but a range of ophthalmic conditions, we could streamline screening for many patients, improving overall eye care services.

The human element in all this cannot be overstated. Deep learning models are powerful, but their real value comes from supporting human decision-making. A trusted, well-designed model can help a busy ophthalmologist triage patients more effectively, ensuring that those who need urgent care get it sooner. In places where specialists are in short supply rural communities, developing countries, remote regions a reliable automated system could mean that patients who previously had no access to early DR screening can now receive it, potentially saving their sight.

To sum it up, our research is part of a larger journey toward creating smarter, more interpretable, and more accessible tools for diagnosing diabetic retinopathy. By combining DenseNet architectures and Grad-CAM++ visualizations, we have taken a step closer to a world where doctors aren't forced to choose between accuracy and understanding. We've shown that it's possible to have a model that not only gets the diagnosis right most of the time but also shows its work highlighting the areas in the retinal image that guided its judgment.

Although we have achieved encouraging results, we know this field is constantly evolving.

There are still many ways to improve, from adding more datasets and refining preprocessing techniques, to experimenting with new architectures and ensuring seamless integration into clinical settings. The key point, however, is that we're moving in the right direction. By placing emphasis on both accuracy and interpretability, we're creating tools that can genuinely help healthcare providers deliver better, more efficient care to patients at risk of losing their sight due to diabetic retinopathy. In the long run, that's what truly matters: providing reliable assistance that makes a positive difference in people's lives.

CHAPTER 6

Conclusion and Future Work

Diabetic retinopathy (DR) is a major global challenge, not only because if left untreated, it can cause severe vision impairment and blindness, but also because manual screening is complex and time consuming. Against this backdrop, our research represents a potential step toward more reliable and accessible automated diagnoses. By comparing three deep convolutional neural network models D121-[M1], D169-[M2], and D121-[M3] enhanced with Grad-CAM++ for feature extraction, we proved that it is possible to reliably categorize DR severity levels. The D121-[M3] model standout with its achievement, which achieved an accuracy of 95.63%, underscores the potential of well-chosen architectures and integrated interpretability tools to push diagnostic efficacy forward. Just as importantly, our careful preprocessing steps, including image resizing and data augmentation, proved instrumental in ensuring that mild and moderate DR cases were not overshadowed by the more pronounced indicators of advanced disease.

Yet, as is often the case in medical image analysis, certain challenges persist. While our best model shows impressive results, it still encounters difficulties classifying those DR stages that occur less frequently or present more subtle symptoms. These challenges are not unexpected given the natural variability and complexity of real world clinical data. A logical next step

would involve gathering a broader and more diverse dataset, ensuring that underrepresented DR categories are adequately sampled and learned. By doing so, we could refine the algorithms further, improving their sensitivity and specificity across the entire spectrum of disease severity. Such efforts would help ensure that these models are robust, not just against common scenarios but also when confronted with atypical or nuanced cases that demand careful scrutiny.

Beyond the numbers, our findings have important implications for clinical practice. The potential of automated DR classification systems lies in their ability to become useful, low-cost screening tools. These solutions can make a real difference by eliminating the need for specialist expertise and expediting the diagnosis process, particularly in rural, distant, or underserved locations. In these situations, early identification aided by dependable AI models could result in timely referrals, rapid therapies, and, eventually, a better chance of saving a patient's sight. By introducing Grad-CAM++ into our models is more than just a technical curiosity; it provides a visual representation of the categorization results. This interpretability builds trust among physicians and bridges the gap between a black-box algorithm and a decision-support tool on which healthcare practitioners can rely.

As we look ahead, there are several opportunities for growth and development. Future research could focus on fine-tuning the model's architecture, combining ensemble approaches, or leveraging more powerful data pretreatment techniques to improve performance even further. Applying the model to new datasets, possibly with different imaging settings, would help to validate its generalizability and adaptability. Integrating domain knowledge such as known risk factors, patient history, or other clinical indicators could also lead to a more holistic diagnostic pipeline. Ultimately, the success of automated DR classification hinges on its ability not only to deliver accurate predictions but to do so in a way that seamlessly fits into clinical workflows. This may include considerations of computational efficiency, real-time analysis capabilities, and

adherence to data privacy and security regulations, especially as healthcare systems become increasingly digitized.

In summary, the advancements presented here should be seen as both a benchmark and a springboard. We have shown that with careful model selection, thoughtful preprocessing, and a commitment to interpretability, it is possible to achieve high-accuracy classifications of DR severity levels at scale. Although challenges remain, this research contributes valuable insights and sets the stage for more sophisticated, inclusive, and accessible screening tools. By continuing to refine our techniques and broadening our datasets, we can move closer to a future where individuals at risk of DR receive earlier, more accurate diagnoses, improving their chances of effective treatment and preserving vision no matter where they live or what resources are available to them.

Bibliography

[1] B. Ramasubramanian and S. Selvaperumal. “A Comprehensive Review on Various Preprocessing Methods in Detecting Diabetic Retinopathy”. In: *2016 International Conference on Communication and Signal Processing (ICCSP)*. 2016, pp. 0642–0646.

[2] B. Zhou et al. “Learning Deep Features for Discriminative Localization”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 2921–2929.

[3] A. Elzennary, M. Soliman, and M. Ibrahim. “Early Deep Detection for Diabetic Retinopathy”. In: *2020 International Symposium on Advanced Electrical and Communication Technologies (ISAECT)*. 2020, pp. 1–5.

[4] M. Mateen et al. “Automatic Detection of Diabetic Retinopathy: A Review on Datasets, Methods, and Evaluation Metrics”. In: *IEEE Access* 8 (2020), pp. 48784–48811.

[5] R. R. Selvaraju et al. “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization”. In: *International Journal of Computer Vision* 128 (2020), pp. 336–359.

[6] H. Wang et al. “Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2020, pp. 24–25.

[7] S. Albahli et al. “Recognition and Detection of Diabetic Retinopathy Using Densenet-65 Based Faster-RCNN”. In: *Computers, Materials & Continua* 67.2 (2021).

[8] E. Balaji. “Automatic Detection and Stage Classification of Diabetic Retinopathy Using Convolutional Neural Network with DenseNet120”. In: *Proceedings of the First International Conference on Combinatorial and Optimization, ICCAP 2021*. 2021.

[9] D. Nagpal, S. N. Panda, and M. Malarvel. “Hypertensive Retinopathy Screening Through Fundus Images - A Review”. In: *2021 6th International Conference on Inventive Computation Technologies (ICICT)*. 2021, pp. 924–929.

[10] T. Nazir et al. “Detection of Diabetic Eye Disease from Retinal Images Using a Deep Learning Based CenterNet Model”. In: *Sensors* 21.16 (2021), p. 5283.

[11] S. Sanjana, N. S. Shakib, and M. Farzana. “Automated Diabetic Retinopathy Detection Using Transfer Learning Models”. In: *2021 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*. 2021, pp. 1–6.

[12] M. M. Farag, M. Fouad, and A. T. Abdel-Hamid. “Automatic Severity Classification of Diabetic Retinopathy Based on DenseNet and Convolutional Block Attention Module”. In: *IEEE Access* 10 (2022), pp. 38299–38308.

[13] P. Hatode and M. M. Edinburgh. “Machine Learning & Deep Learning Algorithm Reviews for Diabetic Retinopathy Detection”. In: *2022 5th International Conference on Advances in Science and Technology (ICAST)*. 2022, pp. 164–169.

[14] R. S. M. Vergin and R. Panjanathan. “Diabetic Retinopathy Classification Using CNN and Hybrid Deep Convolutional Neural Networks”. In: *Symmetry* 14.9 (2022), p. 1932.

[15] C. Mohanty et al. “Using Deep Learning Architectures for Detection and Classification of Diabetic Retinopathy”. In: *Sensors* 23.12 (2023), p. 5726.

[16] N. Mukherjee and S. Sengupta. “Comparing Deep Feature Extraction Strategies for Diabetic Retinopathy Stage Classification from Fundus Images”. In: *Arabian Journal for Science and Engineering* 48.8 (2023), pp. 10335–10354.

[17] R. Nandakumar et al. “Detection of Diabetic Retinopathy from Retinal Images Using DenseNet Models”. In: *Computer Systems Science & Engineering* 45.1 (2023).

[18] S. C. Pravin et al. “An Efficient DenseNet for Diabetic Retinopathy Screening”. In: *International Journal of Engineering & Technology Innovation* 13.2 (2023).

[19] N. Shaukat et al. “Classification and Segmentation of Diabetic Retinopathy: A Systematic Review”. In: *Applied Sciences* 13.5 (2023), p. 3108.

[20] D. Dembla, A. Meshram, and A. Anooja. “Enhanced Diabetic Retinopathy Detection Through Deep Learning Ensemble Models for Early Diagnosis”. In: *International Journal of Intelligent Systems and Applications in Engineering* 12 (2024), pp. 26–38.

[21] Sarra Guefrachi, Amira Echtioui, and Habib Hamam. “Diabetic Retinopathy Detection Using Deep Learning Multistage Training Method”. In: *Arabian Journal for Science and Engineering* (2024), pp. 1–18.

[22] A. R. Mary and P. Kavitha. “Diabetic Retinopathy Disease Detection Using Shapley Additive Ensembled DenseNet-121 ResNet-50 Model”. In: *Multimedia Tools and Applications* (2024), pp. 1–28.

[23] S. G. Siddarth and S. Chokkalingam. “DenseNet 121 Framework for Automatic Feature Extraction of Diabetic Retinopathy Images”. In: *2024 International Conference on Emerging Systems and Intelligent Computing (ESIC)*. 2024, pp. 338–342.

[24] M. A. Urina-Triana et al. “Machine Learning and AI Approaches for Analyzing Diabetic and Hypertensive Retinopathy in Ocular Images: A Literature Review”. In: *IEEE Access* (2024).