People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

KASDI MERBAH UNIVERSITY - OUARGLA

Faculty of New Technologies of Information and communication

Department of Electronics and Telecommunications

Master Thesis

Domain: Science and technology

Field: Automatic and systems

Web-Based Retinal Disease Detection Using Deep Learning and Fundus Imaging

Hernouf Abdelmoumen And Larouci Med Abdeldjallil

Before the Jury Members :

Bensid Khaled (MCA)	President	UKM Ouargla
Chlaoua Rachid (MCB)	Supervisor	UKM Ouargla
Korichi Maarouf (MCA)	Examiner	UKM Ouargla

Academic year: 2024/2025

Abstract

Retinal diseases are a leading cause of preventable vision loss, especially in areas lacking specialized eye care. Early diagnosis is essential for effective treatment and preventing permanent damage. This thesis develops an automated deep learning system for detecting and classifying retinal diseases from fundus images.

We utilized two available datasets in this study: APTOS 2019 for diabetic retinopathy severity grading and ODIR 2019 for multi-disease classification, including glaucoma, age-related macular degeneration (AMD), cataract, pathologic myopia, retinal vein occlusion (RVO), and other conditions.

To enhance training efficiency and improve accuracy, we applied several preprocessing techniques, including Contrast Limited Adaptive Histogram Equalization (CLAHE) and Ben Graham's enhancement method. Various deep learning models including ResNet152, DeiT, Swin-V2, and MaxViT were trained and evaluated under different configurations. Among these models, Swin-V2 demonstrated superior performance in terms of both accuracy and generalization capabilities.

We implemented a two-stage diagnostic pipeline: the first model detects the presence of retinal abnormalities, while a second model, activated upon detection of diabetic retinopathy, estimates the severity level. To ensure practical applicability, we developed a web-based platform that integrates these models into a user-friendly diagnostic tool.

The results demonstrate the potential of combining deep learning approaches with fundus imaging to provide scalable and accurate retinal disease screening. This work establishes a solid foundation for future clinical validation and integration into telemedicine platforms.

Key words:

Deep Learning (DL), Artificial Intelligence (AI),Convolutional Neural Networks(CNNs),Data Processing, Vision Transformers (ViTs), Fundus Imaging, Retinal Diseases, Diabetic Retinopathy (DR).

ملخص

تُعد أمراض الشبكية من الأسباب الرئيسية لفقدان البصر القابل للوقاية، لا سيما في المناطق التي تفتقر إلى رعاية متخصصة في طب العيون. يُعد التشخيص المبكر أمرًا بالغ الأهمية لضمان علاج فعّال والحد من حدوث تلف دائم في الرؤية. تهدف هذه الأطروحة إلى تطوير نظام مؤتمت قائم على تقنيات التعلم العميق للكشف وتصنيف أمراض الشبكية باستخدام صور قاع العين.

استخدمنا في هذه الدراسة مجموعتي بيانات متاحتين: APTOS 2019 لتصنيف درجات شدة اعتلال الشبكية السكري(DR)، وODIR 2019 لتصنيف أمراض شبكية متعددة بما في ذلك الزرق (Glaucoma)، والتنكس البقعي المرتبط بالعمر (AMD)، وإعتام عدسة العين (Cataract)، وقصر النظر المرضي (Pathologic Myopia)، وانسداد الوريد الشبكي (RVO)، وحالات أخرى..

لتحسين كفاءة التدريب وزيادة دقة النماذج، تم تطبيق عدة تقنيات للمعالجة المسبقة، من بينها خوارزمية تحسين التباين CLAHE وطريقة تحسين الصور Ben Graham's تم تدريب وتقييم مجموعة من نماذج التعلم العميق، بما في ذلك ResNet152 وDeiT وSwin-V2 وMaxViT، باستخدام إعدادات مختلفة. وقد أظهر نموذج Swin-V2 أداءً متفوقًا من حيث الدقة والقدرة على التعميم.

تم تنفيذ نظام تشخيصي مكوّن من مرحلتين: يقوم النموذج الأول بالكشف عن وجود أي خلل في الشبكية، وفي حال تم اكتشاف اعتلال الشبكية السكري، يتم تفعيل النموذج الثاني لتقدير مستوى شدته. ولضمان قابلية التطبيق العملي، تم تطوير منصة ويب تدمج هذه النماذج في أداة تشخيصية سهلة الاستخدام. تظهر النتائج إمكانيات واعدة للجمع بين تقنيات التعلم العميق وتصوير قاع العين من أجل تقديم نظام فحص دقيق وقابل للتوسع لأمراض الشبكية. وتُشكّل هذه الدراسة أساسًا متينًا للتحقق السريري مستقبلاً وللتكامل مع منصات الطب عن بُعد.

الكلمات المفتاحية :

التعلم العميق (DL)، الذكاء الاصطناعي (AI)، الشبكات العصبية الالتفافية (CNNs)، معالجة البيانات، محولات الرؤية (ViTs)، تصوير قاع العين، أمراض الشبكية، اعتلال الشبكية السكري (DR).

List of Abbreviations

Abbreviation	**Definition**
AI	Artificial Intelligence
AMD	Age-related Macular Degeneration
APTOS	Asia Pacific Tele-Ophthalmology Society
AUC	Area Under the Curve
BRVO	Branch Retinal Vein Occlusion
CAM	Class Activation Map
CLAHE	Contrast Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Network
CRVO	Central Retinal Vein Occlusion
DR	Diabetic Retinopathy
EER	Equal Error Rate
FOV	Field of View
IQA	Image Quality Assessment
ODIR	Ocular Disease Intelligent Recognition
PDF	Portable Document Format
ROC	Receiver Operating Characteristic
RVO	Retinal Vein Occlusion
VIT	Vision Transformer
XAI	Explainable Artificial Intelligence

Dedication

All praise and gratitude are due to Allah, the Most Gracious, the Most Merciful, whose infinite blessings and guidance have been our constant light, illuminating every step of our journey and enabling us to complete this work. Without His grace, none of this would have been possible.

To our beloved parents, whose unconditional love, tireless sacrifices, and unwavering belief in our abilities have been the cornerstone of our success. providing wisdom in times of confusion, comfort in times of despair, and joy in moments of achievement. Their prayers have paved the way for our accomplishments, and their faith in us has fueled our determination to persevere

To our dear families, whose support and encouragement have been invaluable throughout this journey. Their words of reassurance and presence during challenging times have made every difficulty more bearable. They have celebrated our victories and stood by us during struggles, showing us the true meaning of togetherness and love.

To our supervisor, Professor Chlaoua Rachid, who generously offered his guidance and valuable advice, for which we are deeply grateful.

To our respected professors at University Kasdi Merbah Ouargla, whose knowledge and mentorship have greatly enriched our academic journey.

To our dear friends and colleagues, who have been companions and supporters, sharing this journey with all its details.

We dedicate this humble work to all who have had an impact on our journey, with sincere thanks and appreciation.

Acknowledgment

We would like to express our deepest gratitude to **Dr. Rachid Chlawa** and **Dr. Khalid Ben Side**, whose valuable guidance and the proposal of the initial idea were pivotal in shaping the direction of this thesis. Their support, insight, and encouragement have been instrumental throughout the research journey.

Our sincere thanks go to the **Fentic Department Laboratory** for providing a productive environment and the necessary resources to carry out this work. We also gratefully acknowledge the support of the **Algeria-Cuba Ophthalmology Clinic**, whose contribution added great value to the project, and the **Kasdi Merbah University Incubator** for encouraging the practical realization of this research.

Finally, we would like to express our sincere gratitude to **Kasdi Merbah University Ouargla**, its faculty members, and the administrative staff for providing us with the environment, resources, and guidance necessary to complete this thesis.

Table of Contents

List of Tables

Table **Page**

List of Figures

Figure **Page**

INTRODUCTION

The field of ophthalmology has witnessed significant advancements, enabling ophthalmologists to perform precise clinical examinations and observe symptoms more accurately, thanks to modern technological tools and advanced medical devices. Deep learning techniques have emerged as powerful tools in ophthalmology due to their ability to analyze large volumes of data quickly and accurately, contributing to the early detection of diseases and enhancing the quality of patient care.

By automating the diagnostic process, these advanced models provide accurate predictions, thereby accelerating the achievement of better treatment outcomes. In recent years, the application of deep learning in ophthalmology has attracted considerable academic interest, with numerous studies focusing on the detection of ocular abnormalities through the analysis of fundus images

Conventional fundus images (CFI) is a widely utilized and efficient technique for diagnosing various eye diseases. Its key advantages include rapid execution and ease of use, enabling an ophthalmologists to obtain high-quality images of the fundus with conventional fundus cameras, thus supporting routine clinical examinations. The affordability of CFI ensures its widespread availability in various healthcare facilities, including clinics and hospitals in remote regions with limited financial resources. Furthermore, CFI provides detailed imaging of the central retinal areas, aiding in the early detection of retinal diseases. This technique also requires minimal patient preparation, allowing for quick, non-invasive examinations with minimal discomfort. Despite the emergence of advanced imaging technologies, such as ultra-wide field imaging (UWFI), CFI continues to be a vital diagnostic tool, owing to its simplicity and effectiveness in the retinal diseases detection.

In this thesis, we propose an automated system for the detection of eye diseases using fundus images, leveraging advanced deep learning techniques. The system begins with data augmentation phase to augment the training dataset without requiring the collection of new images by applying various data augmentation techniques, including rotation, zoom, flipping, and shifting [1]. To further enhance image quality, particularly in terms of brightness and contrast, we employ Ben Graham's preprocessing (Ben's) along with the Contrast-Limited Adaptive Histogram Equalization (CLAHE) methods [2, 3]. Additionally, we incorporate several state-of-the-art pre-trained models, such as ResNet152 [4], Data-Efficient Transformer for Image Transformation (DeiT) [5], Swin-ViT Transformer [6], and MaxViT [7], which have demonstrated high efficacy in image classification tasks. These models will be fine-tuned to ensure optimal performance for the specific task of eye disease detection. The primary contribution of this research is the development of an intelligent image classification system for predicting eye diseases through the training and evaluation of cutting-edge deep learning models. Furthermore, the system provides accurate visual representations of regions affected by disease in eye images. This contribution is expected to be valuable to both researchers and professionals in the field of ophthalmology.

The remainder of the thesis is organized as follows. Chapter 1 is devoted to the theoretical study, where the fundamental concepts related to the research topic are reviewed. It also discusses the relevant theoretical frameworks, techniques, and algorithms in a general context. Furthermore, the chapter highlights the foundational principles upon which the proposed system is built, aiming to identify existing research gaps and propose potential solutions.

In Chapter 2, the proposed methodology for a retinal diseases detection based on fundus images is explained. With a rationale for using deep learning models and architectures, which can be used on applications of eye diseases detection. In addition to clarifying all the different proposed system steps in more detail.

In Chapter 3, experiments will be conducted to evaluate the efficiency of the proposed system. This chapter presents the evaluation results, compares the performance of different systems, and highlights the effectiveness of our approach. Based on the experimental outcomes and supporting research, we provide detailed interpretations and assessments of the proposed systems.

In Chapter 4 The thesis concludes with a comprehensive summary, highlighting its key contributions and providing final remarks.
Presents the development and deployment of a web-based diagnostic platform that integrates the trained models into a practical, accessible system for clinical use. This chapter details the platform architecture, user interface design, and system performance in real-world scenarios

Chapter 1

BACKGROUND METHODS

1.1 Introduction

Global estimates indicate that more than 250 million people were affected by retinal disorders in 2020, [8, 9]. In Algeria, a national survey on ocular pathologies conducted in 2008 by the Ministry of Health and Hospital Reform (MSPRH),in collaboration with the National Institute of Public Health (INSP) and the World Health Organization (WHO) [10], reported that retinopathy accounts for 2.4% of blinding eye diseases. Early detection of diabetic retinopathy (DR), along with timely medical intervention, can prevent up to 80% vision loss cases [11].

The early detection of retinal diseases remains a significant global challenge in ophthalmology. In this chapter, we will present the theoretical and technical foundations required to understand a system designed for the detection of retinal diseases. We will begin by reviewing the anatomy and characteristics of the normal eye fundus to establish a reference point for identifying pathological changes. Next, we will explore several common ophthalmic conditions, including diabetic retinopathy (DR), age-related macular degeneration (AMD), retinal vein occlusion,Cataract, Myopia, and Glaucoma. Finally, we will introduce the theoretical foundations of deep learning, which serve as the backbone of our detection system.

1.2 Eye Fundus Anatomy

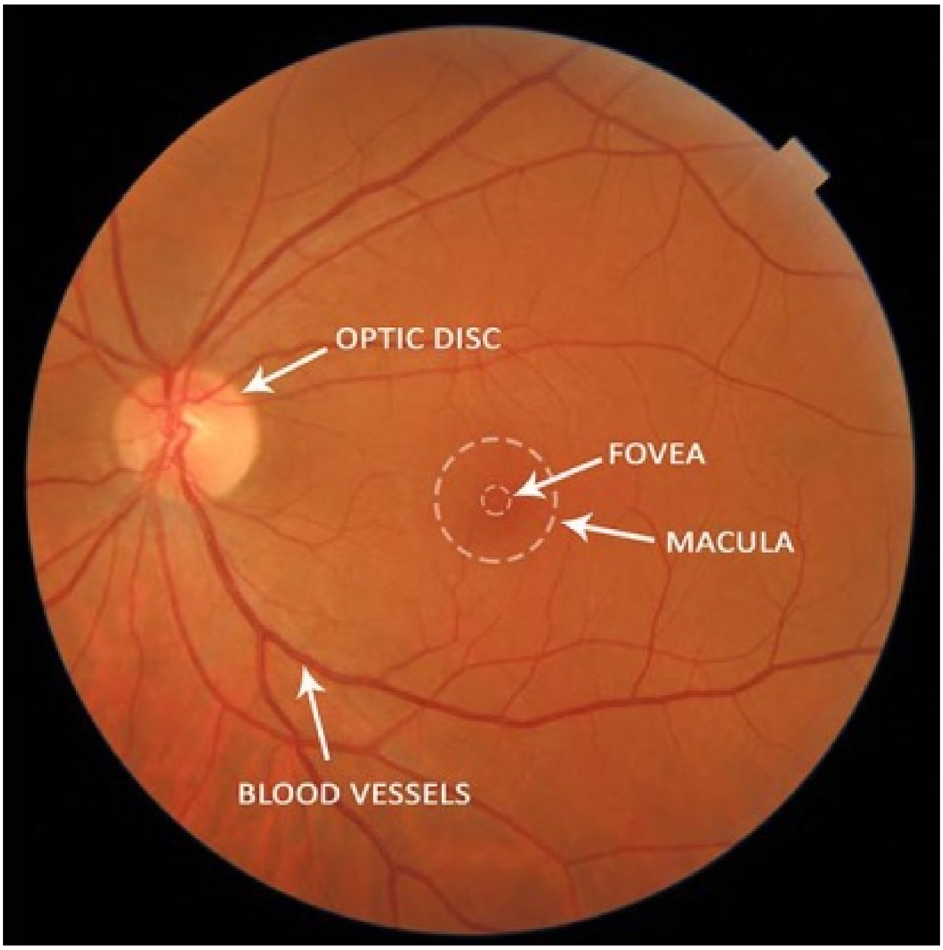


Figure 1.1: Main anatomical structures on a fundus image.

The eye fundus refers to the interior posterior surface of the eye that can be visualized through the pupil using specialized imaging techniques.The normal fundus color ranges from light orange to deep red, depending on the individual's pigmentation, age, and race. The color primarily reflects the choroidal blood supply and the melanin content in the retinal pigment epithelium (RPE). It encompasses several critical structures including the retina, optic disc, macula, fovea, and the network of retinal blood vessels (fig. 1.1). As the only location in the human body where blood vessels can be directly observed non-invasively, the fundus provides a unique window into both ocular health and systemic conditions.

1.2.1 Retinal

The retina is a complex neural tissue that is the innermost layer of the eye responsible for the visual processing that converts light energy from photons into three-dimensional images [12]. Located in the posterior portion of the eyeball, the retina represents the only extension of the brain that can be viewed directly from the outside world,giving clinicians the ability to diagnose many diseases. Development of the retina begins during the fourth week of embryogenesis and continues into the first year of life, making it vulnerable to both genetic and environmental insults. As the most metabolically expensive tissue in the human body, the retina consumes oxygen more rapidly than any other tissue. To support this high metabolic demand, the retinareceives nourishment from a unique dual blood supply that divides it into outer and inner layers for more efficient oxygenation [13].

The background retina appears orange-red due to the visualization of the retinal pigment epithelium (RPE) and the underlying choroidal circulation. It contains the photoreceptors (rods and cones) and the complex neural network that processes visual information. The uniform appearance of the background retina in a healthy eye provides the baseline against which pathological changes are identified. As, the peripheral retina extends from the vascular arcades to the ora serrata (the anterior boundary of the retina). This region contains predominantly rod photoreceptors and is responsible for peripheral and dim-light vision. The peripheral retina is thinner than the central retina and may show slight pigmentary variations in normal eyes. The following fig. 1.2 provides an illustrative of the anatomical structure of the retina, showing its various layers and the specialized cells involved in the visual process.

1.2.2 Optic Disc

The optic disc represents the entry point of the optic nerve into the eye and appears as a round or slightly oval pale area, typically located 3-4 mm nasal to the fovea. With a diameter of approximately 1.5 mm, it serves as the central hub for retinal nerve fibers and blood vessels. Notably, the optic disc lacks photoreceptors, creating the physiological blind spot in our visual field. In a normal fundus, the optic disc has well-defined margins and a small central depression called the cup [14]. The cup-to-disc ratio (CDR) describes the proportion of the cup diameter to the total disc diameter and is a critical parameter in glaucoma assessment. Typically, in a healthy case, cup-to-disc ratio less than 0.4.

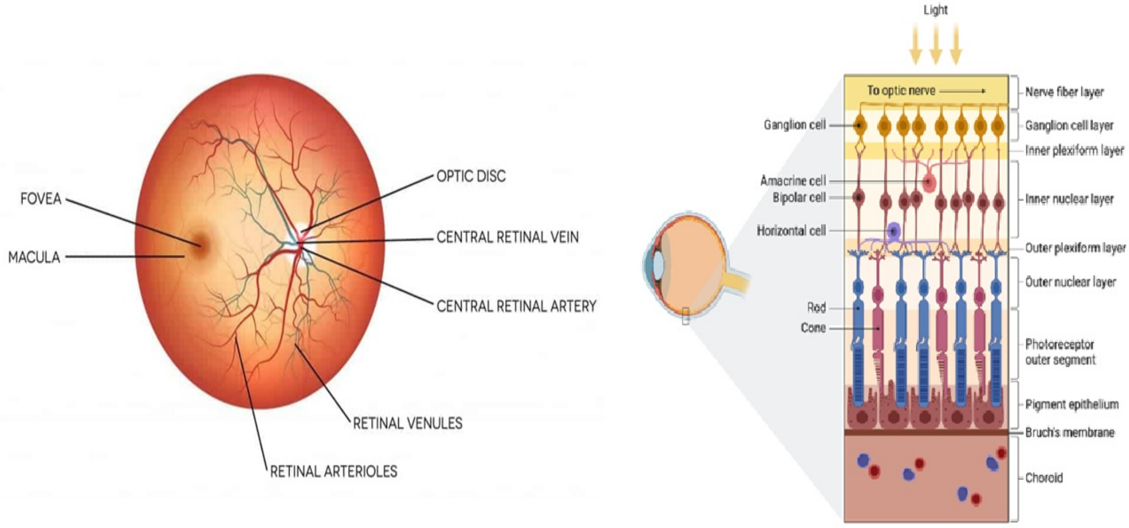


Figure 1.2: Visual process and anatomical structure of the retina.

1.2.3 Macula and Fovea

The macula is a specialized region (approximately 5.5 mm in diameter) of the retina located temporal to the optic disc, responsible for central vision and color perception. It appears slightly darker than the surrounding retina due to the presence of xanthophyll pigments. At the center of the macula lies the fovea a small depression approximately 1.5 mm in diameter that contains the highest concentration of cone photoreceptors, providing the highest visual acuity. The fovea often appears as a small, dark red spot and is avascular, receiving nutrition primarily through the underlying choroidal circulation. A characteristic foveal light reflex is often visible in healthy eyes [14].

1.2.4 Retinal Blood Vessels

The retinal vasculature enters and exits the eye through the optic disc, branching in a characteristic pattern across the retina. Arteries appear lighter red and narrower with a more pronounced light reflex compared to veins, which are darker and wider. The normal ratio of artery to vein width is approximately 2:3. These vessels typically divide dichotomously, becoming progressively smaller toward the periphery, and do not anastomose a characteristic that makes arterial or venous occlusions particularly consequential. Arteries typically cross over veins at arteriovenous crossings, and any deviation from this pattern may indicate abnormal vascular development[15].

1.3 Fundus Image Acquisition Analysis

Fundus photography is an ophthalmic imaging technique used to capture detailed images of the posterior segment of the eye, including the retina, optic disc, macula, and retinal vasculature. This is achieved using a specialized device known as a fundus camera, which typically captures images within a field of view (FOV) ranging from 30° to 50°, centered on the posterior pole [16].

The imaging process generally involves pharmacological pupil dilation using agents such as tropicamide to allow sufficient light to enter the eye. A beam of light from the camera passes through the cornea and lens, reflects off the retina, and is then collected by an objective lens. The reflected light is focused onto a digital sensor that converts the optical signal into a high-resolution digital image [17].

Fundus images are typically high-contrast, color photographs that enable detailed visualization of retinal structures, such as the vascular tree, optic disc margins, and macular region [18]. The resolution of these images generally ranges from 2 to 15 megapixels. Low-intensity flash is used to provide adequate illumination while minimizing patient discomfort. The procedure is conducted in a dimly lit environment, with the patient's head stabilized to minimize motion artifacts. Specialized image analysis software is used post-capture to support diagnostic interpretation [19].

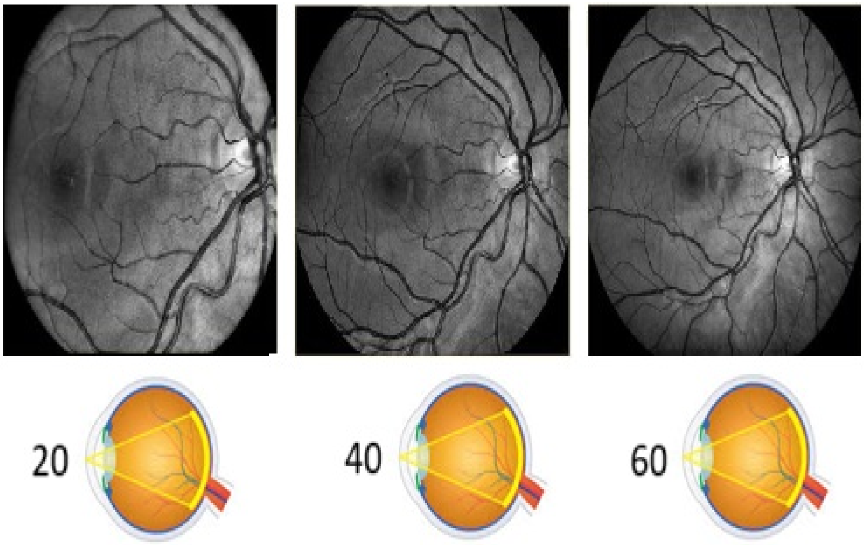


Figure 1.3: Retinal Coverage in Fundus Photography Based on Field of View (FOV).

The field of view in fundus photography determines the extent of the retina visualized in an image (see fig. 1.3). A standard FOV (30°–50°) captures the posterior pole, including the optic disc and macula. Wide-field imaging (up to 100°) allows visualization of the mid-peripheral retina. Ultra-widefield imaging (>100°, up to 200°) enables assessment of the far peripheral retina, revealing subtle pathologies such as retinal tears, holes, or peripheral vascular abnormalities that may be missed in standard imaging [20, 21].

1.4 Retinal Diseases

Retinal diseases encompass a wide range of disorders that affect the retina, the light-sensitive tissue at the back of the eye. These diseases can lead to vision impairment or even complete blindness if not diagnosed and treated early. In this section, we present an overview of the most common retinal pathologies, focusing on their clinical features, causes, and relevance to automated detection systems.

1.4.1 Diabetic Retinopathy (DR)

Diabetic retinopathy (DR) is one of the most common and serious complications of diabetes, affecting the blood vessels in the retina. It begins with micro aneurysms—tiny bulges in the blood vessels of the retina, which can eventually leak fluid or bleed. As the disease progresses, it can lead to more serious complications, where abnormal new blood vessels grow in the retina. These new vessels are fragile and can lead to bleeding in the vitreous, causing vision loss. DR progresses through several stages, these stages can be summarized in the following fig. 1.4 [22, 23]:

1. **Mild Non-Proliferative DR (NPDR):**Characterized by microaneurysms—tiny bulges in retinal blood vessels that may leak fluid.

2. **Moderate NPDR:** Shows larger areas of blood vessel damage and leakage.

3. **Severe NPDR:** Large portions of the retina become affected, with significant blood flow reduction.

4. **Proliferative DR (PDR):** New, fragile blood vessels form (neovascularization), risking vitreous hemorrhage and severe vision impairment.

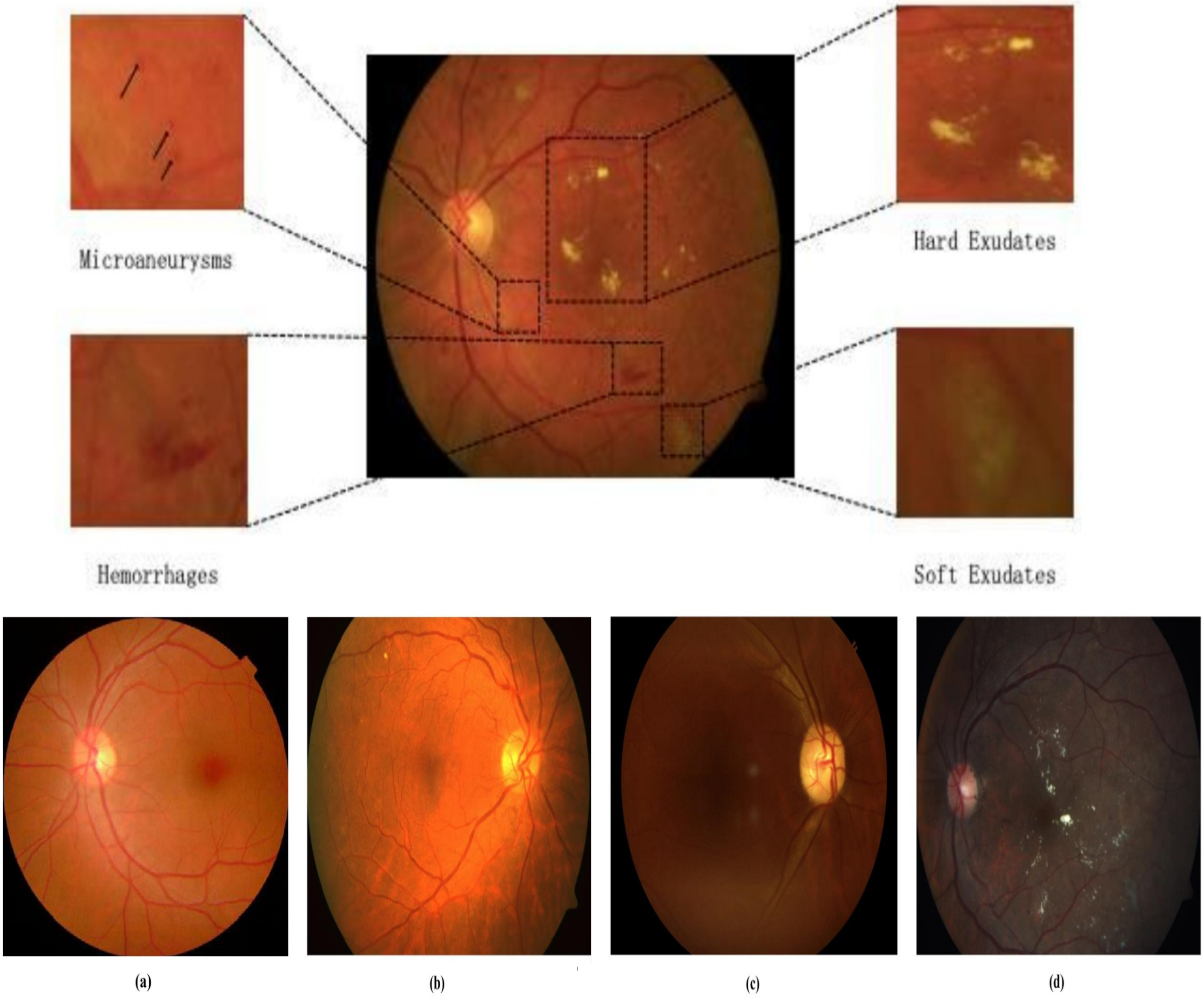


Figure 1.4: Progression of Diabetic Retinopathy DR from Mild to Proliferative Stages; (a) Mild, (b) Moderate, (c) Severe and (d) Proliferative.

1.4.2 Age-related Macular Degeneration (AMD)

AMD is a degenerative disease that affects the macula, the central part of the retina responsible for sharp central vision. This includes reading, driving, and anything that we see when “looking straight ahead”. The rest of the retina is used for peripheral vision. It is most common in older adults and exists in two forms: Dry (atrophic) and Wet (neovascular) [24, 25] such as in fig. 1.5:

- **Dry AMD:** (85-90% of cases) involves the thinning of macular tissues and the accumulation of drusen.
- **Wet AMD:** is characterized by abnormal blood vessel growth under the retina,

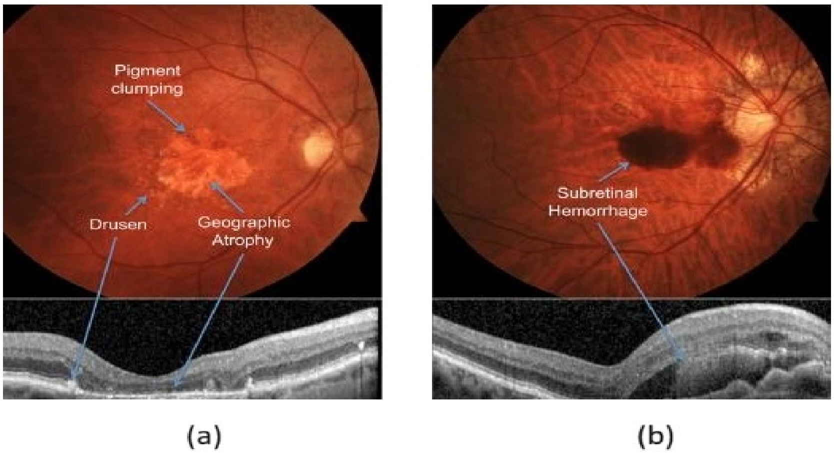


Figure 1.5: Age-related Macular Degeneration (AMD) disease, (a) Dry AMD and (b) Wet AMD.

which can leak fluid and blood, leading to rapid vision loss.

1.4.3 Glaucoma

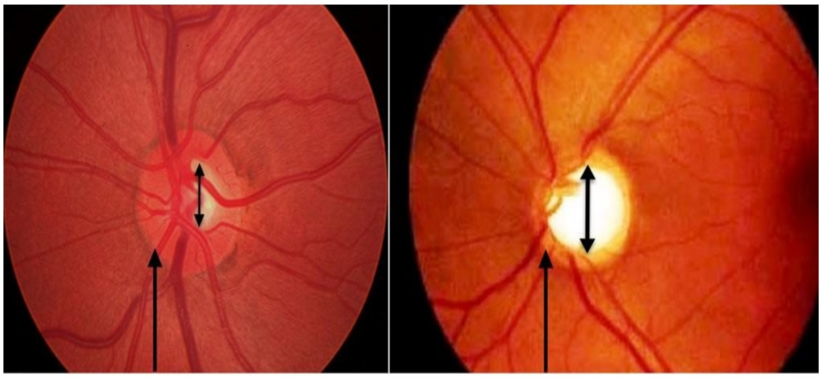


Figure 1.6: Difference Between a Healthy Retina (Left) and a Glaucoma-Affected Retina (Right).

Glaucoma is a group of diseases that damage the optic nerve (fig. 1.6), often due to increased intraocular pressure (IOP). It is a major cause of irreversible blindness. In early stages, glaucoma typically presents no symptoms, which makes regular screening essential. Over time, patients may experience peripheral vision loss that progresses to tunnel vision. Fundus images may show cupping of the optic disc and thinning of the retinal nerve fiber layer [22, 26]. The following Table. 1.1 summarizes the types of glaucoma, their causes, and key clinical features:

Table 1.1: Types of Glaucoma, Causes, and Features.

Type	**Cause / Mechanism**	**Features**
Primary Open-Angle Glaucoma	Gradual dysfunction in fluid drainage with an open angle	Slow progression, no early symptoms, gradual peripheral vision loss
Angle-Closure Glaucoma	Sudden blockage of the drainage angle	Severe eye pain, redness, nausea, blurred vision a medical emergency
Congenital Glaucoma	Congenital defect in eye drainage system in infants	Enlarged eye, excessive tearing, light sensitivity — usually detected early
Secondary Glaucoma	Resulting from other conditions (e.g., injury, inflammation, steroids)	Varies by cause; may resemble open- or angle-closure types

1.4.4 Retinal Vein Occlusion (RVO)

Retinal vein occlusion is the blockage of the central or branch retinal veins, resulting in a backup of blood and fluid. This can cause hemorrhages, macular edema, and ischemia. RVO is often associated with systemic conditions like hypertension, diabetes, and arteriosclerosis. Depending on the location of the blockage, RVO is classified as Central Retinal Vein Occlusion (CRVO) and Branch Retinal Vein Occlusion (BRVO) as presented in fig. 1.7 [23, 27].

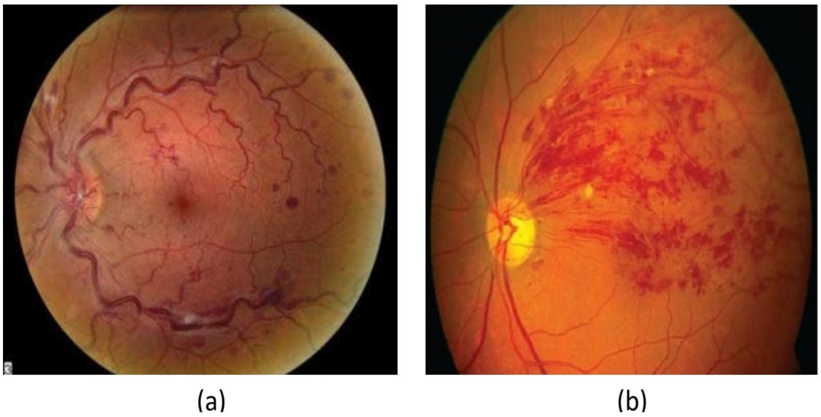


Figure 1.7: Retinal Vein Occlusion (RVO), (a) Central Retinal Vein Occlusion (CRVO) and (b) Branch Retinal Vein Occlusion (BRVO).

1.4.5 Cataract

Cataracts involve the clouding of the eye's natural lens, leading to progressively blurred vision (fig. 1.8). They're primarily age-related but can also result from injury, certain medications, radiation exposure, or congenital factors. Symptoms include blurred vision, increased glare sensitivity, fading colors, and decreased night vision [22]. Main types include:

- **Nuclear cataracts:** Forming in the central lens nucleus, often causing initial myopia before significant visual impairment.
- **Cortical cataracts:** Developing in the lens cortex, creating wedge-shaped opacities that extend from the periphery.

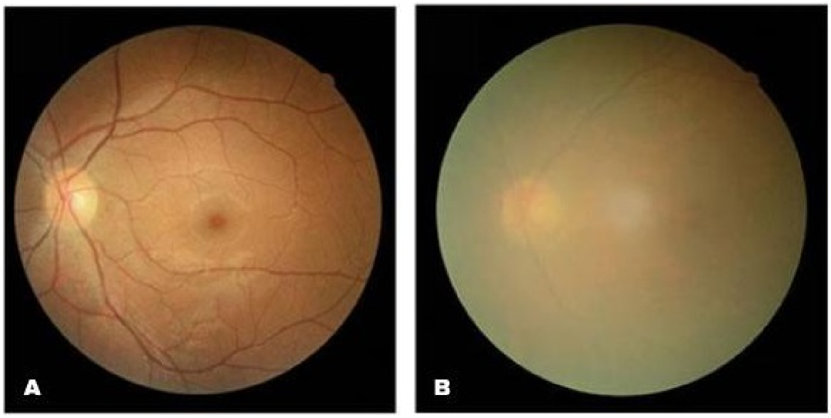


Figure 1.8: (a) Helthy Retina and (b) Retina with Cataract.

1.4.6 Pathologic Myopia

Pathologic Myopia, also known as degenerative myopia, is a severe form of nearsightedness characterized not only by excessive axial elongation of the eye (typically >26.5 mm or a refractive error worse than -6.00 diopters) but also by progressive degenerative changes in the retina, choroid, and sclera. These changes can lead to irreversible vision loss. Due to its structural and progressive nature, it is considered a retinal disease rather than merely a refractive error. Common complications include myopic maculopathy, posterior staphyloma, chorioretinal atrophy, lacquer cracks, choroidal neovascularization (CNV), and retinal detachment [28, 29], see fig. 1.9.

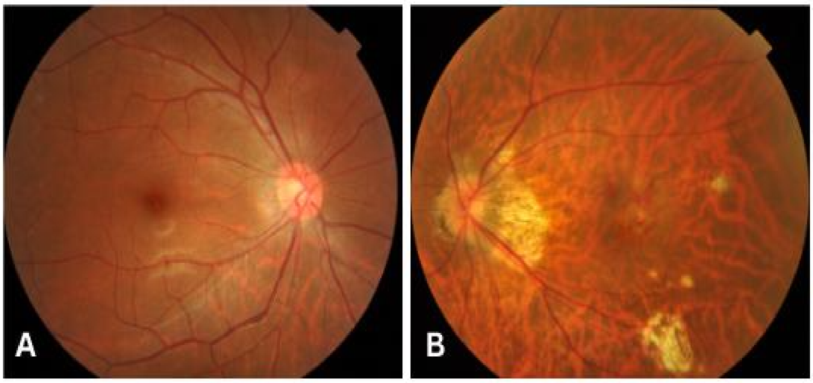


Figure 1.9: (a) Helthy Retina and (b) Retina with Pathologic Myopia.

1.5 Deep Learning Architectures

The advent of artificial intelligence, especially deep learning, has provided promising methods for automated, accurate and scalable detection of retinal diseases. In this section, we try to understand the architectures of deep learning in general, CNN and VIT in particular.

1.5.1 CNN Architecture

Convolutional Neural Networks (CNNs) are among the most prominent deep learning models used in image processing and visual pattern recognition. These networks rely on the concept of using convolutional layers to extract spatial features from images, where small filters are applied to local regions of the image to detect characteristics such as edges or corners in [30].

A typical CNN architecture begins by stacking several **convolutional layers**. After each convolutional layer, nonlinear activation functions such as ReLU are applied to introduce non-linearity into the model. These are often followed by **Pooling layers**, such as Max Pooling or Average Pooling, which serve to reduce the spatial dimensions and the number of computational parameters. This pattern—several convolutional layers with ReLU activations followed by a pooling layer—tends to repeat. As the image moves deeper into the network, its spatial dimensions decrease, but the number of feature maps generally increases due to the convolutional layers. In the final stages, the data is flattened into a vector form using a flatten layer, then passed through **Fully connected layers** to make the final decision (see fig. 1.10). CNNs are widely used in applications

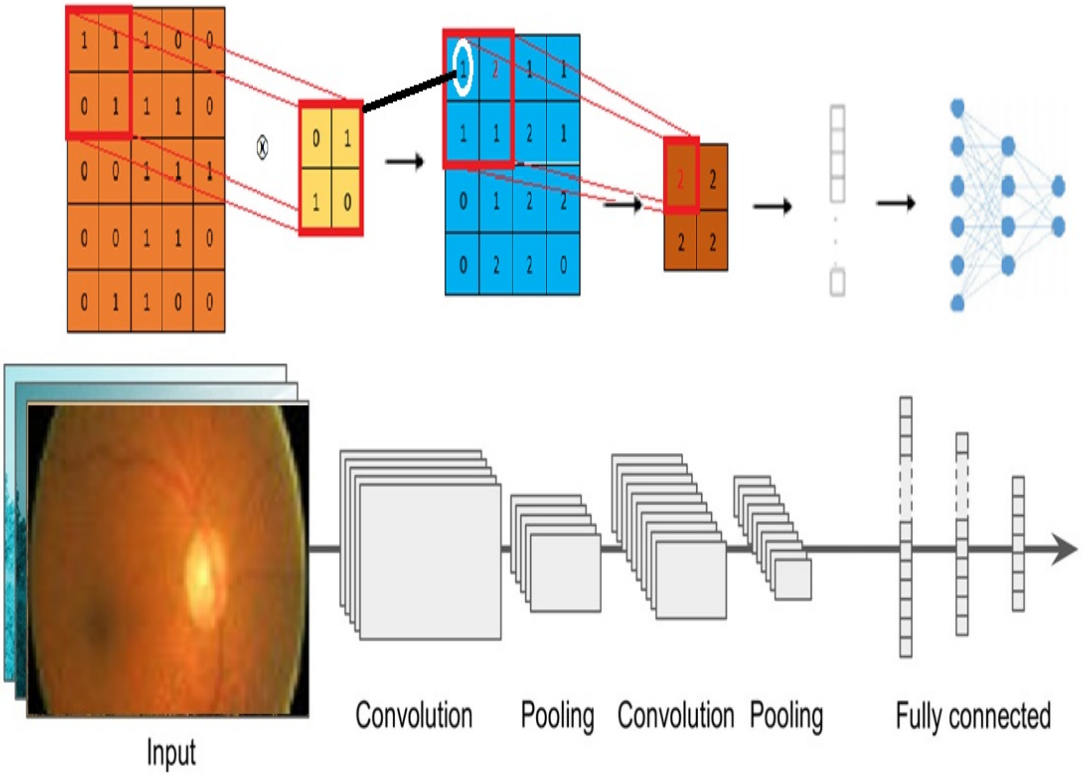


Figure 1.10: Architecture of a Convolutional Neural Network (CNN) for Image Classification, Featuring Convolutional, Pooling, and Fully Connected Layers.

such as face recognition, medical image diagnosis, and autonomous driving systems, due to their high capacity for generalization and pattern recognition in visual data.

1.5.2 VIT Architecture

Transformers are a neural network architecture introduced by Vaswani et al.in [31]. This model relies entirely on attention mechanisms to process sequential data, without the use of recurrence or convolution. The core innovation, self-attention, allows the model to compute contextual relationships between all elements of a sequence simultaneously, enabling greater parallelism and more effective modeling of long-range dependencies. Since then, this architecture has become foundational to a wide range of natural language processing and computer vision tasks, thanks to its scalability and performance efficiency. A key contribution of this architecture is self-attention, which computes interactions among all positions in the sequence at once, regardless of their relative distance.

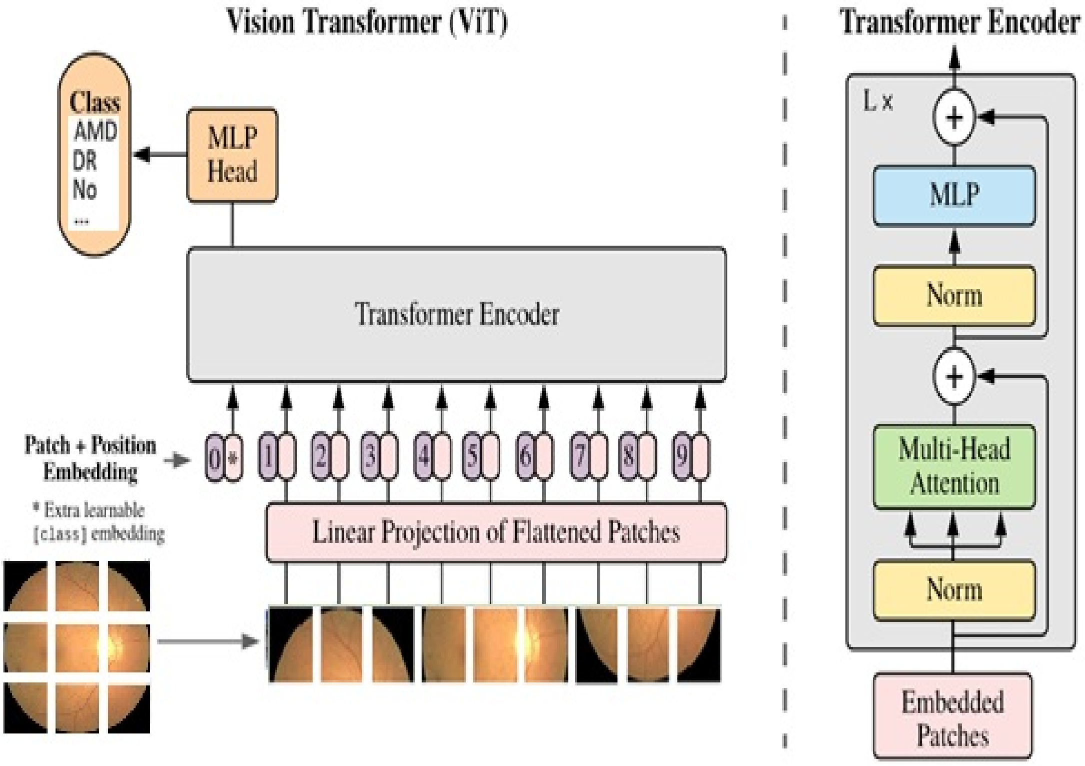


Figure 1.11: Architecture of the Vision Transformer (ViT): Patch Embedding, Positional Encoding, and Transformer-Based Processing for Image Classification.

The Vision Transformer (ViT) is a recent model in the field of computer vision that builds upon the transformer architecture, which has proven highly effective in natural language processing. ViT performs image classification by dividing the image into fixed-size patches rather than processing it as a grid structure like in convolutional neural networks (CNNs). Each patch is flattened into a vector and augmented with a positional embedding to retain spatial ordering, along with a special classification token [CLS] that is later used to derive the final image representation. This sequence of tokens is passed through multiple transformer encoder layers based on multi-head self-attention and feed-forward networks, enabling the model to capture global contextual relationships across image regions.
Finally, the output corresponding to the [CLS] token is used for classification [32]. The fig. 1.11 illustrates the structure and functioning of the Vision Transformer (ViT). ViT is notable for its strong ability to model global relationships within images and has achieved competitive or even superior performance compared to traditional CNN models, especially when trained on large-scale datasets.

1.6 Conclusion

In this chapter, we laid the foundational understanding of eye fundus anatomy, imaging acquisition particularly fundus photography and common retinal diseases. As we discussed the architecture of deep learning models, we highlighted how they work, especially Convolutional Neural Networks and Vision Transformers. By leveraging both local and global features in fundus images, these architectures can detect multiple pathologies with a high degree of precision. This sets the stage for the design and implementation of our proposed AI-based diagnostic system, which we detail in the subsequent chapters.

Chapter 2

PROPOSED WORKFLOW

2.1 Introduction

This chapter outlines the proposed methodology for the automated detection of retinal diseases using fundus images from the APTOS 2019 dataset. With the growing prevalence of diabetic retinopathy and other retinal conditions, early and accurate diagnosis through automated systems has become increasingly vital. Leveraging deep learning techniques, this system aims to enhance diagnostic accuracy and efficiency.

The chapter begins by describing the data preparation process, including data augmentation and preprocessing steps designed to improve image quality and variability. Following this, the architecture and implementation of the deep learning models used in the system are detailed. Each component of the proposed system is carefully designed to contribute to the overall performance and reliability of the system.

2.2 Proposed System Design

In fig. 2.1, we show the block diagram of proposed system for DR disease detection, which has four main parts as database, data augmentation, data preprocessing and deep learning models. In this work, we based our study on the APTOS 2019 dataset. Mostly, medical databases are limited and constrained due to patient privacy concerns. To address these challenges, we propose the use of data augmentation techniques [1], such as rotation, zoom, flipping, and shifting. These techniques enhance the training capabilities of deep learning models. Preprocessing methods [2, 3] are applied to enhance extracting features. Finally, the proposed system can diagnose and classify DR

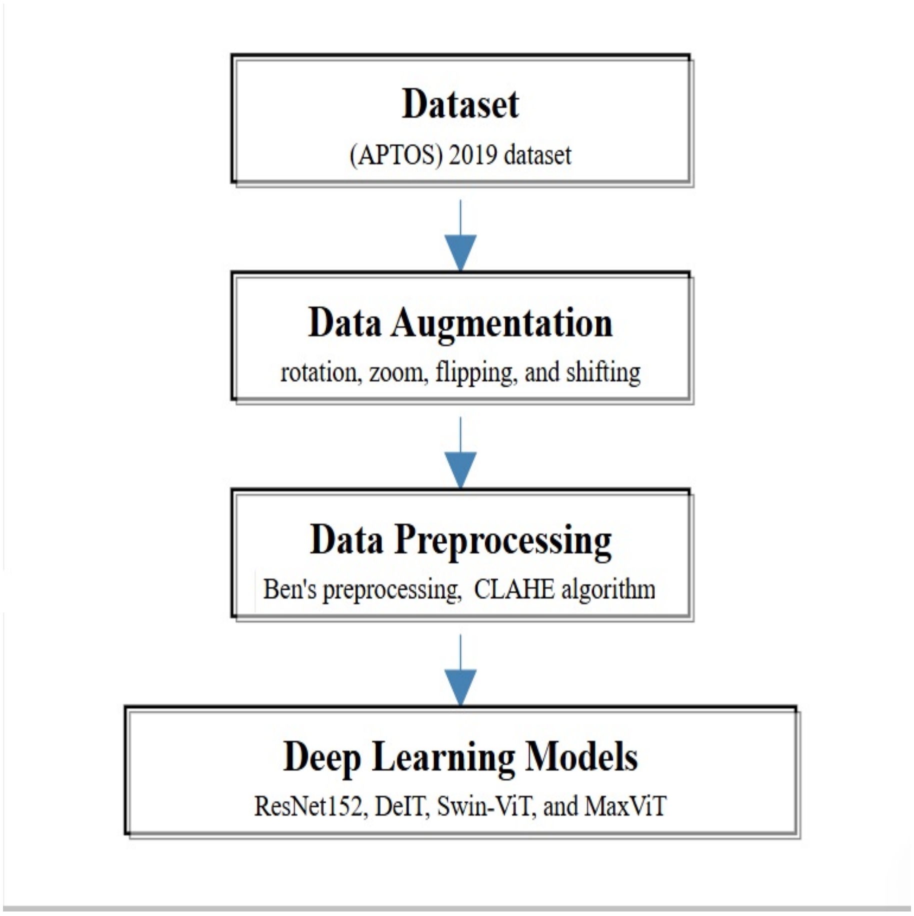

Dataset
(APTOS) 2019 dataset
Data Augmentation
rotation, zoom, flipping, and shifting
Data Preprocessing
Ben's preprocessing, CLAHE algorithm
Deep Learning Models
ResNet152, DeIT, Swin-ViT, and MaxViT

Figure 2.1: Proposed workflow for deep learning-driven retinal disease detection.

disease using deep learning models such as ResNet152, DeIt, Swin-ViT, and MaxViT [4–7].

After a careful evaluation of a set of proposed deep learning models based on APTOS 2019 dataset, the best model that achieve superior performance measures and effective ability in detecting diabetic retinopathy will be determined. Based on this success, we will try to take advantage of this effective model to expand our diagnostic capabilities beyond diabetic retinopathy and generalize the classification to the Odir 2019 dataset, for the purpose of developing a system for detecting and classifying the rest of the diseases also associated with the retina including glaucoma, age-related macular degeneration, and other retinal diseases.

2.3 Datasets Description

Our study utilizes images data from two data bases, the reputable Asia Pacific Tele-Ophthalmology Society (APTOS) 2019 dataset uses for DR disease classification. This dataset was chosen for its broad recognition and the diversity of categories , which support comprehensive analysis and robust comparison of our proposed method against existing models. The APTOS 2019 Kaggle benchmark dataset [33], part of the blindness detection challenge, contains 3,662 retinal fundus images captured under varying imaging conditions by the Aravind Eye Hospital in India. The dataset categorizes images into five levels of diabetic retinopathy severity, ranging from 0 to 4, as illustrated in fig. 2.2. For our analysis, we allocate 80% of the dataset for training and 20% for validation. Additionally, an independent set of 1,928 samples is used for testing.

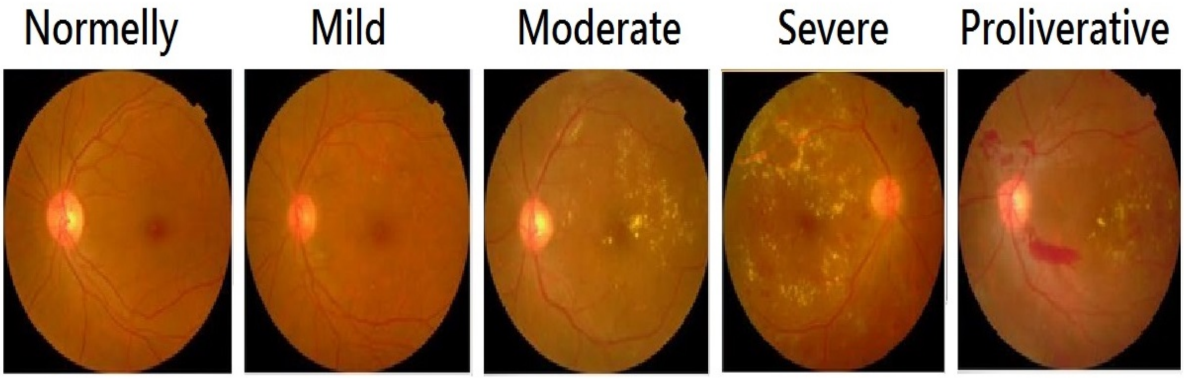


Figure 2.2: Fundus image samples of retinal diseases levels from the APTOS 2019 Dataset.

For the detection of other retinal diseases, our proposed system utilizes the Ocular Disease Intelligent Recognition ODIR 2019 database [34]. The ODIR 2019 dataset of fundus images represents a specialized dataset used for diagnosing eye-related diseases based on the analysis of retinal images. The dataset includes eight different classifications of eye diseases, such as Cataract, Glaucoma, AMD, Myopia, and other retina-related conditions.

The following fig. 2.3 shows an example of fundus images included in the dataset. Out of a total of 4,000 fundus images, 2,820 were allocated for model training. These were divided into two sets: a training set comprising 80% of the images, and a validation set representing the remaining 20% of the training images.

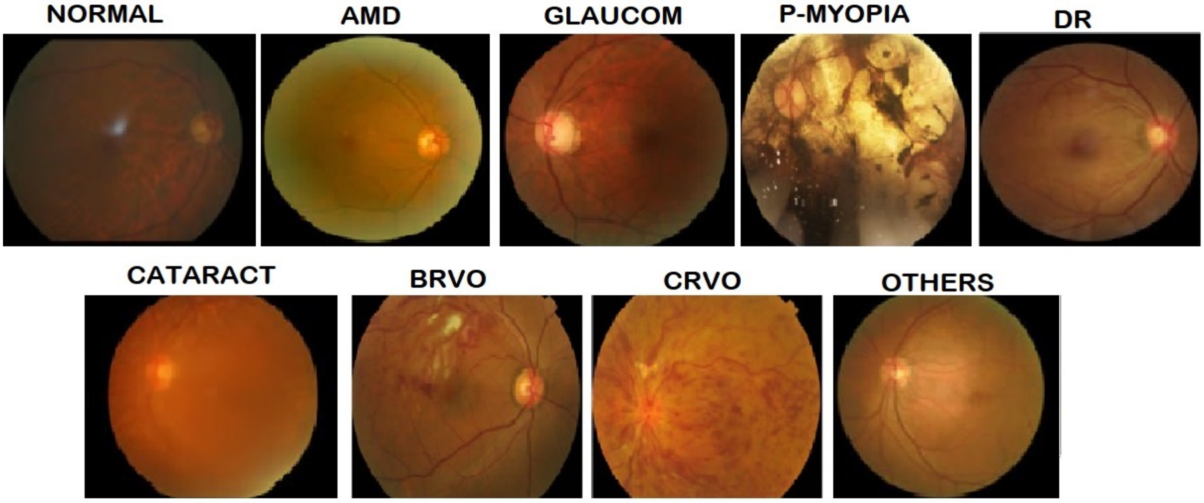


Figure 2.3: Fundus image samples of retinal diseases from the ODIR 2019 Dataset.

2.4 Data Augmentation Strategies

In medical field, the dataset is often limited due to various challenges. To overcome this limitation, we employed a data augmentation techniques. These techniques involve generating variations of the original images while preserving their features. Among these techniques, The following fig. 2.4 shows the used techniques.

2.4.1 Rotation

Rotation technique involves rotating the original images by various angles within a specific range from 0° to 360°, to increase the training data without compromising the image features. In rotation of an image, the original image contains a point (x, y), rotation by an angle θ can be mathematically represented by the following equation [35]:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{2.1}$$

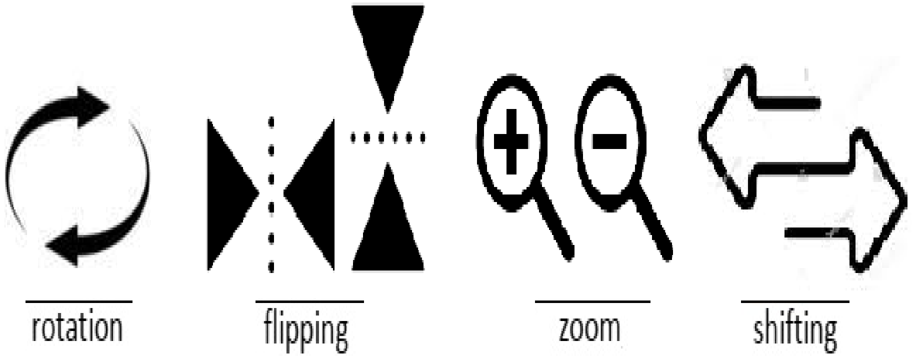


Figure 2.4: The used techniques for dataset augmentation.

Here, (x, y) are the original coordinates, (x', y') are the new coordinates, and θ is the rotation angle.

2.4.2 Flipping

Horizontal and vertical flipping gy transposing the image, by swapping the point (x, y), we can achieve both horizontal and vertical flipping. This straightforward operation effectively doubles the dataset, providing mirrored versions of each image [35]. The following transformation matrixes are used for horizontal and vertical flipping, respectively:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{2.2}$$

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{2.3}$$

2.4.3 Zooming

Zooming in/out is a commonly used data augmentation technique. The image dimensions can be modified using the λ factor, greater than 1 for zooming in, and from 0 to 1 for zooming out [35]. Let I is the image with dimensions $H \times W$, where H is the height and W is the width. The new dimensions of the zoomed in image will be:

$$\begin{aligned} H' &= \lambda \times H, \\ W' &= \lambda \times W. \end{aligned} \tag{2.4}$$

In case of zoom out, the new dimensions will be:

$$\begin{aligned} H' &= \tfrac{1}{\lambda} \times H, \\ W' &= \tfrac{1}{\lambda} \times W. \end{aligned} \tag{2.5}$$

H' is the new height and W' is the new width of the image after zooming.

2.4.4 Shifting

Shifting is a type of data augmentation technique applied to images to increase diversity and improve the model's generalization ability. Horizontal involves shifting left/right or right/left , while vertical involves shifting top/down or down/top [35]. Let I is the image with dimensions $H \times W$, where H is the height and W is the width. Vertical translation (T_v) and horizontal translation (T_h), are used a transformation function the following transformation equations:

$$\begin{aligned} T_v &= (x, y + \Delta_y)\,, \\ T_h &= (x + \Delta_x, y)\,. \end{aligned} \tag{2.6}$$

Where, Δ_x represents the horizontal shift and Δ_y represents the vertical shift.

2.5 Data Preprocessing

Preprocessing of data improves the images quality, making it easier to extract subtle patterns. Among various techniques, our study uses the both Ben's and CLAHE methods, successfully. Ben's preprocessing technique is a complementary approach that focuses on enhancing the visibility of fine retinal structures in fundus images [2]. This method applies specialized filtering and intensity normalization to reduce noise while preserving important diagnostic features. Particularly effective for addressing variations in image quality commonly found in clinical settings, Ben's preprocessing improves the detection of subtle retinal changes that may indicate early-stage pathologies. For the seconde technic which is Contrast-limited adaptive histogram equalization (CLAHE). This algorithm designed to improve the illumination distribution in an image while mitigating the effects of contrast enhancement [4]. It is based on histogram equalization techniques, but with improvements to prevent excessive contrast in certain areas. CLAHE is applied to the image using local divisions or small blocks, and the statistical distribution (histogram) is calculated for each block individually. Subsequently, the contrast in each region is enhanced using the adjusted histogram. When combined with CLAHE, this dual

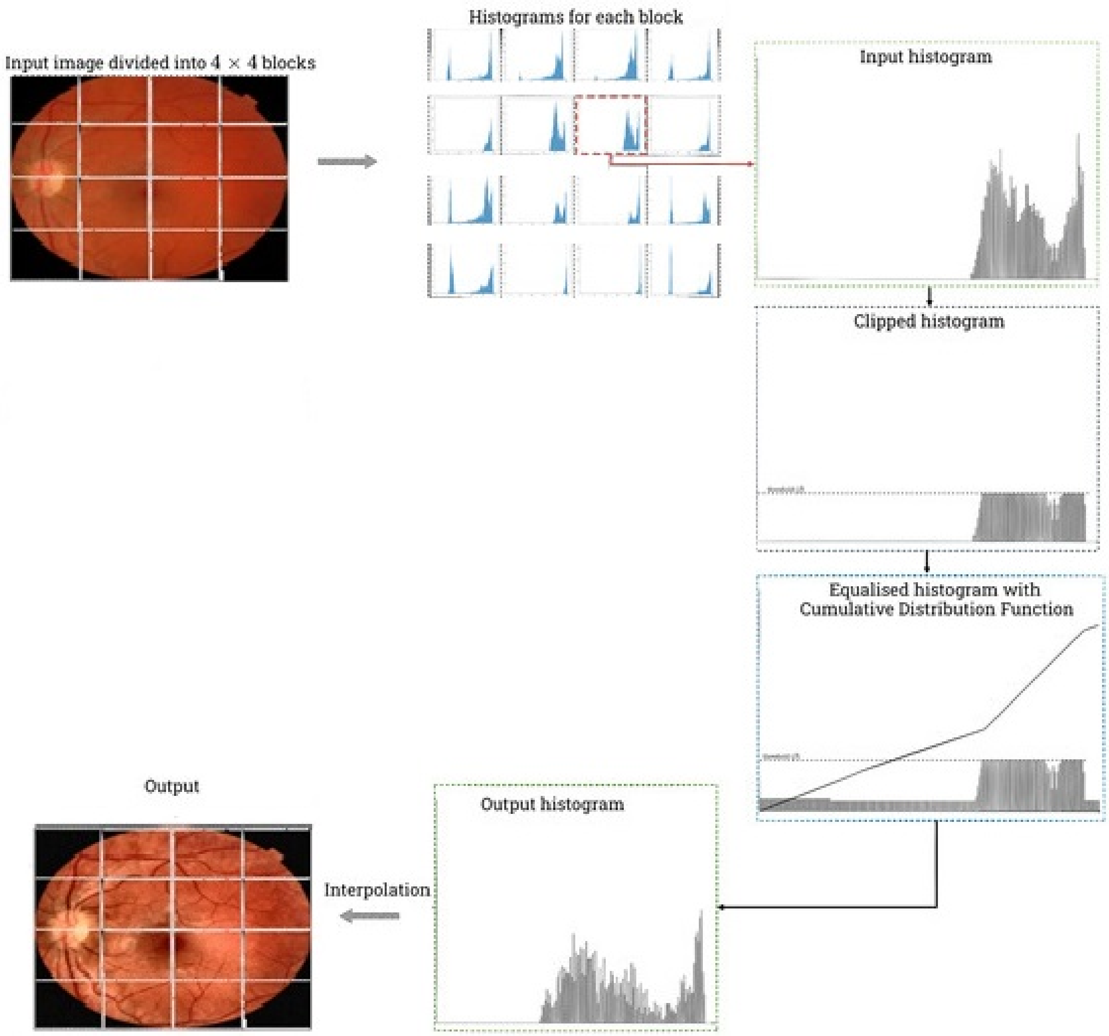

Input image divided into 4 × 4 blocks
Histograms for each block
Input histogram
Clipped histogram
Equalised histogram with
Cumulative Distribution Function
Output histogram
Interpolation
Output

Figure 2.5: Steps of the CLAHE algorithm to improve data quality.

preprocessing approach significantly enhances the quality of input data for deep learning models. CLAHE algorithm can be summarized in Table . 2.1. The steps have been displayed in Fig. 2.5.

2.6 Deep Learning Models

In this work, we utilized the recent deep learning methods, ResNet152 based on Convolutional Neural Networks (CNNs), and transformer-based models such as DeIT, Swin-ViT, and MaxViT.

Table 2.1: CLAHE algorithm steps.

Steps	Description
Step 1:	Divide Image into Tiles
Step 2:	Compute Local Histogram
Step 3:	Clip the Histogram
Step 4:	Redistribute Clipped Pixels
Step 5:	Equalize the Histogram
Step 6:	Interpolate Between Tiles
Step 7:	Reconstruct Image

2.6.1 ResNet-152 Model

ResNet-152 algorithm consists of 152 layers, a large number compared to traditional networks, which allows it to learn intricate and precise features from the data. Moreover, ResNet-152 leverages the concept of residual connections, where the original input is added to the output calculated by the deeper layers. This approach addresses the vanishing gradient problem, which can occur when training deep networks [4]. By incorporating residual connections, the network is able to learn more efficiently. Mathematically, the residual connection can be expressed as:

$$y = F\left(x, \{W_i\}\right) + x. \tag{2.7}$$

Where:

- y is the output of the residual block.
- $F\left(x, \{W_i\}\right)$ represents the learned function of the input x through the layers with weights W_i.
- x is the input to the residual block, and the addition of x ensures that the original input is retained, facilitating the gradient flow during back-propagation.

This design of ResNet-152 helps mitigate the vanishing gradient issue, enabling more effective training of very deep networks.

2.6.2 DeIT Transformer

DeIT model is based on the Vision Transformer architecture. In this setup, the larger "Teacher" model is trained on a large dataset, while the smaller "Student" model learns from the outputs of the "Teacher" during training [5]. The image is divided into small patches of fixed size, such as $P \times P$. These patches are flattened and transformed into vectors, which are then passed through Transformer layers using a self-attention mechanism. Each patch is represented as a token and processed with self-attention to obtain suitable representations of the full image. In DeiT, the larger "Teacher" model, which is pre-trained or larger in size, is used to instruct the smaller "Student" model. The smaller model is trained using both traditional loss functions (Cross-Entropy Loss) and a distillation loss that measures the difference between the outputs of the smaller model and the larger model using a metric like KL-Divergence [5]. Mathematically, Prediction Loss (Cross-Entropy Loss) can be noted as:

$$LCE = -\sum_{c=1}^{C} y_c log\left(\hat{y}_c\right). \tag{2.8}$$

Where, y_c is the target of class c and $\hat{y}_c$ is the predicted probability that the sample belongs to class c , which is the output of the model. KL-Divergence is used to measure the difference between the probability distributions produced by the larger (Teacher) and smaller (Student) models:

$$KL(P_{tch} \parallel P_{std}) = \sum_{c=1}^{C} P_{tch}(c) \log\left(\frac{P_{tch}(c)}{P_{std}(c)}\right). \tag{2.9}$$

Where, P_{tch} and P_{std} represent the probability distributions produced by the larger and smaller models, respectively. Finally, the total loss is computed by summing the two losses:

$$L_{total} = LCE + \lambda \cdot KL\left(P_{tch} \parallel P_{std}\right). \tag{2.10}$$

Where, λ is the weighting factor between the prediction loss and distillation loss.

2.6.3 Swin Vision Transformer

The Swin-ViT introduces improvements over the traditional Vision Transformer based on Self-attention within windows and shifting attention across windows, enhancing the model's efficiency and speed when handling large images [6]. In Swin-ViT, the image is divided into fixed-size windows, with each window being processed using a Self-attention mechanism that computes the interdependencies between pixels within that window. However, Self-attention at the level of individual windows may be insufficient in some cases, which is why shifted window attention is employed. This approach involves shifting the windows between network layers to promote interactions between neighboring windows, thereby improving the model's ability to capture long-range dependencies [6].

$$\text{Attention}\left(Q, K, V\right) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \tag{2.11}$$

Where, Q (Query), K (Key), V (Value) and d_k is dimension of keys.

2.6.4 Multi-axis Vision Transformer

MaxViT is a model belonging to the Transformer class, designed to enhance the capability of processing data across multiple axes, particularly in the context of images. The core idea behind MaxViT is to employ a Multi-axis Attention mechanism to improve the model's performance in image interpretation by focusing on several axes or directions, rather than relying on fixed axes as in traditional models. In MaxViT, attention is computed across multiple axes, which can be mathematically expressed as follows [7]:

$$\text{Attention}_{multi-axis}\left(Q, K, V\right) = \sum_{axis} \text{Attention}\left(Q_{axis}, K_{axis}, V_{axis}\right). \tag{2.12}$$

2.7 Conclusion

The proposed system is based on the fundus images datasets. In this chapter, we present the methodology for the detection of retinal diseases, along with the various components of the system. The process begins with the implementation of data augmentation techniques, followed by a preprocessing step to enhance the quality of the input data. Subsequently, deep learning models are employed for retinal diseases classification. The effectiveness of the system will be further validated in the next chapter through a detailed analysis of the results.

Chapter 3

Experimental Evaluation

3.1 Introduction

In this chapter, we present a comprehensive evaluation of our proposed deep learning system for ocular disease classification. The first part focuses on diabetic retinopathy detection using the APTOS 2019 dataset, while the second part tests the model's generalization to a broader dataset covering nine ocular diseases. We begin by outlining the key evaluation metrics, including Accuracy, Cohen's Kappa, F1 Score, and AUC, along with two clinically significant metrics—Sensitivity and Specificity—which are computed and discussed to assess both the model's ability to detect disease and avoid false alarms. We describe the training configuration, including the choice of architectures, optimization strategy, and the use of preprocessing techniques such as Ben's method combined with CLAHE. Four models—ResNet152, SwinV2, DeiT3, and MaxViT—are compared under three experimental conditions: using raw images, augmented data, and the proposed enhanced preprocessing. To interpret model decisions, Grad-CAM visualizations are included. Additionally, we compare our system's performance with existing literature on the same dataset. Finally, the model is evaluated on the extended ocular disease dataset, with detailed results including classification reports, confusion matrices, ROC curves, and overall diagnostic performance.

3.2 Evaluation Metrics

To assess the performance of the trained model in detecting diabetic retinopathy, we employed four primary evaluation metrics: Accuracy, Cohen's Kappa, F1 Score, and Area Under the Curve (AUC). We adopted the Quadratic Weighted Kappa variant because it more heavily penalizes large misclassifications—reflecting the ordinal nature of DR grades from 0 to 4—adjusts for chance agreement in a manner analogous to clinicians' inter-rater reliability assessments, is widely accepted in medical research for measuring agreement, and served as the official metric in the APTOS 2019 Kaggle competition to ensure direct comparability with benchmark submissions.

- **Accuracy:** Accuracy represents the percentage of correctly classified samples out of the total number of samples. It is one of the simplest performance metrics but can be misleading in cases of class imbalance [36].

- **Cohen's Kappa:** Cohen's Kappa measures the agreement between predicted labels and true labels, accounting for the possibility of agreement occurring by chance. It is considered a more reliable metric than accuracy, especially in situations with class imbalance.

- **F1 Score:** The F1 Score is the harmonic mean of precision and recall. It is often used when there is class imbalance and when both false positives and false negatives are considered important.

- **Area Under the Curve (AUC):** The AUC represents the model's ability to distinguish between different classes. A higher AUC indicates better classification performance. AUC is commonly used in medical tasks due to its importance in evaluating model sensitivity and stability,this parameter are from [36].

3.3 Training Settings

In this section, we present the training configurations adopted for building the diabetic retinopathy classification models. The experiments were conducted in an environment equipped with a GPU, using the PyTorch library to design and train the models. The Albumentations library was used for image augmentation, and Stratified K-Fold Cross-Validation was applied to ensure balanced class distribution across each fold. A series of experiments were conducted on all the models evaluated in this work, namely: ResNet152, SwinV2, DeiT3, and MaxViT, to determine the optimal hyperparameters that provide the best performance. The results showed that the ideal training settings were nearly identical across all models, allowing us to standardize the configuration for comparison. The final selected settings are summarized below, and the detailed results of the experiments, along with plots and tables justifying these choices:

- **Epochs:** Defines how many times the entire dataset is passed through the model during training.
- **Batch Size:** Determines how many samples are processed together before the model updates its weights.
- **Learning Rate:** Controls how quickly the model adjusts its parameters during training.
- **Loss Function:** Used for multi-class classification to measure prediction accuracy against true labels.

3.4 Analysis of Training Settings Experiments

To determine the most suitable training settings for all models used in this work, several experiments were conducted to test different values for key hyperparameters. The aim was to identify the configurations that provide the best performance in terms of model generalization, stability, and evaluation metrics.

Below are the experimental results for each hyperparameter:

3.4.1 Batch Size Impact

Table 3.1: Accuracy with varying batch sizes.

Batch Size	ResNet152	SwinV2	DeiT3	MaxViT
32	84%	84%	84%	85%
64	83%	83%	83%	84%
96	83%	82%	83%	83%
128	83%	82%	83%	83%

3.4.1.1 Analysis of Batch Size Impact on Model Performance

To study the effect of batch size on the performance of the four selected models, we fixed the number of training epochs (Epoch = 25) and the learning rate (Learning Rate = 1e-4), and experimented with various batch size values: 32, 64, 96, and 128. We chose to start with a batch size of 32 because it is a commonly used default in most machine learning frameworks such as PyTorch. Additionally, it provides a good balance between training stability, convergence speed, and GPU memory usage, making it a logical starting point for performance optimization. As shown in the table, the best results were obtained with a batch size of 32, where the MaxViT model achieved 85% accuracy, and the other models (ResNet152, SwinV2, and DeiT3) reached 84%. When the batch size was increased, performance either declined or remained unchanged, indicating that larger values may not be optimal in this context. These findings are visually confirmed by the bar chart fig 3.1, which clearly highlights the superior overall performance at a batch size of 32 compared to the other tested values.

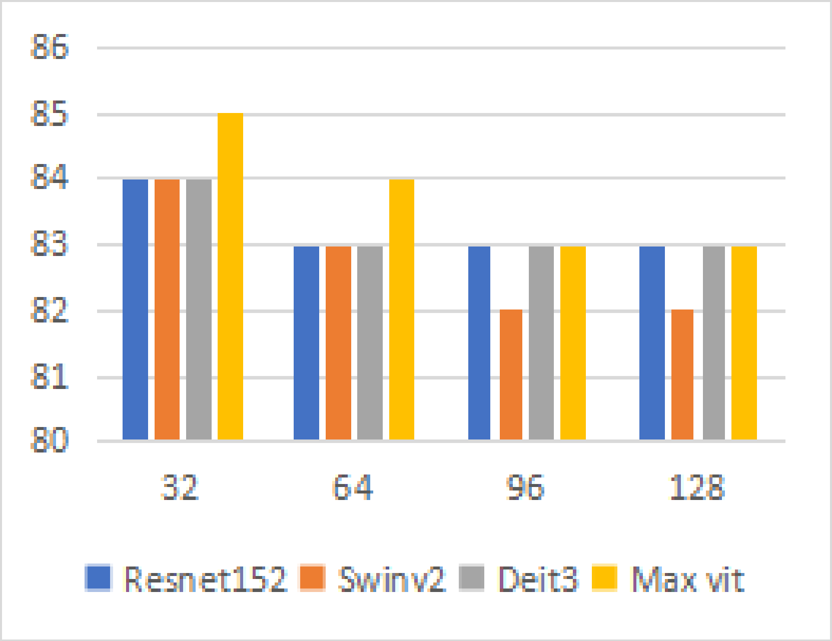


Figure 3.1: Accuracy comparison of models at different Batch size.

3.4.2 Epoch

Table 3.2: Accuracy vs. Epochs.

Epochs	ResNet152	SwinV2	DeiT3	MaxViT
25	84%	84%	84%	85%
50	83%	83%	83%	83%
75	83%	83%	83%	83%
100	83%	83%	83%	83%

3.4.2.1 Effect of Epoch Count on Model Accuracy

To assess the impact of the number of training epochs on model performance, we conducted experiments with 25, 50, 75, and 100 epochs, while keeping other parameters fixed (Batch Size = 32, Learning Rate = 1e-4). The results in the table show that the best accuracy for all models was achieved at 25 epochs, with MaxViT reaching 85%, while others achieved 84%. Increasing the number of epochs did not lead to performance improvement and may have introduced overfitting or plateauing. This is visually confirmed in fig 3.2, where performance stabilizes or slightly declines beyond 25 epochs.

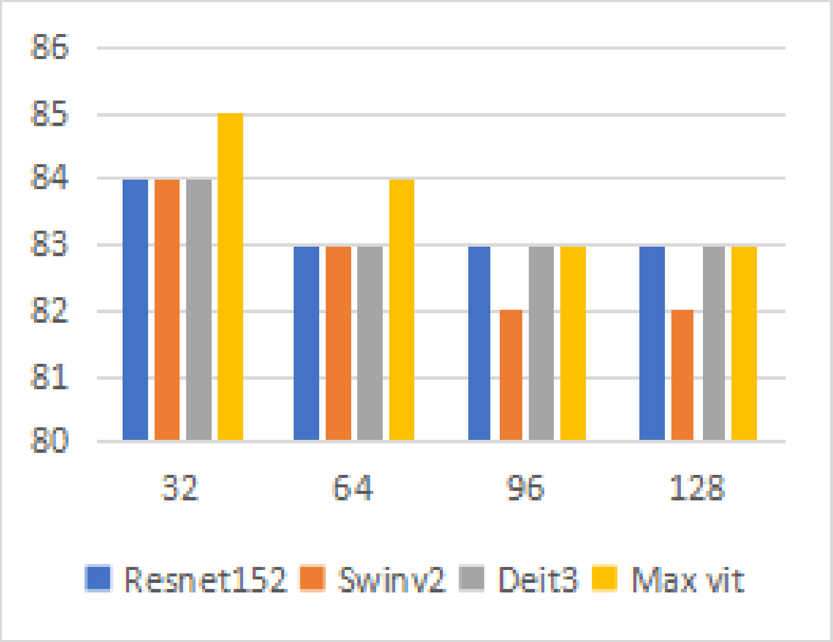


Figure 3.2: Model accuracy at different epoch counts.

3.4.3 Learning Rate Impact

Table 3.3: Accuracy vs. Learning Rate.

Learning Rate	ResNet152	SwinV2	DeiT3	MaxViT
1×10^{-2}	71%	68%	55%	74%
1×10^{-3}	81%	74%	71%	84%
1×10^{-4}	84%	83%	84%	85%

3.4.3.1 Effect of Learning Rate on Model Accuracy

We evaluated the impact of different learning rates (1e-2, 1e-3, and 1e-4) on model accuracy, keeping the batch size fixed at 32 and the number of epochs at 25. The results clearly indicate that a learning rate of 1e-4 yielded the best performance across all models, with MaxViT reaching 85% accuracy, followed by ResNet152 and DeiT3 with 84%. Higher learning rates (especially 1e-2) resulted in significantly lower accuracy, likely due to unstable or suboptimal convergence. These findings are illustrated in fig 3.5, confirming that 1e-4 is the optimal learning rate for stable and accurate training in our experiments.

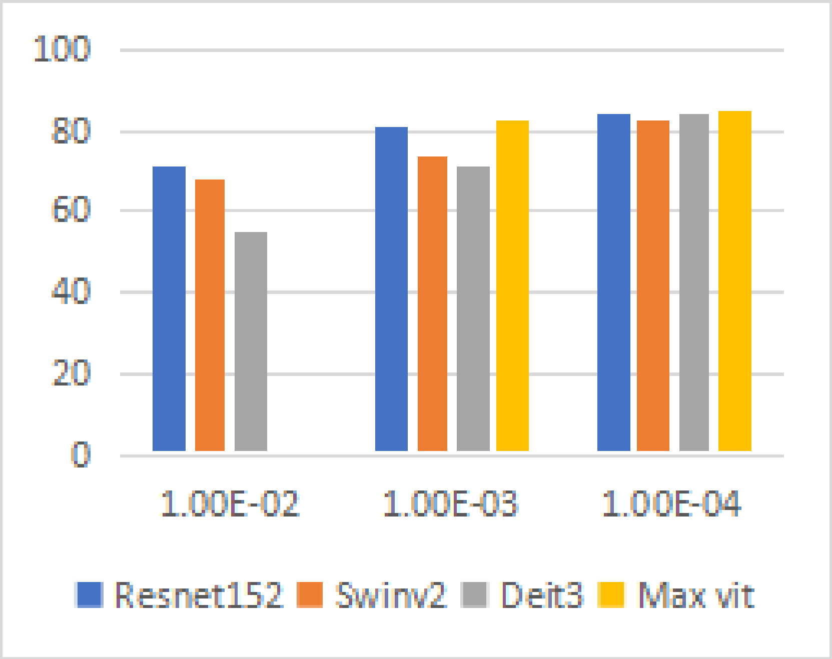


Figure 3.3: Comparison of model accuracy across different learning rates.

3.5 Model Performance Comparison

In this study, the performance of four state-of-the-art deep learning models for diabetic retinopathy classification—ResNet152, SwinV2, DeiT3, and MaxViT—was evaluated. The models were tested across three separate experiments, each applying a different preprocessing technique. The objective was to analyze the effect of each preprocessing method on model performance using key evaluation metrics: Accuracy, Quadratic Weighted Kappa, F1 Score, and AUC.

3.5.1 First Experiment: (Raw Data)

In this experiment, the models were trained and evaluated using the original retinal images without any preprocessing applied. tab 3.4 presents the performance of the four models after training on the augmented dataset, evaluated using the same metrics as in the first experiment.

Table 3.4: Comparison of models on various performance metrics Raw Data.

Model	AUC Score	F1 Score	Kappa Score	Accuracy
Resnet152	93.03%	83.36%	87.93%	84.29%
Swinv2	93.51%	83.48%	88.88%	84.15%
Deit3	93.44%	80.40%	85.09%	81.97%
Max vit	94.85%	84.11%	88.44%	84.70%

Analysis

Under raw data conditions, all models demonstrate strong classification capabilities, but their strengths vary by metric. MaxViT achieves the highest AUC (94.85%), F1 Score (84.11%), and Accuracy (84.70%), indicating exceptional discriminative power and balance between precision and recall. SwinV2, however, attains the highest Quadratic Weighted Kappa (88.88%), our primary metric for this study, which measures agreement with the ground truth after correcting for chance. This suggests SwinV2 provides the most reliable ordinal predictions among the models. ResNet152 delivers consistently solid performance across metrics, while DeiT3 shows comparatively lower F1 and Accuracy, reflecting less effective raw-feature extraction.

3.5.2 Second Experiment: Data Augmentation

In the second experiment, various data augmentation techniques were applied to the raw retinal images to increase data diversity and improve model generalization. This included transformations such as rotation, flipping, and scaling. tab 3.6 presents the performance of the four models after training on the augmented dataset, evaluated using the same metrics as in the first experiment.

Table 3.5: Comparison of models on various performance metrics(data augmentation) .

Model	**AUC Score**	**F1 Score**	**Kappa Score**	**Accuracy**
Resnet152	94.69%	82.17%	88.62%	84.15%
Swinv2	95.44%	85.52%	90.91%	86.20%
Deit3	94.45%	80.69%	86.85%	82.65%
Max vit	95.39%	84.94%	90.16%	85.79%

Analysis:

In this experiment, the application of data augmentation led to noticeable improvements in model performance compared to the raw data scenario. All four models benefited from the increased variability in the training data, with enhanced results across all evaluation metrics. ResNet152 showed a slight improvement in AUC (94.69%) and maintained a competitive Kappa score (88.62%), indicating a more stable performance. SwinV2 achieved the highest Kappa score (90.91%), marking it as the top-performing model in this experiment according to our primary metric. It also recorded the highest F1 score (85.52%) and Accuracy (86.20%), which further confirms its robustness. MaxViT closely followed SwinV2, with a high Kappa score of 90.16 and AUC of 95.39%, suggesting strong classification capabilities. DeiT3 showed the least improvement among the models, with a Kappa score of 86.85% and the lowest accuracy (82.65%), despite a solid AUC of 94.45%. Based on the Quadratic Weighted Kappa, SwinV2 is considered the best-performing model in this experiment. Its consistent strength across all metrics, especially in agreement with true labels, highlights its effectiveness when trained on augmented data.

3.5.3 Third Experiment: Full Processing

In the final experiment, we evaluated the models using our proposed preprocessing pipeline, which combines Ben's color enhancement technique with CLAHE (Contrast Limited Adaptive Histogram Equalization). This method was designed to enhance important features in retinal images while improving contrast and reducing noise. The goal was to investigate whether this tailored preprocessing strategy would further improve classification performance. tab 3.6 presents the results.

Table 3.6: Comparison of models on various performance metrics(Full Processing).

Model	AUC Score	F1 Score	Kappa Score	Accuracy
Resnet152	95.11%	84.61%	90.28%	85.52%
Swinv2	95.62%	88.11%	92.95%	88.11%
Deit3	94.88%	83.58%	90.83%	84.56%
Max vit	95.80%	86.58%	91.09%	86.89%

Analysis:

In the third experiment, all models demonstrated noticeable improvements compared to the previous experiments, indicating the effectiveness of the proposed preprocessing strategy. Among the models: SwinV2 achieved the highest Kappa Score of 92.95%, which is the primary evaluation metric in this study, indicating superior agreement with ground truth labels. It also recorded the highest F1 Score (88.11%) and matched the highest Accuracy (88.11%), reinforcing its strong performance across multiple metrics. MaxViT followed closely with a Kappa Score of 91.09% and the highest AUC Score (95.80%), showing excellent capability in distinguishing between classes. ResNet152 and DeiT3 also showed solid performance, but their Kappa Scores (90.28% and 90.83% respectively) were slightly lower than those of SwinV2 and MaxViT. Considering all evaluation metrics—especially Quadratic Weighted Kappa, which is the main metric adopted in both this study and the APTOS 2019 Kaggle competition—SwinV2 stands out as the best-performing model in this experiment.

3.6 Grad-CAM Visualizations

Interpreting Model Decisions Using Grad-CAM

Gradient-weighted Class Activation Mapping (Grad-CAM) is a widely used visualization technique that helps interpret the decision-making process of deep convolutional neural networks. It highlights the regions in the input image that the model considers most relevant when predicting a specific class.

In this study, Grad-CAM was manually implemented using PyTorch, without relying on external libraries. The technique works by registering forward and backward hooks to extract both the feature maps and gradients from a specific layer of the model—specifically, the norm layer in the base transformer models used. These values are then used to generate a class-discriminative heatmap that localizes important regions for the target prediction.

The mathematical principle behind Grad-CAM involves computing the gradients of the class score y^c with respect to the feature maps A^k [37].

These gradients are globally average-pooled to obtain the importance weights α_k^c, which are then combined with the feature maps to produce the final heatmap as follows:

$$\mathrm{ReLU}\left(\sum_k \alpha_k^c A^k\right) = L_{\text{Grad-CAM}} \tag{3.1}$$

This allows for highlighting only the features that positively influence the prediction, helping to visually confirm whether the model is focusing on medically relevant areas such as the macula, hemorrhages, or retinal vessels.

To gain a deeper understanding of how the trained models make their predictions, Grad-CAM (Gradient-weighted Class Activation Mapping) was utilized to generate heatmaps that highlight the regions each model focused on during classification. Grad-CAM serves as a critical interpretability tool in the context of medical imaging, as it allows us to verify whether the model relies on medically relevant features such as the macula, hemorrhages, blood vessels, or other damaged regions in the retina. In this study, Grad-CAM was applied to the four models used—ResNet152, SwinV2, DeiT3, and MaxViT—and the resulting visualizations were analyzed using one representative image from each of the five diabetic retinopathy classes (0 to 4). For each model, five Grad-CAM images are presented, corresponding to the five severity levels.

Ben’s Preprocessing + CLAHE Results

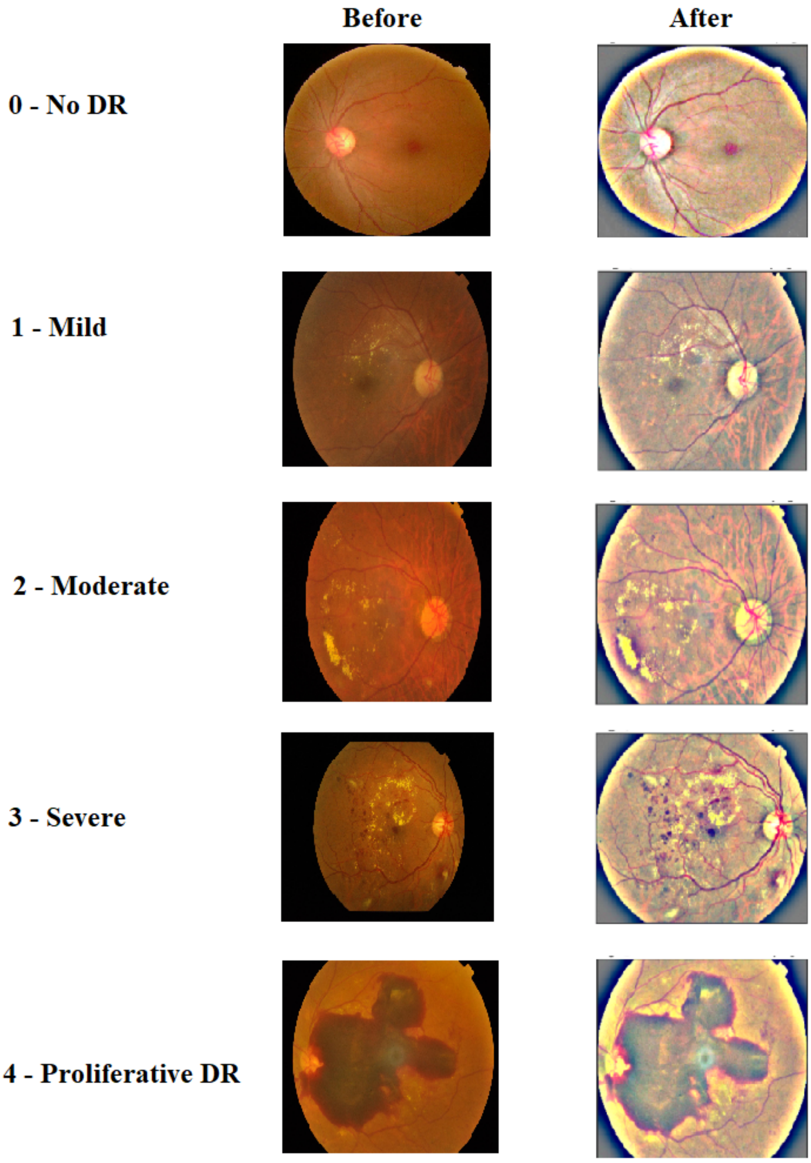


Figure 3.4: Comparison of fundus images before and after Proposed Prepossessing.

Results

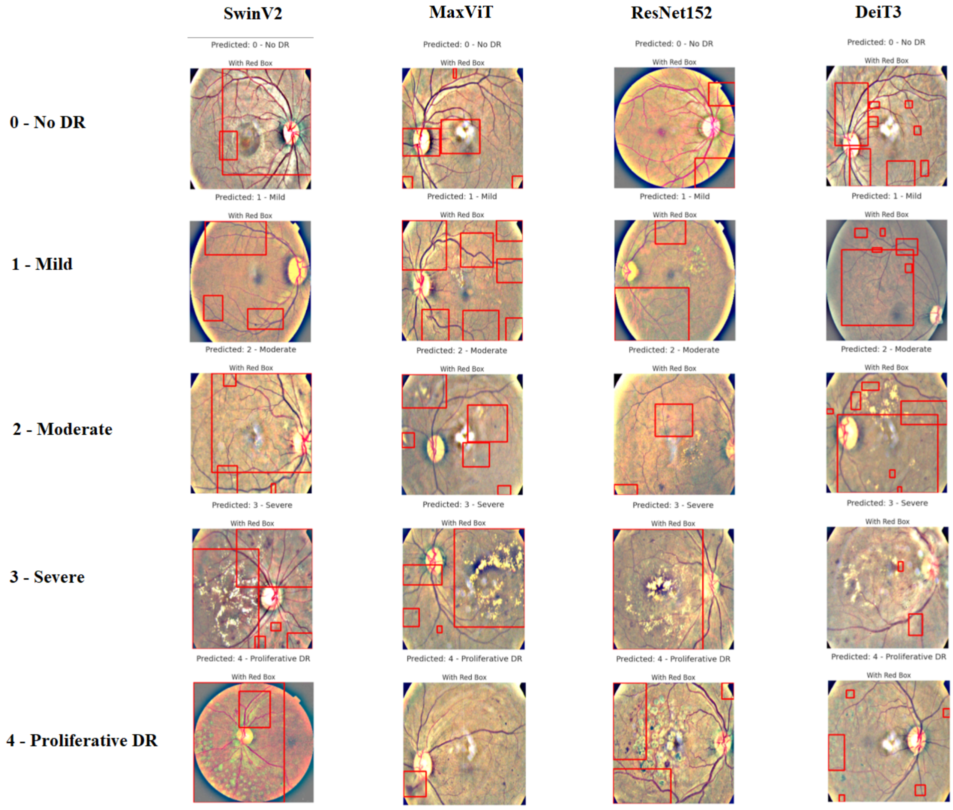


Figure 3.5: Grad-CAM visualizations across DR grades and models.

This fig 3.5 illustrates the Grad-CAM heatmaps generated by the four evaluated models (SwinV2, MaxViT, ResNet152, and DeiT3) for each diabetic retinopathy (DR) grade: 0 (No DR) to 4 (Proliferative DR). The red boxes indicate the regions the model focused on when making its prediction.

Observation

Across all models, the central regions of the retina—especially around the macula and optic disc—are highlighted, which aligns with clinical importance. The models generally perform well in localizing relevant features, such as hemorrhages and exudates, particularly in higher severity levels (grades 3 and 4), Notably:

- **SwinV2 and MaxViT** demonstrate sharper and more focused activations, particularly in severe and proliferative DR cases, suggesting better spatial attention.

- **ResNet152** provides balanced localization but with slightly less sharp focus compared to transformer-based models.
- **DeiT3** occasionally highlights broader or more dispersed regions, which may indicate lower localization precision in some cases.

These visualizations support the interpretability of the models and validate that the predictions are based on medically relevant regions of interest.

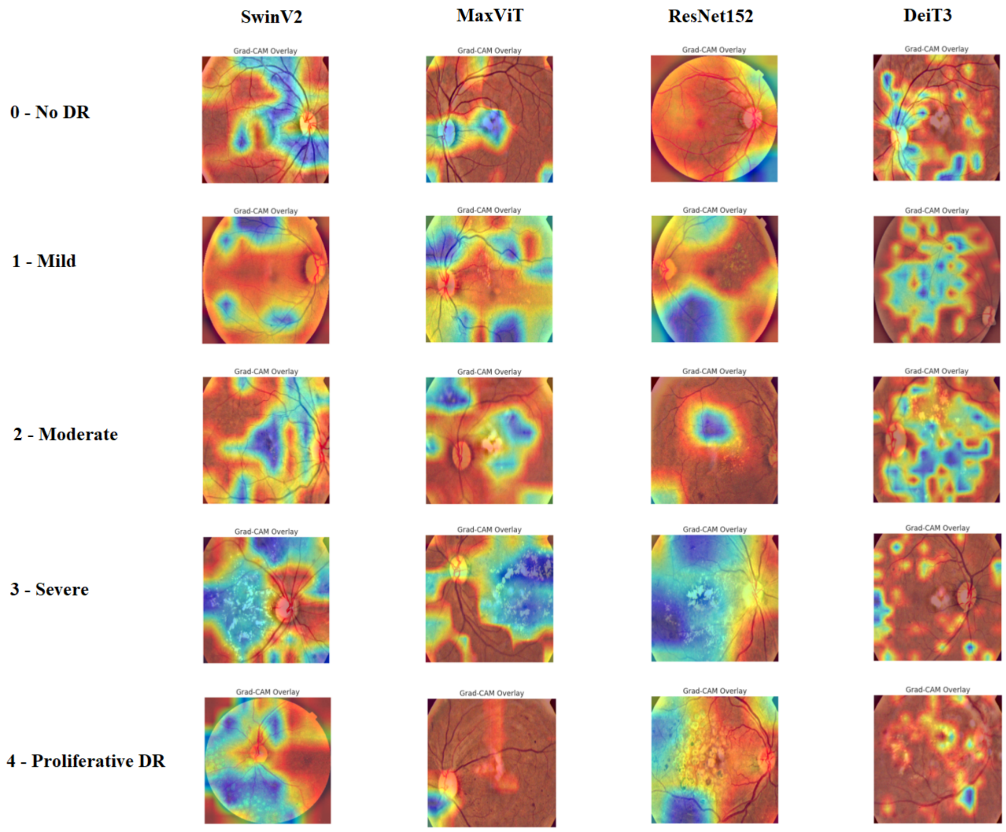


Figure 3.6: Grad-CAM overlay examples from different models.

This fig 3.6 displays Grad-CAM heatmap overlays for each diabetic retinopathy (DR) class using four different models: SwinV2, MaxViT, ResNet152, and DeiT3. These overlays highlight the most influential regions in the fundus images that contributed to the model's prediction.

Observation

In these visualizations, blue regions indicate areas of highest relevance according to the model's internal gradients, while red areas are less relevant. Notable insights include:

- **SwinV2 and MaxViT** tend to focus on distinct pathological regions, especially in moderate to severe DR cases, showing good spatial discrimination.
- **ResNet152** often concentrates on central retinal regions with sharper and more localized activation in earlier stages.
- **DeiT3** generally highlights more scattered areas, which may reflect broader attention but also potential noise in focus.

These heatmaps enhance transparency in model decisions, making it easier to validate whether the models are attending to medically meaningful features.

3.7 Optimal Model Performance and Outcome Evaluation

This section provides a detailed evaluation of the performance of the best-performing model, SwinV2, trained using the proposed preprocessing method (Ben's preprocessing + CLAHE). The assessment includes multiple metrics such as the classification report, confusion matrix, sensitivity, specificity, and ROC curves. These metrics collectively offer a comprehensive understanding of the model's ability to accurately classify the five severity levels of diabetic retinopathy based on the APTOS 2019 dataset.

3.7.1 Classification Report

Table 3.7: Classification report per DR class including Precision, Recall, Specificity, F1-Score, and Support.

Class	Precision	Recall (Sensitivity)	Specificity	F1-Score	Support
0 - No DR	98%	99%	91%	98%	361
1 – Mild	76%	64%	97%	69%	74
2 - Moderate	80%	91%	93%	85%	199
3 - Severe	72%	46%	98%	56%	39
4 - Proliferative DR	78%	73%	97%	75%	59

3.7.2 Confusion Matrix

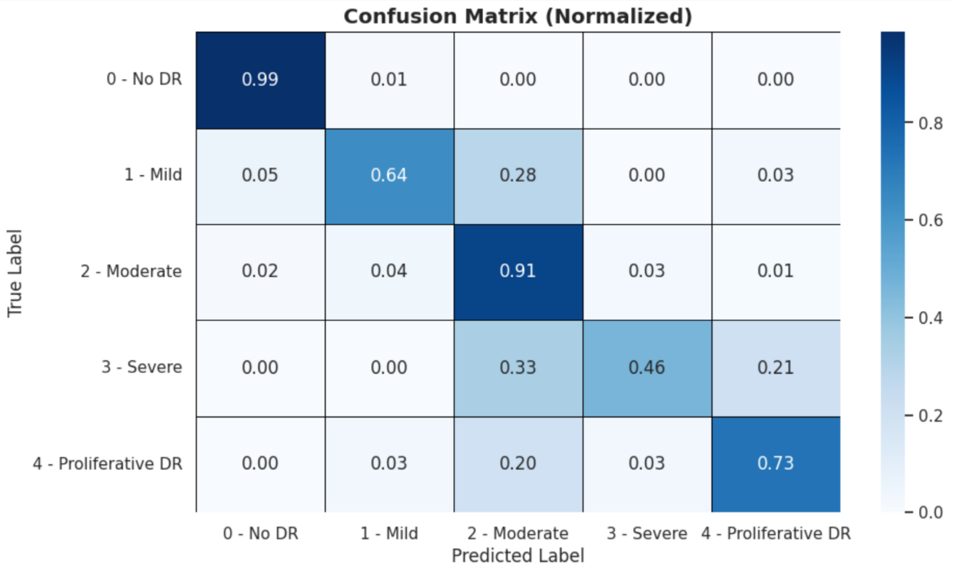


Figure 3.7: Confusion Matrix for DR classification.

3.7.3 ROC Curves

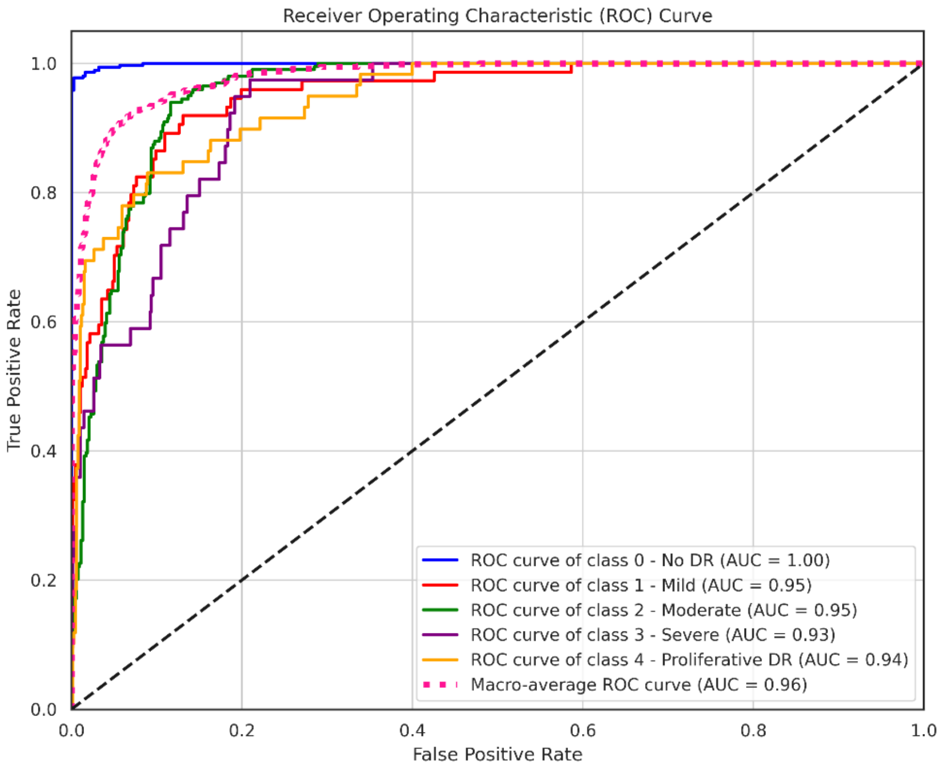


Figure 3.8: ROC Curves for all classes.

3.8 Comprehensive Results Analysis

The proposed model's performance was thoroughly evaluated using multiple standard evaluation tools, including the classification report, confusion matrix, overall sensitivity and specificity, and the Receiver Operating Characteristic (ROC) curves for each of the five diabetic retinopathy (DR) classes.
As presented in tab 3.7 the classification report highlights the model's consistent and balanced performance across all DR categories. The model achieved an overall precision of 88.80%, an overall sensitivity (recall) of 88.24%, and an F1-score of 88.38%, indicating its strong capability in correctly classifying both majority and minority classes. In addition, The overall specificity reached 96.38%,which reflects the model's high reliability in correctly identifying non-DR cases and minimizing false positives—an essential aspect in medical diagnostics. The confusion matrix fig 3.7 further demonstrates the model's classification accuracy, with a clear concentration of correct predictions along the diagonal. The model showed outstanding performance in detecting Class 0 (No DR) and Class 4 (Proliferative DR), with minimal misclassifications. The other classes—Class 1 (Mild), Class 2 (Moderate), and Class 3 (Severe)—were also predicted with high reliability, highlighting the model's ability to discern between subtle variations in retinal pathology. Additionally, the ROC curves fig 3.8 provide further validation of the model's effectiveness.

The curves for all five classes show high Areas Under the Curve (AUC), each exceeding the commonly accepted threshold of 90% for high-performing classifiers. This reflects excellent trade-offs between true positive rates and false positive rates across the board, confirming the model's robust discriminative ability. In conclusion, the proposed system demonstrates strong and balanced diagnostic capabilities for multi-class diabetic retinopathy classification. The combination of advanced feature extraction using SwinV2 and refined preprocessing strategies (Ben's preprocessing + CLAHE) played a crucial role in achieving these outcomes. These promising results underline the system's potential for real-world deployment in automated DR screening and diagnostic decision support.

3.9 Comparison With State-of-the-Art

Overview of Selected Studies

In this section, we compare the performance of our proposed model with other recent studies that used the APTOS 2019 dataset to classify diabetic retinopathy into five classes. The comparison focuses on key evaluation metrics such as Kappa, Accuracy, and AUC, and also considers the models and preprocessing techniques used in each study. This helps to clearly show how our approach performs relative to existing work.

Performance Comparison

Table 3.8: Comparative performance proposed models of existing approaches on APTOS19.

Study	Model Used	Preprocessing	Accuracy	Kappa
Bodapati al [38]	Fusion of features from Xception and VGG16 with DNN	Not specified	81.7%	71.1%
Tymchnko al [39]	Convolutional Neural Network with multi-stage transfer learning	Not specified	Not reported	92.55%
Fan al [40]	CNN with Adaptive Multi-stage Feature Fusion	Data Augmentation	85.32%	77.26%
Our Proposed	SwinV2 MaxViT ResNet152 DeiT3	Ben's + CLAHE	88.11% 86.89% 85.52% 84.56%	92.95% 91.09% 90.28% 90.83%

Result Analysis

Tab 3.8 presents a comprehensive comparison between our proposed models and several state-of-the-art approaches for five-class diabetic retinopathy classification. The comparison includes the models used, preprocessing strategies, and their performance in terms of accuracy and Cohen's Kappa score. Bodapati et al. (2020) implemented a hybrid approach by fusing features from Xception and VGG16 through a deep neural network. However, they did not specify any preprocessing strategy. Their model achieved an accuracy of 81.7% and a Kappa score of 71.1%, which reflects moderate agreement and highlights the limitations in feature representation and data handling. Tymchenko et al. (2020) employed an Xception-based convolutional neural network with data augmentation. In addition, they incorporated external datasets to expand the training set. Despite this enhancement, the study did not report the accuracy, and only provided a Kappa score of 92.55%. Although this indicates strong agreement, the lack of accuracy reporting limits the comprehensiveness of performance evaluation, especially when compared to models trained and tested on the same dataset without external additions. Fan et al developed a convolutional neural network enhanced with Adaptive Multi-stage Feature

Fusion, and utilized data augmentation as a preprocessing strategy. Their model achieved 85.32% accuracy and 77.26% Kappa score. While these results show improvement over earlier methods, the reliance on traditional CNN architecture may limit the model's ability to effectively capture global and hierarchical image features. In contrast, our proposed models—particularly the SwinV2 model combined with Ben's preprocessing and CLAHE—achieved superior results, reaching 88.11% accuracy and 92.95% Kappa score. This demonstrates the effectiveness of modern transformer-based architectures in capturing long-range dependencies and detailed visual cues in retinal images. Furthermore, the additional tested models—MaxViT, ResNet152, and DeiT3—also exhibited competitive performance, all outperforming previously published works in both metrics. These findings emphasize the effectiveness of combining modern deep learning architectures with suitable preprocessing strategies to improve the performance of multi-class retinal disease classification. Rather than integrating multiple models, each architecture was evaluated independently, demonstrating that transformer-based models—when coupled with appropriate image enhancement techniques—can achieve state-of-the-art results.

3.10 Generalization to Other Eye Diseases

After identifying the best-performing model for DR classification using the APTOS 2019 dataset, we extended our investigation to evaluate the model's generalization capabilities on a broader set of ocular diseases. Specifically, this phase focuses on assessing the model's ability to accurately classify nine distinct ocular disease categories, representing a diverse range of eye conditions with varying visual characteristics. To maintain consistency and ensure fair evaluation, we employed the same training strategy, hyperparameters, and evaluation metrics used in the previous experiments. The objective here is not to draw direct comparisons between datasets but to validate the robustness and adaptability of the proposed system when exposed to new and heterogeneous diagnostic categories. The following subsections present both global performance metrics—such as accuracy, Cohen's Kappa score, F1-score, and AUC—and detailed diagnostic results, including the classification report, confusion matrix, and ROC curves.

Global Metrics

To assess the overall diagnostic effectiveness of the model on the expanded ocular disease dataset, we calculated four primary performance metrics: Accuracy, Cohen's

Kappa Score, F1-Score, and Area Under the ROC Curve (AUC). These metrics offer a comprehensive overview of the model's generalization capability beyond diabetic retinopathy.

In term of accuracy, the model achieved an overall accuracy of 87.98%, indicating that a high proportion of images were correctly classified across the nine disease categories. Also for Kappa Score, the Cohen's Kappa score reached 93.91%, reflecting a strong agreement between the model predictions and the ground truth labels, while accounting for random chance. This high score reinforces the reliability of the model in multi-class medical. For F1 Score, the macro-averaged F1 score was 86.73%, balancing both precision and recall across all classes. This value highlights the model's ability to maintain a fair performance even in the presence of class imbalance or visually similar conditions. Lastly, the macro-averaged AUC was 98.56%, indicating an excellent ability to discriminate between classes across all decision thresholds. An AUC close to 100% confirms the model's robustness in differentiating pathological signs in retinal images.

The obtained results indicate that the SwinV2 model maintains a high level of performance even when applied to wider range of ocular diseases, demonstrating its efficiency and robustness. Moreover, these results support the effectiveness of the proposed system architecture in diagnosing various eye-related pathological conditions.

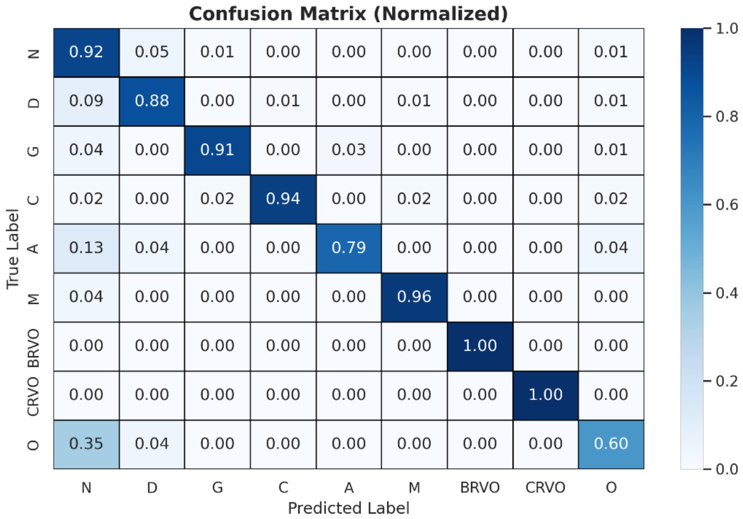
Detailed Diagnostic Performance

In this subsection, we examine the model's behavior on each of the nine ocular disease categories by presenting the full classification report, the normalized confusion matrix, and the ROC curves.

Table 3.9: Classification Report – Generalization to Other Eye Diseases.

Class	**Precision**	**Recall**	**Specificity**	**F1-Score**	**Support**
0 - Normal (N)	78%	92%	97%	84%	435
1 - Diabetes (D)	92%	88%	98%	90%	424
2 - Glaucoma (G)	91%	91%	99%	91%	69
3 - Cataract (C)	91%	94%	98%	92%	64
4 - AMD (A)	95%	79%	99%	87%	53
5 - Myopia (M)	91%	96%	100%	93%	51
6 - BRVO	100%	100%	100%	100%	9
7 - CRVO	100%	100%	100%	100%	4
9 - Other	89%	60%	97%	72%	182

3.10.1 Confusion matrix



Confusion Matrix (Normalized)

True Label \ Predicted Label	N	D	G	C	A	M	BRVO	CRVO	O
N	0.92	0.05	0.01	0.00	0.00	0.00	0.00	0.00	0.01
D	0.09	0.88	0.00	0.01	0.00	0.01	0.00	0.00	0.01
G	0.04	0.00	0.91	0.00	0.03	0.00	0.00	0.00	0.01
C	0.02	0.00	0.02	0.94	0.00	0.02	0.00	0.00	0.02
A	0.13	0.04	0.00	0.00	0.79	0.00	0.00	0.00	0.04
M	0.04	0.00	0.00	0.00	0.00	0.96	0.00	0.00	0.00
BRVO	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
CRVO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
O	0.35	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.60

Figure 3.9: Confusion matrix for generalized retinal diseases.

3.10.2 ROC Curve and AUC

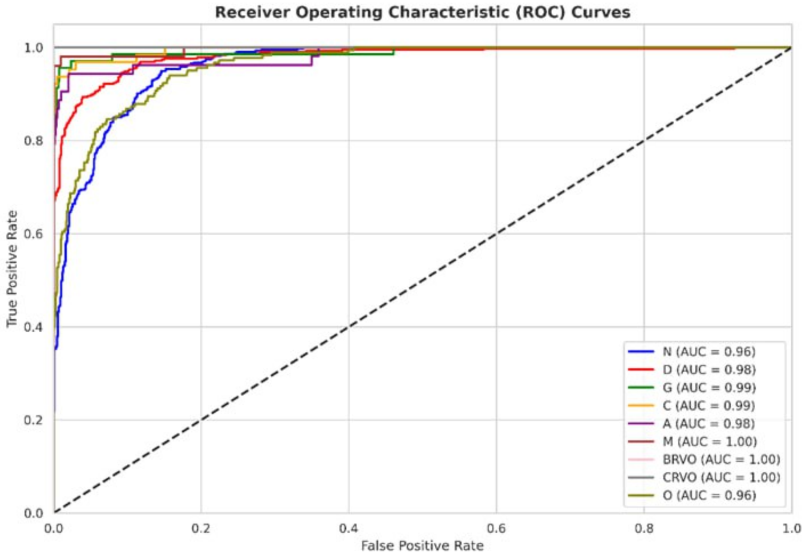


Figure 3.10: ROC Curves for generalized retinal diseases classification.

3.10.3 Comprehensive Results Analysis

The generalized classification model was comprehensively evaluated using a wide array of diagnostic tools, including the full classification report, the confusion matrix, global sensitivity and specificity values, and individual Receiver Operating Characteristic (ROC) curves for each of the nine ocular disease categories.
As presented in tab 3.9 the classification report reveals the model's robust and reliable performance across a heterogeneous set of eye diseases. The model achieved an overall precision of 89.14%, overall sensitivity (recall) of 86.00%, and an F1-score of 88.48%, highlighting its strong ability to correctly classify both common and rare ocular conditions. Furthermore, the overall specificity reached 99.00%, reinforcing the model's reliability in correctly identifying non-diseased classes and minimizing false positives—an essential factor in clinical applications where overdiagnosis must be avoided. The normalized confusion matrix fig 3.9 illustrates a clear dominance of correct predictions along the diagonal, suggesting that the model consistently identifies the correct class with minimal confusion between categories. High-precision results were particularly observed in BRVO and CRVO, with perfect scores (100%) in both sensitivity and specificity, despite the small number of support images—demonstrating the model's capability to generalize well, even on limited samples. Meanwhile, classes such as No DR (N) and Diabetic Retinopathy (D) also achieved notable results, with sensitivities of 92% and 88%, respectively, showing that the model continues to excel in detecting diabetic-related conditions. For Myopia (M) and Glaucoma (G), the model maintained high recall and specificity, further indicating that it can capture diverse pathological patterns within fundus images. However, slight performance variability was observed in the Other Conditions (O) class, which recorded a sensitivity of 60%. This can be attributed to the heterogeneity of visual patterns and underlying pathologies grouped within this "catch-all" category. Despite this, the high specificity (97%) ensures that the model remains cautious in assigning such broad diagnoses, a desirable trait in clinical screening.

The ROC curves presented in fig 3.10 further validate the classifier's diagnostic strength. Each curve lies well above the diagonal baseline, with all AUC values exceeding 93%. The macro-averaged AUC of 98.56% confirms the system's exceptional ability to distinguish between multiple eye diseases across varying decision thresholds. These results affirm the model's strong discriminative power in real-world screening scenarios.
In summary, the proposed model—leveraging the SwinV2 architecture and enhanced image preprocessing techniques (Ben's preprocessing + CLAHE)—demonstrates outstand-

ing diagnostic capabilities across a diverse spectrum of ocular diseases. Its performance remains stable and reliable even in the face of class imbalance and inter-class visual similarities. This robustness positions the system as a powerful and scalable diagnostic support tool, capable of enhancing early detection and classification in a wide range of ophthalmic applications.

Experimental conclusion

In this chapter, a comprehensive experimental study was presented to evaluate the performance of a set of advanced deep learning models in classifying the stages of diabetic retinopathy using the APTOS 2019 dataset. The study consisted of three different experimental phases: using raw data, applying data augmentation techniques, and finally the proposed system, which relies on enhanced preprocessing using CLAHE and Ben's preprocessing methods.

The results demonstrated that the SwinV2 model achieved the best performance across all evaluation metrics, especially in terms of Kappa score and classification accuracy, outperforming the other models (MaxViT, ResNet152, DeiT3). Furthermore, the model analysis was enriched through the Grad-CAM technique, which highlighted SwinV2's strong ability to focus precisely on relevant regions in the images, indicating its efficiency in decision interpretation. Additionally, the proposed system's results were compared with previous studies in the literature and showed superior performance, despite relying solely on the original dataset without incorporating external data. The final results were also analyzed through the confusion matrix and the classification report, confirming the model's capability to accurately identify different disease stages, with notable strengths in sensitivity and specificity across most classes.
Therefore, the proposed system can be considered a promising step toward the development of accurate and reliable AI-based diagnostic tools in the medical field, particularly for the early detection of diabetic retinopathy.

3.11 Our proposed Web-Based Diagnostic Platform

3.11.1 Overview of the Deployment Strategy

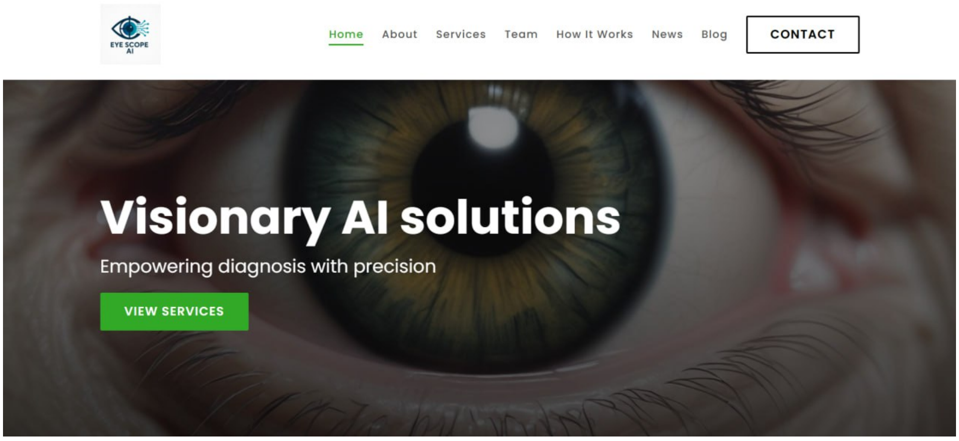
To bridge the gap between research and practical implementation, we deployed our best-performing deep learning models into a web-based diagnostic platform. The system is designed to facilitate automated detection and staging of retinal diseases using fundus images. It consists of a two-stage classification workflow where the first model detects the presence of any of six retinal conditions or confirms a normal retina, and a second model performs diabetic retinopathy (DR) staging if DR is detected.

3.11.2 Model Selection for Deployment

The SwinV2 model was selected for deployment due to its superior performance in both multi-disease and DR-specific tasks. On the APTOS 2019 dataset, it achieved an accuracy of 85%, a F1-score of 84.94%, and an AUC of 95.39% when using enhanced preprocessing. For the ODIR 2019 dataset, SwinV2 also led all performance metrics in general disease classification. These results demonstrated the model's generalization capability and made it an ideal candidate for real-time deployment.

3.11.3 System Workflow

The user will be greeted with Platform home page wich contain info and Serveries



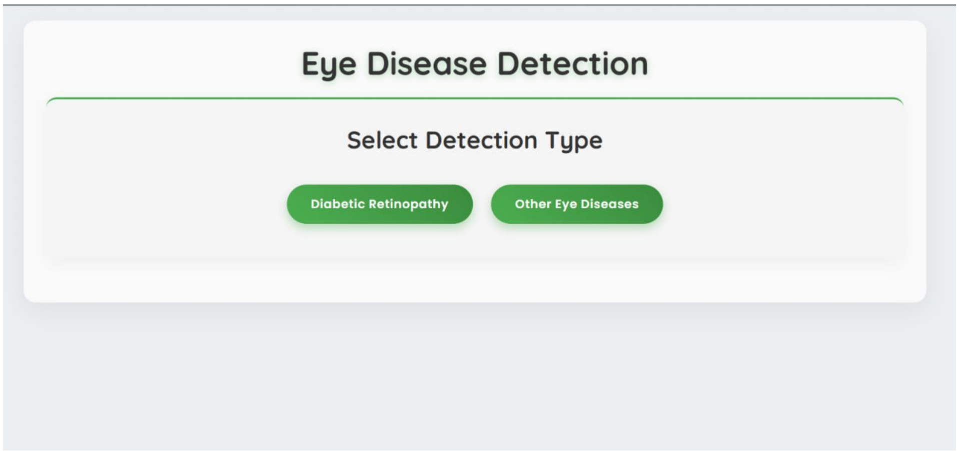
Visionary AI solutions

Empowering diagnosis with precision

Figure 3.11: platform home page.

The services we providing varies from ai diagnostics, Schedule appointment to Enhanced treatment planning .

The workflow of the diagnostic platform is structured as follows: The user uploads a fundus image via the web interface.

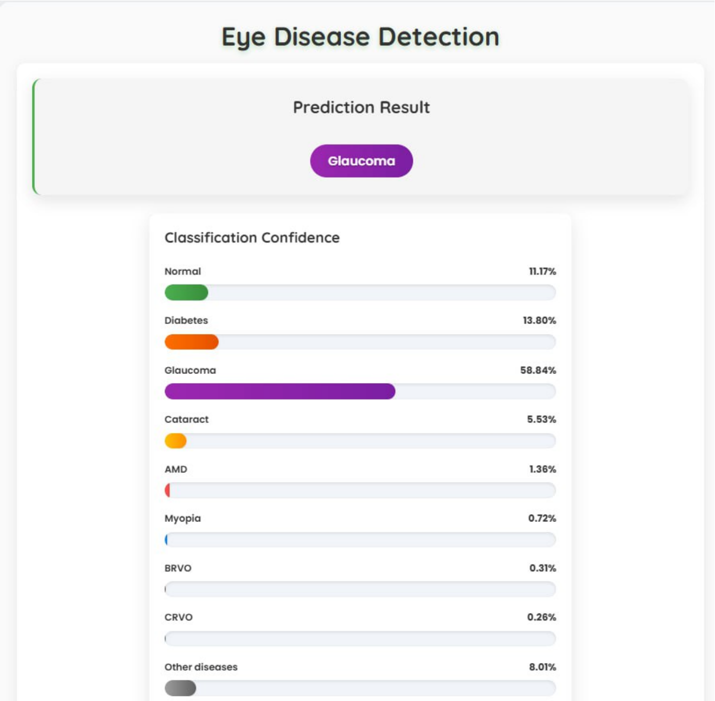


Eye Disease Detection

Select Detection Type

Figure 3.12: Diagnose interface.

1. **Stage 1:** The image is classified by the general disease detection model into one of seven categories: Normal, Diabetic Retinopathy, Glaucoma, AMD, Cataract, Pathologic Myopia, or RVO.



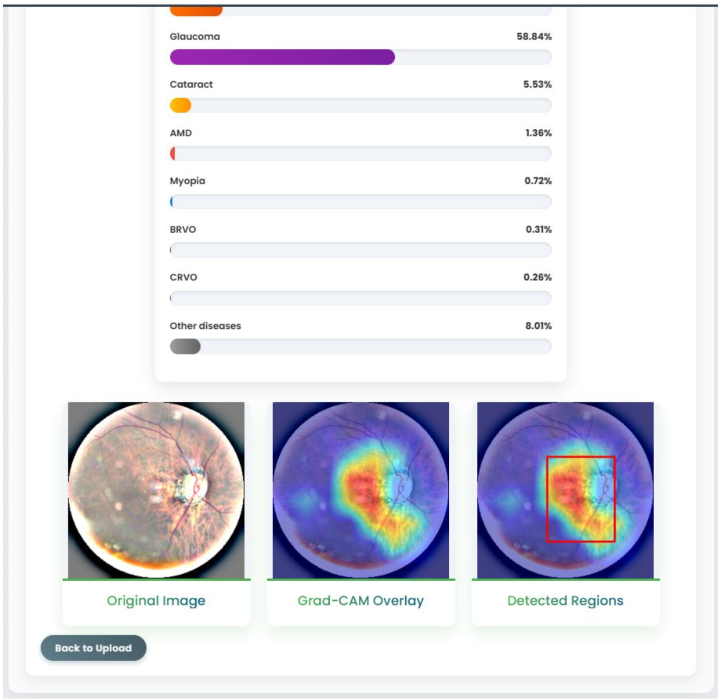
Eye Disease Detection

Prediction Result

Classification Confidence

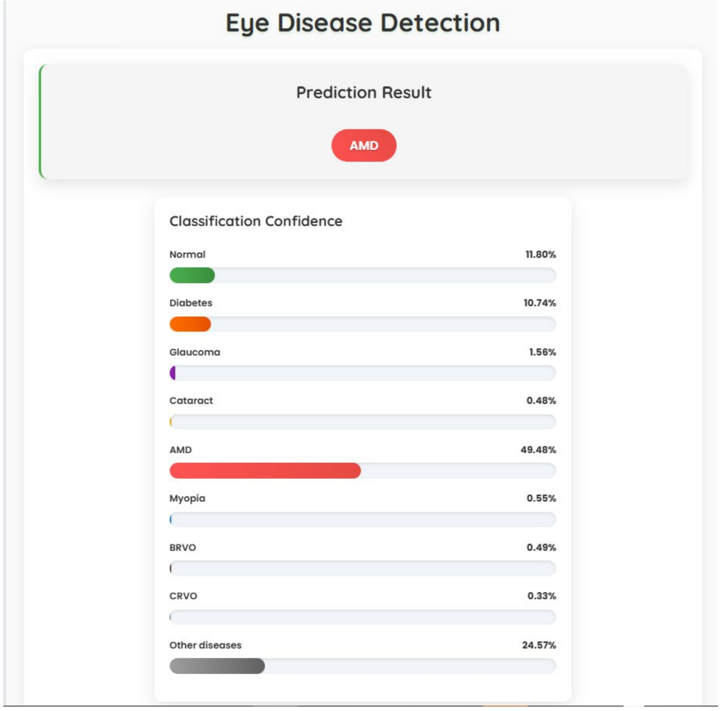
Normal	11.17%
Diabetes	13.80%
Glaucoma	58.84%
Cataract	5.53%
AMD	1.36%
Myopia	0.72%
BRVO	0.31%
CRVO	0.26%
Other diseases	8.01%

Figure 3.13: Example of Glaucoma detection percentage.



Glaucoma	58.84%
Cataract	5.53%
AMD	1.36%
Myopia	0.72%
BRVO	0.31%
CRVO	0.26%
Other diseases	8.01%

Figure 3.14: Exemple of detection Glaucoma visualization.



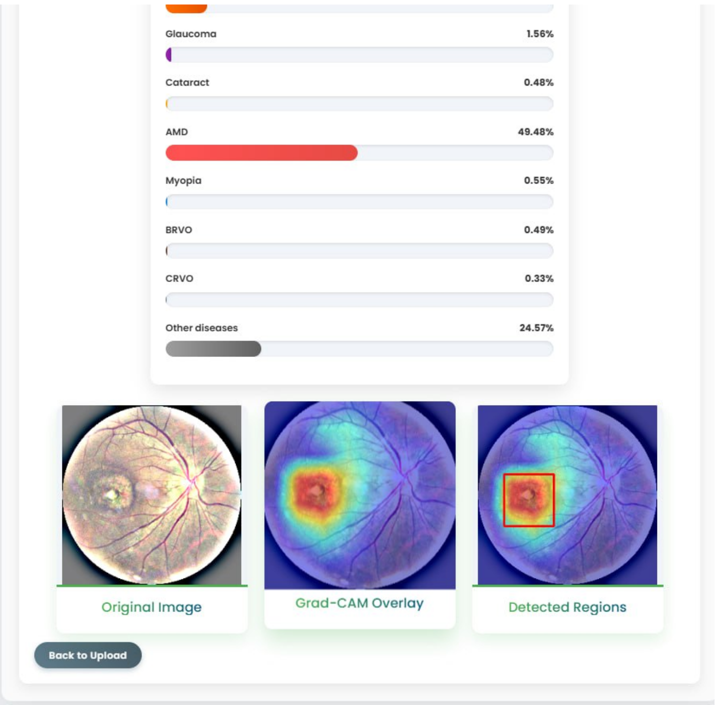
Eye Disease Detection

Prediction Result

Classification Confidence

Normal	11.80%
Diabetes	10.74%
Glaucoma	1.56%
Cataract	0.48%
AMD	49.48%
Myopia	0.55%
BRVO	0.49%
CRVO	0.33%
Other diseases	24.57%

Figure 3.15: Example of AMD detection percentages.



Glaucoma 1.56%

Cataract 0.48%

AMD 49.48%

Myopia 0.55%

BRVO 0.49%

CRVO 0.33%

Other diseases 24.57%

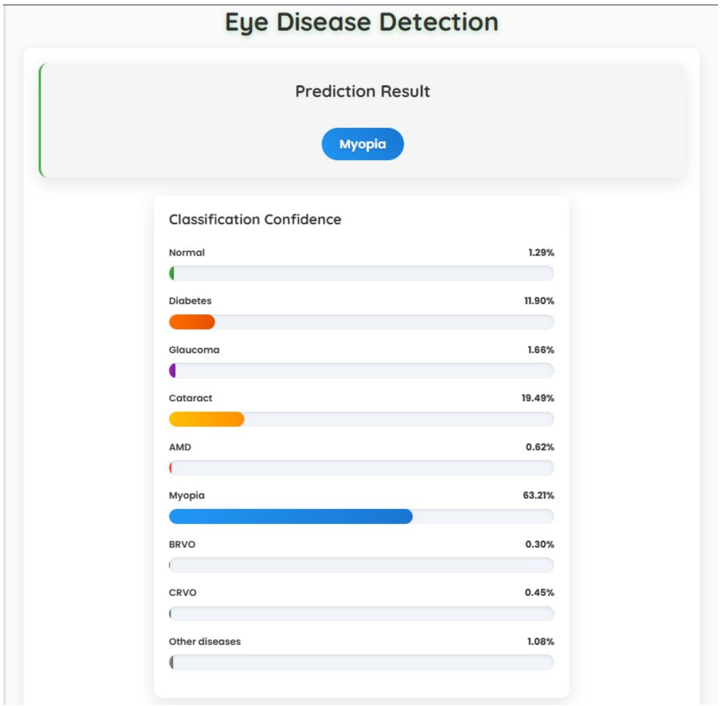
Original Image

Grad-CAM Overlay

Detected Regions

Back to Upload

Figure 3.16: Example of AMD detection visualization.



Eye Disease Detection

Prediction Result

Myopia

Classification Confidence

Normal 1.29%

Diabetes 11.90%

Glaucoma 1.66%

Cataract 19.49%

AMD 0.62%

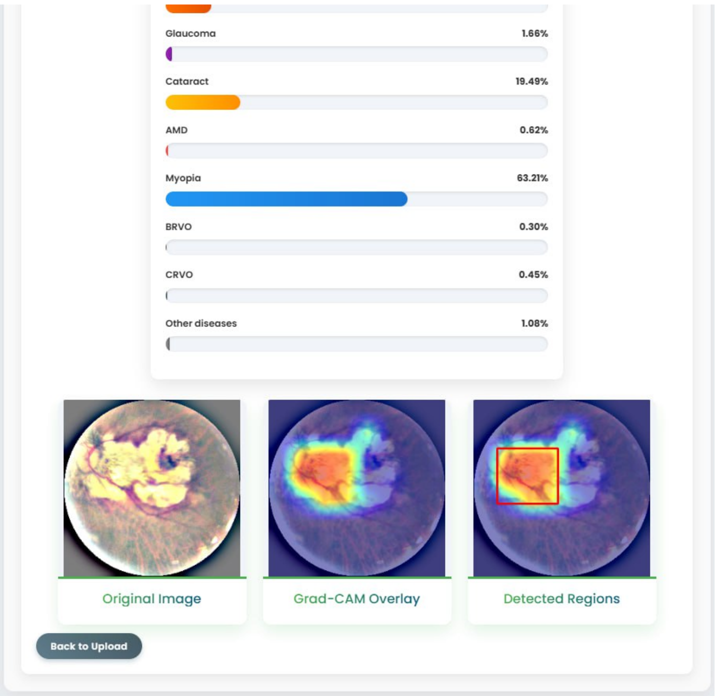
Myopia 63.21%

BRVO 0.30%

CRVO 0.45%

Other diseases 1.08%

Figure 3.17: Example of Myopia detection percentage.



Glaucoma 1.66%

Cataract 19.49%

AMD 0.62%

Myopia 63.21%

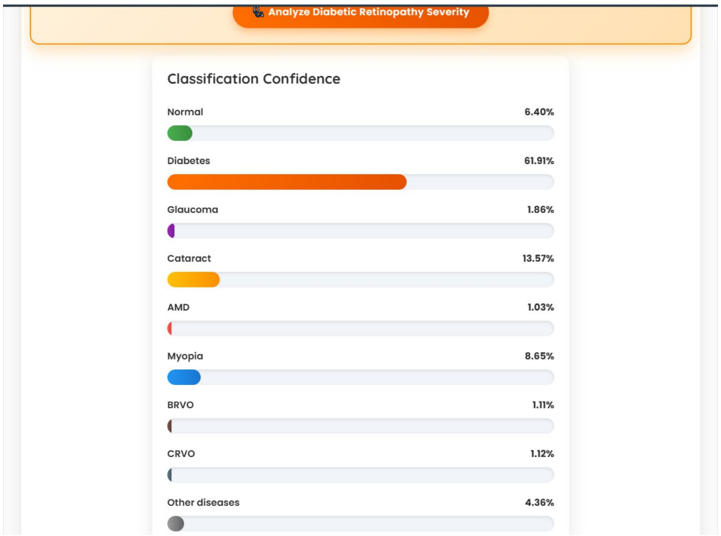
BRVO 0.30%

CRVO 0.45%

Other diseases 1.08%

Figure 3.18: Example of Myopia detection visualization.

2. If the predicted class is DR, **Stage 2** is triggered: the DR-staging model is called to determine the specific grade (0 to 4).



Classification Confidence

Normal 6.40%

Diabetes 61.91%

Glaucoma 1.86%

Cataract 13.57%

AMD 1.03%

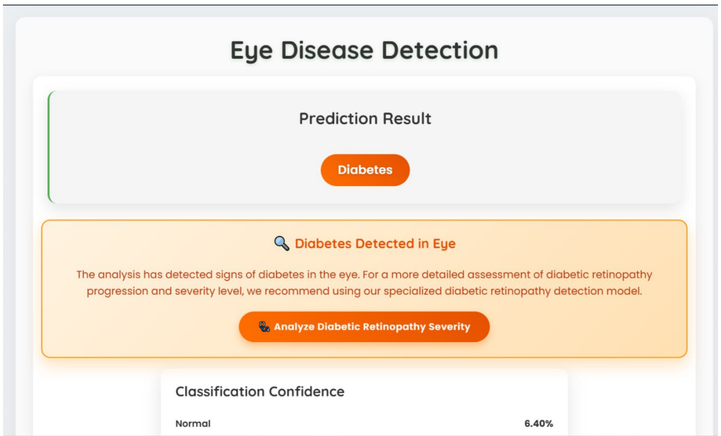
Myopia 8.65%

BRVO 1.11%

CRVO 1.12%

Other diseases 4.36%

Figure 3.19: Example of DR detection percentage.



Eye Disease Detection

Prediction Result

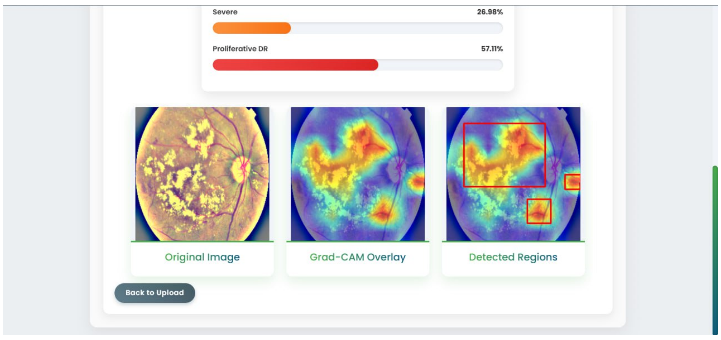
🔍 Diabetes Detected in Eye

The analysis has detected signs of diabetes in the eye. For a more detailed assessment of diabetic retinopathy progression and severity level, we recommend using our specialized diabetic retinopathy detection model.

Classification Confidence

Normal 6.40%

Figure 3.20: DR detection.



Severe 26.98%

Proliferative DR 57.11%

Figure 3.21: Example of DR detection visualization.

3. The platform displays the result and provides Grad-CAM visual explanations to indicate the model's focus area as shown .

This cascaded design ensures computational efficiency while maintaining diagnostic precision.

3.11.4 Implementation Details

The platform was built using a lightweight and scalable technology stack:

- **Frontend:** Developed using HTML and CSS within the Flask framework, providing a simple and interactive user interface
- **Backend:** Implemented in Python using the PyTorch library. Image preprocessing is performed using Ben's method and CLAHE to enhance image quality in alignment with the model's training conditions.

- **Hosting:**The platform runs locally and is accessed via an ngrok link, offering easy testing and demonstration without requiring full cloud deployment.

Both models were exported using TorchScript and loaded into the platform backend to ensure compatibility and performance. Image preprocessing (CLAHE and Ben's method) is applied at upload time to match the model training conditions.

3.11.4.1 Mods

Mode 1: Specializes in classifying only diabetic retinopathy (DR) severity levels, ranging from grade 0 (no disease) to grade 4 (proliferative DR).
Mode 2: Capable of classifying six different eye diseases, including diabetic retinopathy.

3.11.5 User Interface Features

The diagnostic interface is intuitive and designed for ease of use by non-technical users. It includes:

- A simple image upload panel.
- Real-time prediction with disease name and DR grade.
- Grad-CAM overlays for enhanced interpretability.
- Clear output text and probability/confidence levels.

3.11.6 Benefits and Limitations

This platform offers several practical advantages:

- Real-time, automated diagnosis based on state-of-the-art deep learning.
- Modular design that supports easy model updates and scalability.
- Potential for deployment in clinics, rural screening programs, or educational settings.

Current limitations:

- The system has not been clinically validated with real-world patient data.
- Model performance may degrade with poor-quality or out-of-distribution images.
- Integration with hospital databases .

3.12 Conclusion

This chapter highlighted the successful development and deployment of an AI-powered web-based diagnostic platform tailored for retinal disease detection. The system demonstrated significant advantages, including real-time automated analysis, a modular design for flexible updates, and potential applicability in diverse settings such as clinics, remote screening programs, and educational environments. The two-stage model architecture ensured diagnostic precision while optimizing computational efficiency.

Despite these strengths, certain limitations remain. The system has not yet undergone clinical validation with real-world patient data, which is essential for regulatory approval and clinical adoption. Additionally, the model's performance may be affected by poor-quality or out-of-distribution images, and integration with hospital databases is still under development.

Overall, this work represents a promising contribution toward practical, AI-based diagnostic solutions in ophthalmology, bridging the gap between research and clinical utility while laying the groundwork for future improvements and clinical integration.

Chapter 4

CONCLUSION

The thesis presented the design, development, and evaluation of a computer-based system for automatic detection of retinal diseases using fundus imaging. By leveraging modern deep learning techniques, we built a diagnostic framework capable of identifying multiple retinal conditions, including diabetic retinopathy, glaucoma, AMD, cataract, and others. The system was trained and tested on two widely recognized datasets—APTOS 2019 and ODIR 2019—using various preprocessing strategies to enhance image quality and model performance.

Among the models explored, SwinV2 achieved the most consistent and accurate results across both datasets. The model showed strong generalization for multi-disease classification and high sensitivity in diabetic retinopathy grading, particularly when supported by CLAHE and Ben's preprocessing techniques.

To demonstrate the practical use of our work, we deployed the best-performing models into a functional web-based platform. This platform provides an accessible interface for users to upload fundus images and receive real-time diagnostic predictions, with automatic staging when diabetic retinopathy is detected. The system was designed to be intuitive, modular, and adaptable for future extensions.

Overall, this work provides a complete diagnostic pipeline—from dataset preparation and model training to platform deployment—that can serve as a foundation for real-world medical screening tools. While the results are encouraging, further testing in clinical environments and integration with broader healthcare systems are recommended as next steps.

Bibliography

[1] C. Shorten and T. M. Khoshgoftaar.
A survey on image data augmentation for deep learning.
Journal of Big Data, 6(1):60, 2019.

[2] Ben Graham.
Kaggle diabetic retinopathy detection competition preprocessing methods, 2015.
Accessed: 2025-05-28.

[3] K. Zuiderveld.
Contrast limited adaptive histogram equalization.
In Graphics Gems IV, 1994.

[4] Zhang X. Ren S. He, K. and J. Sun.
Deep residual learning for image recognition.
In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[5] Cord M. Douze M. Touvron, H. and H. Jégou.
Training data-efficient image transformers and distillation through attention.
In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, volume 139, pages 10347–10357, 2021.

[6] Lin Y. Cao Y. Hu H. Liu, Z. and Y. Wei.
Swin transformer: Hierarchical vision transformer using shifted windows.
In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10012–10022, 2021.

[7] Fan H. Chen, L. and J. Wu.
Maxvit: Multi-axis vision transformer.
In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7580–7589, 2021.

[8] Tien Y. Wong et al.
Global prevalence and major risk factors of diabetic retinopathy.
The Lancet, 383(9933):2182–2193, 2014.

[9] International Diabetes Federation.
Idf diabetes atlas, 10th edition, 2021.
Accessed: 2025-05-14.

[10] M. T. Nouri, D. Hartani, M. Tiar, L. Chachoua, Y. Terfani, and S. Badji.
Enquête nationale sur les pathologies oculaires cécitantes en algérie en 2008.
Revue internationale du trachome et de pathologie oculaire tropicale et subtropicale et de santé publique, 87:141–163, 2010.
Conducted by MSPRH and INSP in collaboration with WHO.

[11] Fatima Bezzina and Karima Bereksi Reguig.
Diabetic retinopathy: Prevalence and risk factors in type 2 diabetic patients of sidi bel abbes, algeria.
South Asian Journal of Experimental Biology, 12(3):422–428, 2022.

[12] David M. Berson.
Phototransduction in ganglion-cell photoreceptors.
Pflügers Archiv, 454(5):849–855, 2007.

[13] Richard H. Masland.
The tasks of amacrine cells.
Visual Neuroscience, 29(1):3–9, 2012.

[14] American Academy of Ophthalmology.
Retina and Vitreous: Basic and Clinical Science Course.
American Academy of Ophthalmology, San Francisco, CA, 2023.
2023–2024 Edition.

[15] American Academy of Ophthalmology.
Fundamentals and Principles of Ophthalm.
American Academy of Ophthalmology, San Francisco, CA, 2019.

[16] Neil Patton, Tariq Aslam, Thomas MacGillivray, Ian J Deary, Baljean Dhillon, Robert H Eikelboom, Kanagasingam Yogesan, and Ian J Constable.
Retinal image analysis: Concepts, applications and potential.
Progress in Retinal and Eye Research, 25(1):99–127, 2006.

[17] Michael D Abràmoff, Meindert Niemeijer, M. S. A. Suttorp-Schulten, Max A Viergever, Stephen R Russell, and Bram van Ginneken.
Automated analysis of retinal images for detection of referable diabetic retinopathy.
JAMA Ophthalmology, 128(6):750–756, 2010.

[18] David Wong, Jiang Liu, Joo Hwee Lim, Fen Yin, Huiqi Li, Ning Tan, Carol Y Cheung, and Tien Yin Wong.
Computerized retinal image analysis to improve diagnosis of diabetic retinopathy: A review.
Medical Image Analysis, 15(1):169–196, 2011.

[19] Andrew Bastawrous, Marco E Giardini, Neil M Bolster, Tunde Peto, Neil Shah, Ian A T Livingstone, Heather A Weiss, Shirley Hu, Hillary Rono, Hannah Kuper, and Matthew J Burton.
Clinical validation of a smartphone-based adapter for optic disc imaging in kenya.
JAMA Ophthalmology, 134(2):151–158, 2016.

[20] Patrick J. Saine and Marshall E. Tyler.
Ophthalmic Photography: Retinal Photography, Angiography, and Electronic Imaging.
Butterworth-Heinemann Medical, 2nd edition, 2002.

[21] V. Kumar, A. Surve, D. Kumawat, B. Takkar, S. Azad, R. Chawla, D. Shroff, A. Arora, R. Singh, and P. Venkatesh.
Ultra-wide field retinal imaging: A wider clinical perspective.
Indian Journal of Ophthalmology, 69(4):824–835, Apr 2021.

[22] Stewart Muchuchuti and Serestina Viriri.
Retinal disease detection using deep learning techniques: A comprehensive review.
Journal of Imaging, 9(4):84, 2023.

[23] Heinrich Heimann, Ulrich Kellner, and Michael H. Foerster.
Atlas of Fundus Angiography.
Georg Thieme Verlag, Stuttgart, Germany, 2006.

[24] Yoshihiro Yonekawa, Joan W. Miller, and Ivana K. Kim.
Age-related macular degeneration: Advances in management and diagnosis.
Journal of Clinical Medicine, 4(2):343–359, 2015.

[25] National Eye Institute.

Facts about age-related macular degeneration.
nei.nih.gov/health/maculardegen/armd$_{f}acts$, 2018.
Accessed January 6, 2019.

[26] Robert N. Weinreb, Tin Aung, and Felipe A. Medeiros.
The pathophysiology and treatment of glaucoma: A review.
JAMA, 311(18):1901–1911, 05 2014.

[27] David J. Browning.
Pathophysiology of retinal vein occlusions.
In *Retinal Vein Occlusions*, pages 33–52. Springer, New York, 2012.

[28] Kyoko Ohno-Matsui, Pei-Chang Wu, Kenji Yamashiro, Kamonwan Vutipongsatorn, Yu-Yun Fang, Carol M G Cheung, Timothy Y Y Lai, Yasushi Ikuno, Sarah Y Cohen, and Richard F Spaide.
Pathologic myopia.
Nature Reviews Disease Primers, 5(1):1–20, 2019.

[29] Tetsuya Ueta, Shintaro Makino, Yuji Yamamoto, Hiroki Fukushima, Sayaka Yashiro, and Masayoshi Nagahara.
Pathologic myopia: an overview of the current understanding and interventions.
Global Health & Medicine, 2(3):151–155, Jun 2020.

[30] Aurélien Géron.
Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems.
O'Reilly Media, 2nd edition, 2019.

[31] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin.
Attention is all you need.
In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[32] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby.
An image is worth 16x16 words: Transformers for image recognition at scale.
In *International Conference on Learning Representations*, 2021.

[33] Aptos 2019 blindness detection.
Available online at: kaggle.com/c/aptos2019-blindness-detection, 2019.

[34] Andrew Mohammad.
Ocular disease recognition (odir-5k), 2019.
Accessed: 2025-05-25.

[35] C. Shorten and T. M. Khoshgoftaar.
A survey on image data augmentation for deep learning.
Journal of Big Data, 6(1):60, 2019.

[36] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, and É. Duchesnay.
Accuracy score — scikit-learn 1.2.2 documentation.
/scikit-learn, 2011.

[37] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra.
Grad-cam: Visual explanations from deep networks via gradient-based localization.
In *Proceedings of the IEEE International Conference on Computer Vision*, pages 618–626, 2017.

[38] J. Dinesh Bodapati, N. Veeranjaneyulu, Shaik N. Shareef, Saeed Hakak, Muhammad Bilal, P. K. R. Maddikunta, and Oksam Jo.
Blended multi-modal deep convnet features for diabetic retinopathy severity prediction, 2020.
arXiv preprint arXiv:2006.11357.

[39] B. Tymchenko, P. Marchenko, and D. Spodarets.
Deep learning approach to diabetic retinopathy detection.
arXiv preprint arXiv:2003.02261, 2020.

[40] Y. Fan, Y. Zhang, and X. Li.
Enhancing diabetic retinopathy classification using deep learning.
Journal of Healthcare Engineering, 2023:Article ID 10406759, 2023.
Article ID 10406759.