PEOPLE`S DEMOCRATIC REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
ECHAHID CHEIKH LARBI TEBESSI UNIVERSITY –TEBESSA-
EXACT SCIENCES NATURAL AND LIFE SCIENCES FACULTY
DEPARTMENT OF COMPUTER SCIENCE

MASTER THESIS

Domain: Mathematics and Computer Science
Field: Computer Science
Option: Systems and Multimedias

Automated Glaucoma Diagnosis from Fundus Images Using Deep Learning

Defended by:

Mohamed Ali AHMED SISTA

Supervised by:
Pr. Chawki DJEDDI
Dr. Yaakoub BOUALLEG

In front of the jury composed of:

Abdeljalil GATTAL	Professor	Echahid Cheikh Larbi Tebessi University	**President**
Lakhdar LAIMECHE	Professor	Echahid Cheikh Larbi Tebessi University	**Examiner**

Defended on **10/06/2025**
Note: ……. Mention: ……

Dedication

To my family,

whose unwavering love and infinite patience
anchored me through every challenge

To my beloved parents,

whose countless sacrifices and steadfast belief
in my potential shaped this journey

And to the countless sleepless nights—

faithful companions in my pursuit of knowledge—
may this work reflect the resilience they demanded
and the dreams they nurtured

Acknowledgments

I would like to express my deepest gratitude to all those who have contributed to the completion of this thesis.

First and foremost, I extend my sincere appreciation to my supervisor, **Prof. Chawki DJEDDI** and co-supervisor, **Dr. Yaakoub BOUALLEG**, for their meticulous guidance, exceptional technical expertise, and constant availability to address the numerous challenges encountered during this research.

I would also like to extends my gratitude to **Prof. Abdeljalil GATTAL**, president of the examination committee, and **Prof. Lakhdar LAIMECHE**, examiner, for their thorough evaluation, constructive critiques, and valuable recommendations. Their professional expertise and thoughtful feedback have greatly strengthened the final outcome of this work.

Finally, I am grateful to my family and friends for their unconditional support, understanding, and encouragement throughout this academic endeavor.

[AHMED SISTA Mohamed Ali]
[Juin 6th 2025]

Abstract

Glaucoma is a leading cause of irreversible blindness worldwide, often progressing without early symptoms and resulting in significant vision loss if left undiagnosed. Early detection is therefore critical, but current manual screening methods are time-consuming and require specialized expertise, limiting accessibility in many regions. This thesis presents a fully automated framework for glaucoma diagnosis using deep learning on retinal fundus images. The proposed method integrates three main phases: image enhancement with Contrast-Limited Adaptive Histogram Equalization (CLAHE), hierarchical segmentation of the optic disc and cup using a fine-tuned Segment Anything Model 2 (SAM2), and glaucoma classification based on neuroretinal rim analysis with a Vision Transformer (ViT). Experimental results on the ORIGA and REFUGE datasets demonstrate high segmentation accuracy, with Dice scores of 95.84% and 90.63% for the optic disc, and over 90% for the optic cup. For classification, the ViT model achieves an F1-score of 95.02% and a precision of 97.39% on the REFUGE dataset, outperforming several traditional convolutional neural networks. The framework is efficient, interpretable, and suitable for real-time clinical deployment. While specificity remains a challenge due to class imbalance and lack of data augmentation, the results highlight the potential of combining advanced segmentation and transformer-based classification for robust, scalable glaucoma screening. This work lays a strong foundation for future improvements and broader clinical application in automated eye disease diagnosis

Keywords: Glaucoma, Deep Learning, Fundus Images, Automated Diagnosis, Image Segmentation, Vision Transformer, SAM2, Neuroretinal Rim, Medical Image Analysis, Artificial Intelligence

Résumé

Le glaucome est l'une des principales causes de cécité irréversible dans le monde, progressant souvent sans symptômes précoces et entraînant une perte de vision importante en l'absence de diagnostic. La détection précoce est donc essentielle, mais les méthodes de dépistage manuel actuelles sont longues et nécessitent une expertise spécialisée, ce qui limite leur accessibilité dans de nombreuses régions. Cette thèse propose un nouveau cadre automatisé pour le diagnostic du glaucome à partir d'images du fond d'œil, basé sur l'apprentissage profond. La méthode développée s'articule en trois phases principales: l'amélioration des images par l'égalisation d'histogramme adaptative à contraste limité (CLAHE), la segmentation hiérarchique du disque optique et de la cupule à l'aide d'un modèle SAM2 ajusté, et la classification du glaucome par analyse de la région du bord neuro-rétinien grâce à un Vision Transformer (ViT). Les résultats expérimentaux sur les bases de données ORIGA et REFUGE montrent une grande précision de segmentation, avec des scores Dice de 95,84% et 90,63% pour le disque optique, et plus de 90% pour la cupule. Pour la classification, le modèle ViT atteint un score F1 de 95,02% et une précision de 97,39% sur REFUGE, surpassant plusieurs réseaux de neurones convolutifs classiques. Le cadre proposé est efficace, interprétable et adapté à un déploiement clinique en temps réel. Bien que la spécificité reste perfectible en raison du déséquilibre des classes et de l'absence d'augmentation de données, les résultats démontrent le potentiel de la combinaison d'une segmentation avancée et d'une classification par transformeur pour un dépistage robuste et évolutif du glaucome. Ce travail pose ainsi les bases pour de futures améliorations et une application clinique élargie dans le diagnostic automatisé des maladies oculaires.

Mots clés: Glaucome, Apprentissage profond, Fond d'œil, Diagnostic automatisé, Segmentation d'image, Vision Transformer, SAM2, Bord neuro-rétinien, Analyse d'images médicales, Intelligence artificielle

الملخص

الجلوكوما هي واحدة من الأسباب الرئيسية للعمى غير القابل للعلاج على مستوى العالم، حيث تتقدم غالبًا بدون أعراض مبكرة وتؤدي إلى فقدان كبير في الرؤية إذا لم يتم تشخيصها. لذلك، يعد الكشف المبكر أمرًا حاسمًا، لكن الطرق اليدوية الحالية للفحص تستغرق وقتًا طويلاً وتحتاج إلى خبرة متخصصة، مما يحد من إمكانية الوصول إليها في العديد من المناطق. تقدم هذه الأطروحة إطارًا آليًا بالكامل لتشخيص الجلوكوما باستخدام التعلم العميق على صور قاع العين. الطريقة المقترحة تدمج ثلاث مراحل رئيسية: تحسين الصورة باستخدام تقنية تسوية الهيستوجرام المتكيفة المحدودة التباين (CLAHE)، والتجزئة الهرمية للقرص البصري والكأس باستخدام نموذج Segment Anything Model 2 المُعد مسبقًا، وتصنيف الجلوكوما بناءً على تحليل الحافة العصبية الشبكية باستخدام معالج الرؤية Vision Transformer . أظهرت النتائج التجريبية على مجموعتي بيانات ORIGA و REFUGE دقة عالية في التجزئة، حيث بلغت درجات Dice 95.84٪ و90.63٪ للقرص البصري، وأكثر من 90٪ للكأس البصري. أما بالنسبة للتصنيف، حقق نموذج ViT درجة F1-score بلغت 95.02٪ ودقة 97.39٪ على مجموعة بيانات REFUGE، متفوقًا على عدة شبكات عصبية تلافيفية تقليدية. الإطار فعال، قابل للتفسير، مناسب للتطبيق السريري في الوقت الفعلي. وعلى الرغم من أن التحدي ما زال قائماً فيما يتعلق بالخصوصية بسبب عدم توازن الفئات ونقص زيادة البيانات، إلا أن النتائج تسلط الضوء على إمكانات الجمع بين تقنيات التجزئة المتقدمة والتصنيف المعتمد على المحولات لفحص الجلوكوما بشكل قوي وقابل للتوسع. يضع هذا العمل أساسًا قويًا للتحسينات المستقبلية والتطبيق السريري الأوسع في تشخيص أمراض العيون تلقائيًا.

الكلمات المفتاحية: الجلوكوما، التعلم العميق، صور قاع العين، التشخيص الآلي، تجزئة الصور، محول الرؤية، SAM2، الحافة العصبية الشبكية، تحليل الصور الطبية، الذكاء الاصطناعي.

List of Figures

1.1 Illustration of the human eye structure. 5
1.2 Goldmann applanation tonometer positioned on the slit lamp (A) with its cone prism (B) (on the top right); the two arcs appear correctly aligned (B) (on the bottom right) [1]. 13
1.3 Illustration of gonioscopy [2]. 13
1.4 Illustration of gonio prism [3]. 14

2.1 Fundus image highlighting glaucoma-relevant structures: optic disc (OD), cup (OC), and ISNT quadrants (inferior, superior, nasal, temporal)[4]. . . 17
2.2 Ultrahigh-resolution ophthalmic Optical Coherence Tomography image [5]. 18
2.3 Illustration of fundus artificial intelligence-based diagnosis: (a) Machine Learning-based, (b) Deep Learning-based [6]. 19
2.4 Fundus image with the region of interest around the optic disc; (a) Ocular Fundus Image, (b) Region of interest extracted from an ocular fundus image. [7]. 20
2.5 Convolutional Neural Network internal layers [8]. 22
2.6 Overview of the Vision Transformer model [9]. 24
2.7 Illustration of the segment anything model 2 architecture [10]. 27
2.8 Key challenges in medical AI and how they relate to one another and clinical care. Node color represents the type of challenge: conceptual (orange), technical (green), or humanistic (pink)[11]. 28
2.9 How accredited expert groups–developers, validators, and operational staff–can help overcome the key challenges in medical artificial intelligence [11]. 29

3.1 Example of SLIC image (a) and Nomalized graph cut image (b)[12]. . . . 32
3.2 Overview of the proposed SLSR-Net model. 34
3.3 illustration of framework proposed by Govindharaj et al. [13]. 40
3.4 Illustration of the method proposed by Naidana et al. [14] 41

4.1 Illustration of the proposed glaucoma classification framework. 47
4.2 Illustration of contrast enhancement using Contrast Limited Adaptive Histogram Equalization technique on ORIGA dataset fundus image, normal (a) and enhanced (b) 48

4.3 Optic disc segmentation result from finetuned Segment Anything Model 2 (SAM2), (a) input image, (b) output optic disk region. 50
4.4 Optic cup segmentation result from finetuned Segment Anything Model 2 (SAM2), (a) input optic disk image, (b) output optic cup region. 51
4.5 Illustration of rim area image. 51
4.6 Sample images with optic disk/cup groundtruth masks from REFUGE fundus images dataset. 53
4.7 Random samples from the ORIGA fundus image dataset for glaucoma analysis. 54
4.8 Fine-tuning performances of Segment Anything Model 2 on the ORIGA dataset for optic disc segmentation. 57
4.9 Fine-tuning performances of Segment Anything Model 2 on the ORIGA dataset for optic cup segmentation. 58
4.10 Fine-tuning performances of Segment Anything Model 2 on the REFUGE dataset for optic disc segmentation. 59
4.11 Fine-tuning performances of Segment Anything Model 2 on the REFUGE dataset for optic cup segmentation. 60
4.12 Finetuning performances of Vision Transformer model on REFUGE dataset. 63
4.13 Confusion Matrix of the REFUGE dataset test set: (a) without optic disc/cup segmentation and (b) with optic disc/cup segmentation 63

List of Tables

2.1 Comparison of conventional machine learning and deep learning. 21
2.2 Key Differences Between Convolutional Neural Networks and Vision Transformers. 25

3.1 Limitations of state-of-the-art classification methods. 38
3.2 Limitations of state-of-the-art classification methods. 44

4.1 Summary of ORIGA and REFUGE Datasets 54
4.2 Fine-tuning Hyperparameters for SAM2 Optic Disc and Cup Segmentation 57
4.3 Segmentation performance comparison on ORIGA dataset 58
4.4 Segmentation performance comparison on REFUGE dataset 61
4.5 Fine-tuning hyperparameters for Vision Transformer model 62
4.6 Classification results on the REFUGE dataset 64

Contents

Abstract **iv**

List of Figures **v**

List of Tables **vi**

General Introduction **1**

1 Human Vision and Medical Glaucoma Disease **4**
1.1 Anatomy of the Human Eye and Visual Pathway 4
1.1.1 Retina 6
1.1.2 Optic Nerve 6
1.1.3 Macula 6
1.1.4 Blood Vessels 7
1.1.5 Optic Nerve Head 7
1.1.6 Visual Pathway 7
1.2 Background on Glaucoma Disease 8
1.2.1 Types of Glaucoma 8
1.2.1.1 Open Angle Glaucoma 8
1.2.1.2 Angle-closure glaucoma 9
1.2.1.3 Congenital glaucoma 9
1.2.2 Epidemiology, Pathophysiology, and Risk Factors of Glaucoma . . 10
1.2.2.1 Epidemiology 10
1.2.2.2 Pathophysiology 10
1.2.2.3 Risk Factors 11
1.2.3 Current Diagnostic Techniques 11
1.2.3.1 Comprehensive Eye Examination 12
1.2.3.2 Intraocular Pressure Measurement 12
1.2.3.3 Central Corneal Thickness Assessment 12
1.2.3.4 Anterior Chamber Angle Evaluation 12
1.2.3.5 Optic Nerve and Retinal Nerve Fiber Layer Assessment . 14

2 Artificial Intelligence and Ophthalmology **16**
2.1 Medical Imaging Technologies in Ophthalmology . . . 16
2.1.1 Fundus Photography . . . 17
2.1.2 Optical Coherence Tomography . . . 18
2.2 Role of Artificial Intelligence in Glaucoma Diagnosis . . . 18
2.3 Deep Learning Basics . . . 20
2.3.1 Machine Learning vs Deep Learning . . . 20
2.3.2 Convolutional Neural Networks . . . 21
2.3.3 Attention mechanism . . . 22
2.3.4 Vision Transformers . . . 23
2.3.5 Foundation Models . . . 26
2.4 Challenges in medical AI . . . 27

3 State of The Art **31**
3.1 Fundus Image Segmentation . . . 31
3.2 Fundus Image Classification . . . 39

4 Contribution and Experimental Study **46**
4.1 Proposed Glaucoma Diagnosis Framework . . . 47
4.1.1 Phase 1: Pre-Processing . . . 47
4.1.2 Phase 2: Optic Disc And Optic Cup Segmentation . . . 49
4.1.2.1 Fine-tuning Segment Anything Model 2 for Optic Disc Segmentation . . . 49
4.1.2.2 Fine-tuning Segment Anything Model 2 For Optic Cup Segmentation . . . 50
4.1.3 Phase 3: Glaucoma classification . . . 51
4.2 Experimental Study . . . 52
4.2.1 System Configuration . . . 52
4.2.2 Datasets . . . 52
4.2.2.1 REFUGE Dataset . . . 53
4.2.2.2 ORIGA Dataset: . . . 53
4.2.3 Evaluation Metrics . . . 54
4.2.3.1 Segmentation Metrics . . . 54
4.2.3.2 Classification Metrics . . . 55
4.2.4 Experimental Results . . . 56
4.2.4.1 Optic Disc and Optic Cup Segmentation . . . 56
4.2.4.2 Glaucoma Classification . . . 62
4.3 Conclusion . . . 65

General Conclusion **67**

Bibliographie **76**

General Introduction

Glaucoma is a silent and progressive disease that damages the optic nerve, often without early warning signs [15]. It is a leading cause of irreversible blindness worldwide, affecting more than 70 million people by 2020, with projections exceeding 112 million by 2040 [16]. The disease is especially concerning because it progresses without symptoms until significant vision loss has occurred. This makes early detection both difficult and essential. The prevalence among adults aged 40–80 is about 3.5%, but the risk is higher in certain populations, such as those of African, Asian, or Hispanic descent [17]. Furthermore, limited access to healthcare in many regions means that many cases remain undiagnosed and untreated, resulting in preventable blindness [18].

The impact of glaucoma extends beyond the clinic. It imposes heavy social and economic burdens on individuals, families, and health systems, especially as populations age. While high intraocular pressure (IOP) is a well-known risk factor, other elements such as genetics, oxidative stress, and vascular health also play a role in the disease's development and progression [19]. These complexities make early and accurate diagnosis a major challenge for clinicians.

Traditionally, glaucoma diagnosis relies on manual screening by ophthalmologists. This process involves careful examination of the optic nerve head, measurement of IOP, and visual field testing. However, these methods are time-consuming, require specialist expertise, and are subject to human error and variability. As a result, there is a clear need for automated, objective, and scalable diagnostic tools that can assist clinicians and reach underserved populations.

Recent advances in artificial intelligence (AI), especially deep learning (DL), have opened new possibilities for automated glaucoma diagnosis. Deep learning models, such as convolutional neural networks (CNNs) and vision transformers (ViTs), have shown great promise in analyzing medical images including retinal fundus photographs and optical coherence tomography (OCT) scans to detect subtle changes associated with glaucoma. These models can extract and interpret complex patterns, such as changes in the cup-to-disc ratio and thinning of the neuroretinal rim, which are key biomarkers for the disease.

Despite their potential, AI-based approaches face several challenges. First, the quality and quantity of annotated medical images are often limited, which can hinder model training and generalization. Second, many deep learning models are seen as "black boxes," making it difficult for clinicians to understand or trust their decisions. This lack of interpretability is a significant barrier to clinical adoption. Third, automated systems must be robust to variations in image quality, patient demographics, and disease presentation, which are common in real-world settings. Finally, the integration of these tools into clinical workflows requires careful validation and regulatory approval.

This thesis aims to address these challenges by developing a novel, automated framework for glaucoma diagnosis from retinal fundus images using deep learning. The proposed method is designed to be both accurate and efficient, with a focus on clinical relevance and interpretability. The framework consists of three main phases:

- **Pre-processing :** The input fundus images are enhanced using Contrast-Limited Adaptive Histogram Equalization (CLAHE). This technique improves local contrast, making subtle anatomical features such as the optic disc and cup more visible for analysis.
- **Segmentation:** A hierarchical, two-step segmentation approach is used, leveraging a fine-tuned Segment Anything Model 2 (SAM2). First, the optic disc is segmented from the enhanced image. Then, within the segmented disc, the optic cup is isolated. This sequential process ensures anatomical accuracy and reduces background noise, focusing the analysis on clinically relevant regions.
- **Classification:** The neuroretinal rim area, defined as the region between the segmented disc and cup, is extracted and resized. A Vision Transformer (ViT) model is then fine-tuned on these rim images to classify each case as glaucomatous or healthy. The use of ViT allows the model to capture global context and subtle structural changes, which are critical for early detection.

The key innovations of this work include the two-step segmentation strategy, which prioritizes anatomical accuracy, and the use of feature fusion based on disc-cup relationships. The optimized SAM2 architecture balances the demands of medical imaging with computational efficiency, enabling real-time inference suitable for clinical use. By focusing on interpretable biomarkers and leveraging state-of-the-art deep learning techniques, the framework aims to provide a practical solution for scalable glaucoma screening.

To validate the proposed approach, extensive experiments are conducted on two widely used datasets: ORIGA and REFUGE. These datasets provide annotated fundus images with ground truth for both segmentation and classification tasks, allowing for rigorous evaluation and comparison with existing methods. The performance is assessed using standard metrics such as Dice coefficient, Intersection over Union (IoU), accuracy, precision, recall, F1-score, specificity, and area under the ROC curve (AUC-ROC).

The results demonstrate that the proposed framework achieves high segmentation accuracy for both optic disc and cup, with Dice scores exceeding 90% on both datasets. For classification, the fine-tuned ViT model attains competitive precision and F1-scores, outperforming several traditional CNN-based approaches. Importantly, the method achieves these results without data augmentation, highlighting its robustness and efficiency.

Nevertheless, certain limitations remain. The specificity of the classification model is lower than its sensitivity, indicating challenges in correctly identifying healthy cases. This is likely due to the class imbalance in the datasets and the absence of data augmentation, which could improve generalization. Future work will explore targeted augmentation strategies and hybrid model architectures to address these issues.

The structure of this thesis is as follows:

- **Chapter 1 :** provides a medical background, covering eye anatomy, types of glaucoma, and current diagnostic methods.
- **Chapter 2 :** discusses the technological foundations of AI, deep learning, and vision transformers, with a focus on their application in medical imaging.

- **Chapter 3 :** reviews the state-of-the-art in automated glaucoma diagnosis, including both segmentation and classification approaches.
- **Chapter 4 :** presents the proposed framework in detail, describes the experimental setup, and reports the results of the evaluation.

  The thesis concludes with a summary of the main findings and suggestions for future research directions.

Chapter 1

Human Vision and Medical Glaucoma Disease

Introduction

Glaucoma is a major reason people lose their sight permanently around the world. Significantly, this disease often begins without any clear signs. It slowly damages the optic nerve before a person notices any changes in their vision. Consequently, this chapter gives a basic medical summary of glaucoma. This summary will help in understanding the disease more deeply.

This chapter looks into human vision and provides medical details about glaucoma. To begin, we will discuss the structure of the human eye. In particular, we will examine the retina and the optic nerve, as these are crucial for sight . Then, we will focus on the optic nerve head (ONH). The ONH is a critical area when doctors diagnose glaucoma. Clinicians often assess the ONH by examining the cup-to-disc ratio (CDR). This CDR is a crucial clue for detecting glaucoma.

Next, we will examine how glaucoma affects the eye and its prevalence. This includes describing how the optic nerve head gets damaged. It also explains why it is hard to detect glaucoma early. Lastly, the chapter will outline the current diagnostic methods and clinical tests that doctors use to detect glaucoma. This information provides a necessary background for later discussions.

1.1 Anatomy of the Human Eye and Visual Pathway

The human eye is a vital sensory organ that enables individuals to perceive and interpret their surroundings. Located within the protective bony orbit of the skull, the eye is

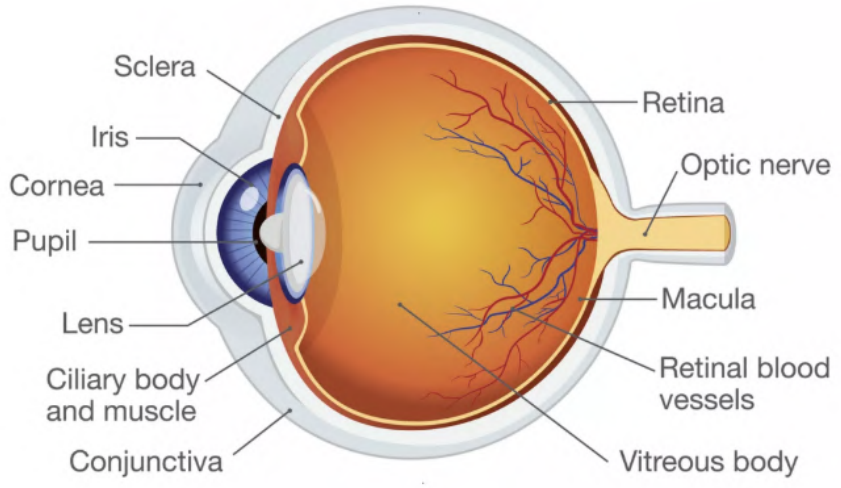


Figure 1.1: Illustration of the human eye structure.

a complex, spherical structure composed of multiple internal and external layers, each contributing to its visual function. Its detailed anatomy supports a wide range of visual processes, including light detection, color perception, and spatial awareness. A schematic illustration of the eye's anatomical features is shown in Figure 1.1, which provides a global overview of its main components.

Externally, the sclera forms a tough, outer layer. This layer maintains the eye's shape and protects its contents. At the front, the cornea is a clear, dome-like surface. It directs incoming light into the eye. Then, the iris, which is the colored part of the eye, surrounds the pupil. The iris adjusts the pupil's size to control how much light enters. Behind the pupil, the lens changes its curve. This change helps to focus light onto the retina.

Inside the eye, the retina is full of special cells called photoreceptors, which are rods and cones. These cells convert light into electrical signals. Subsequently, these signals travel to the brain through the optic nerve[20]. The macula is at the center of the retina. It contains the fovea, which is key for sharp, central vision. Furthermore, the vitreous humor, a clear gel, fills the space between the lens and retina. This gel helps the eye keep its shape. Beneath the retina lies the choroid. The choroid is a layer with many blood vessels. It brings nutrients to the retinal tissues and absorbs extra light, which reduces reflections inside the eye [21].

Therefore, understanding the eye's structure and function is very important. It is crucial for diagnosing and managing eye diseases like glaucoma. For instance, increased pressure inside the eye can harm the optic nerve. This damage can lead to vision loss, highlighting how important a healthy retina is for good vision [22].

1.1.1 Retina

The retina is a layered tissue of nerves. It lines the back part of the eye and is vital for vision. Essentially, it performs phototransduction. This means it changes light into electrical signals. It does this using a network of special neurons. These neurons include photoreceptors (rods and cones), bipolar cells, horizontal cells, amacrine cells, and ganglion cells. These are all organized into clear layers. Specifically, rods are responsible for vision in low-light conditions (scotopic vision), while cones support color vision and sharp vision in bright light (photopic vision). There are three types of cones. Each type is sensitive to different light wavelengths: red, green, or blue. This allows us to see colors. Cones are most crowded in the fovea, where our vision is sharpest [23, 24].

1.1.2 Optic Nerve

The optic nerve is a crucial part of our visual system. It is made of about 1.2 million axons. These axons come from retinal ganglion cells (RGCs). They all come together at the optic disc. The optic disc is an area on the retina that has no photoreceptors. The optic nerve itself has four parts: intraocular, intraorbital, intracanalicular, and intracranial. Inside the nerve, there are also supporting structures. These include glial cells, connective tissue, and blood vessels. These blood vessels, fed by the central retinal artery and ciliary arteries, provide stability, nutrients, and help send signals efficiently [25].

1.1.3 Macula

The macula is the central area of the retina responsible for high-acuity and color vision. At its core lies the fovea, which contains the greatest concentration of cone photoreceptors but lacks retinal ganglion cells (RGCs). This unique arrangement minimizes light distortion and enhances visual clarity. Around the fovea, the parafoveal region houses the highest density of RGCs, many of which contribute to the papillomacular bundle—a major component of the optic nerve. Although the fovea contains few RGCs, it has strong representation in the visual cortex, emphasizing its role in fine visual detail processing. The organization of the macula—concentrated cones centrally and displaced RGCs peripherally—supports precise central vision necessary for tasks like reading. Damage to the macula significantly impairs visual function, highlighting its importance in daily life [23, 24].

1.1.4 Blood Vessels

The eye's blood supply is primarily maintained by two vascular systems: the retinal and choroidal circulations. The inner layers of the retina, especially the nerve fiber and ganglion cell layers, receive nourishment from the central retinal artery and its branches, such as the radial peripapillary capillaries. In contrast, the outer retina, retinal pigment epithelium, and optic nerve head are supplied by the choroidal circulation, which is largely fed by the short posterior ciliary arteries. Within the choroid, the choriocapillaris forms a dense and highly permeable network of vessels that facilitates efficient nutrient and waste exchange. Deoxygenated blood is collected by the vorticose veins, which drain near the equatorial region of the eye into the systemic venous system.

Preserving the structural and functional integrity of these vascular systems, along with the blood-retinal barrier—formed by endothelial cells and stabilized by glial cells—is crucial for healthy retinal function. Damage or dysfunction in these vascular components can lead to major ocular diseases, including glaucoma, diabetic retinopathy, and age-related macular degeneration [23, 24].

1.1.5 Optic Nerve Head

The optic nerve head (ONH), also called the optic disc, is the place where about 1.2 million retinal ganglion cell (RGC) axons leave the eye and form the optic nerve [25]. This area is located in the nasal part of the retina. Importantly, the ONH does not have any photoreceptors, which means it creates a natural blind spot in our visual field. The ONH is made up of several layers: the nerve fiber layer at the surface, the prelaminar zone, the lamina cribrosa, and the retrolaminar area. Each layer plays a key role in supporting RGC axons and responding to mechanical stress [25, 26].

In clinical observation, doctors examine the ONH, they focus on a central pale area called the cup. This cup is surrounded by a pinkish neuroretinal rim. The size of the cup compared to the whole disc (called the cup-to-disc ratio (CDR)) is very important for diagnosing ocular diseases. If the pressure inside the eye gets too high, it can cause changes at the ONH, especially in the lamina cribrosa. The lamina cribrosa is a porous tissue that supports the axons. These changes make the ONH a main site for axonal injury in glaucoma. This is why the ONH is a major focus for imaging technologies used in eye diagnosis and monitoring.

1.1.6 Visual Pathway

The visual pathway begins when photoreceptors respond to light by initiating biochemical cascades that alter neurotransmitter release. These signals are transmitted through

bipolar cells, with modulation by horizontal and amacrine cells, which enhance contrast, motion sensitivity, and temporal processing. Retinal ganglion cells then integrate these inputs into center-surround receptive fields, selectively encoding light increments (ON pathways) and decrements (OFF pathways). The axons of these ganglion cells form the optic nerve, which carries processed visual information to the thalamus. The macula, densely populated with cone photoreceptors, is essential for high-acuity vision, while the optic disc serves as the exit point for ganglion cell axons. Damage to these structures—such as the loss of ganglion cells in glaucoma—disrupts visual signaling, highlighting their critical role in both vision and clinical diagnosis [23, 24].

1.2 Background on Glaucoma Disease

Glaucoma refers to a group of eye diseases that cause progressive damage to the optic nerve. This damage is crucial to vision, as it can lead to vision loss and, if left untreated, permanent blindness. Although elevated intraocular pressure (IOP) is a major risk factor [22], glaucoma can also occur with normal IOP, which complicates its diagnosis and management [27, 18]. The disease typically leads to structural changes in the optic disc, such as increased cupping and thinning of the retinal nerve fiber layer (RNFL). These changes correlate with progressive visual field loss. If left untreated, glaucoma can advance to total blindness [19, 28], highlighting the importance of early detection and treatment.

1.2.1 Types of Glaucoma

Glaucoma is classified mainly by the status of the anterior chamber angle, its underlying cause, and its progression pattern. The two main types are open-angle glaucoma and angle-closure glaucoma. Other forms, such as congenital glaucoma, are less common but also important [28].

1.2.1.1 Open Angle Glaucoma

Open-angle glaucoma is the most common type of glaucoma worldwide [29, 30, 18]. In this condition, the anterior chamber angle appears normal on examination. However, the outflow of aqueous humor is reduced, which often leads to elevated intraocular pressure (IOP) and gradual damage to the optic nerve [28, 29, 27]. The most prevalent subtype is primary open-angle glaucoma (POAG), which occurs in the absence of any identifiable secondary cause [28, 27, 18]. A variant of POAG is normal-tension glaucoma (NTG), in which optic nerve damage progresses despite IOP levels that fall within the normal statistical range [31]. In addition, secondary forms of open-angle glaucoma can arise

from specific conditions such as pigment dispersion syndrome, pseudoexfoliation, uveitis, or prior ocular trauma [28, 30].

In open-angle glaucoma, visual loss tends to be gradual and bilateral, but may be more pronounced in one eye. Common signs and symptoms include

- Gradual loss of peripheral vision, often affecting both eyes.
- Reduced night vision and difficulty seeing in low light.
- Impaired contrast sensitivity, especially noticeable in dim settings.
- Slowed reading speed or compensatory head movements to navigate visual field loss.
- Early deficits in blue-yellow color vision, though red-green perception is usually preserved until later stages.

1.2.1.2 Angle-closure glaucoma

Angle-closure glaucoma is characterized by a narrowed or closed anterior chamber angle, restricting aqueous humor drainage and can result in acute or chronic increases in IOP [28, 32]. Acute angle-closure glaucoma is a medical emergency, marked by a sudden blockage of aqueous outflow and a rapid rise in IOP, potentially causing irreversible vision loss if not treated promptly [28]. Primary angle-closure glaucoma (PACG) typically results from mechanisms such as pupillary block or anatomical variations like plateau iris syndrome [28]. Secondary causes of angle closure include neovascularization of the angle or lens-related factors, such as phacomorphic glaucoma due to cataract progression [28].

In angle-closure glaucoma, symptoms arise suddenly and may include blurred vision, severe eye pain, redness, halos around lights, headache, nausea, and vomiting—often linked to a rapid rise in intraocular pressure.

1.2.1.3 Congenital glaucoma

Congenital glaucoma, also referred to as primary congenital glaucoma (PCG), is a rare disorder that presents at birth or during early childhood [29, 30]. It arises from developmental abnormalities in the anterior segment of the eye, which impair the normal drainage of aqueous humor and lead to elevated IOP. Without timely diagnosis and treatment, the condition can result in significant visual impairment [29, 30].

1.2.2 Epidemiology, Pathophysiology, and Risk Factors of Glaucoma

Epidemiology refers to the frequency with which a disease occurs in different groups of people and helps us understand who is the most likely to be affected. Pathophysiology explains the biological changes that lead to the disease and how it harms the body. In the case of glaucoma, this involves damage to the optic nerve, often linked with increased eye pressure. Risk factors are characteristics or conditions that make someone more likely to develop a disease. Understanding these three areas: how common glaucoma is; how it develops; and what increases the chance of getting it; is essential for early detection, treatment planning, and public health efforts. Together, they form a foundation for both clinical practice and further research into preventing vision loss due to glaucoma.

1.2.2.1 Epidemiology

Glaucoma is a major public health concern and the second leading cause of blindness globally, after cataracts [15]. It is also the most common cause of irreversible vision loss worldwide. By 2020, over 70 million people had glaucoma, with projections reaching nearly 112 million by 2040 [16]. However, these figures may underestimate the true burden, especially in low-income regions where diagnosis rates are low and access to care is limited [17]. The disease mainly affects adults aged 40 to 80, with an overall prevalence of about 3.5% in this group.

In the U.S., around 2.2 million people have open-angle glaucoma, though nearly half of these cases remain undiagnosed due to its asymptomatic early stages [18]. Age is a key risk factor, as the likelihood of developing glaucoma rises sharply with advancing years, making it closely linked to aging populations. Geographic and ethnic differences also influence its occurrence. Primary open-angle glaucoma (POAG) is more common in North America and Europe, while primary angle-closure glaucoma (PACG) is more frequent in parts of Asia, such as Mongolia, Singapore, China, and India [28]. These disparities suggest the need for targeted screening strategies and healthcare planning tailored to specific populations.

1.2.2.2 Pathophysiology

Glaucoma is a progressive optic neuropathy characterized by the degeneration of retinal ganglion cells (RGCs) and their axons. This leads to structural changes such as optic disc cupping, thinning of the retinal nerve fiber layer (RNFL), and permanent visual field loss. In many cases, the disease involves impaired drainage of aqueous humor, often due to increased resistance in the trabecular meshwork or blockage at the anterior chamber angle. This results in elevated intraocular pressure (IOP), which exerts mechanical

stress on the optic nerve head (ONH). This stress disrupts axonal transport and triggers apoptosis of RGCs.

Beyond mechanical effects, the vascular theory of glaucoma suggests that high IOP can reduce blood flow to the ONH, causing hypoxia and ischemic damage. However, other IOP-independent mechanisms also contribute to the neurodegenerative process. These include oxidative stress, inflammation, mitochondrial dysfunction, deficiency in neurotrophic factors, and broader neurodegenerative changes along the central visual pathway [19, 33, 18].

1.2.2.3 Risk Factors

Several risk factors for glaucoma have been identified, grouped into demographic, clinical, genetic, and ocular categories. Age is a major non-modifiable risk factor, with individuals aged 60 years or older being at significantly higher risk [17]. Ethnicity also plays a role; people of African, Asian, or Hispanic descent tend to have higher prevalence rates and more severe forms of the disease. For instance, those of African or Caribbean origin often develop primary open-angle glaucoma (POAG) earlier and more aggressively than other groups [17].

Family history is another key factor, showing a strong hereditary influence. Having a first-degree relative with glaucoma greatly increases one's risk, highlighting the need for routine eye exams in such families. Clinically, elevated intraocular pressure (IOP) remains the only modifiable risk factor so far. However, high IOP does not always lead to glaucoma, as some patients develop the condition even with normal IOP levels, especially in normal-tension glaucoma.

Ocular features such as thin central corneal thickness (CCT), optic disc asymmetry, and an increased cup-to-disc ratio are linked to higher risk. In primary angle-closure glaucoma (PACG), additional risk factors include female gender, Asian ethnicity, hypermetropia, and shallow anterior chambers. Corticosteroid use either (topical or systemic) can also trigger glaucoma in genetically predisposed individuals.

Recent studies suggest that environmental stressors, including oxidative stress, may contribute to the development and progression of glaucoma [19]. While these findings are still emerging, they provide valuable insights into the disease's complex biological mechanisms.

1.2.3 Current Diagnostic Techniques

The diagnosis of glaucoma begins with a detailed patient history. Family history of glaucoma increases risk [19, 28]. For diagnosed patients, prior intraocular pressure

(IOP) levels and disease onset help assess optic nerve vulnerability. General medical conditions such as diabetes mellitus, which may lead to neovascular glaucoma—and systemic medications like steroids or beta-blockers are also relevant. Past ocular events, including trauma or surgery, can further influence IOP and glaucoma risk [19, 28].

1.2.3.1 Comprehensive Eye Examination

A comprehensive eye examination (CEE) includes visual acuity testing, pupillary response evaluation, and slit lamp biomicroscopy [34]. Visual acuity remains largely unaffected until late stages of glaucoma. A relative afferent pupillary defect (RAPD), often detected via the swinging flashlight test, may indicate unilateral or asymmetric optic nerve damage. Slit lamp biomicroscopy allows clinicians to examine anterior segment structures, identifying signs such as exfoliation material that may contribute to elevated IOP [34].

1.2.3.2 Intraocular Pressure Measurement

Elevated IOP is a major risk factor for glaucoma [19, 28, 34]. Goldmann applanation tonometry is the gold standard for IOP measurement, assessing the force needed to flatten a defined corneal area, Figure 1.2 showcases a Goldmann applanation tonometer . Rebound tonometers, such as iCare, offer comparable results but may be less reliable at higher IOPs [28, 34]. Interpretation of IOP should consider central corneal thickness (CCT) and optic nerve status rather than relying solely on standard ranges (11–21 mmHg). Diurnal variation in IOP suggests the need for multiple readings throughout the day [28, 34].

1.2.3.3 Central Corneal Thickness Assessment

Central corneal thickness (CCT) is an independent risk factor for progression from ocular hypertension to primary open-angle glaucoma (POAG) [34]. CCT is typically measured using pachymetry, with an average of approximately 550 μm [34]. Thinner corneas may underestimate IOP, and thicker ones may overestimate it when using Goldmann tonometry. However, correction formulas have not consistently improved predictive accuracy, so CCT is used alongside unadjusted IOP values [34].

1.2.3.4 Anterior Chamber Angle Evaluation

Determining whether the anterior chamber angle is open or closed is essential for glaucoma classification [34]. Gonioscopy, the gold standard, uses a prismatic lens to visualize angle

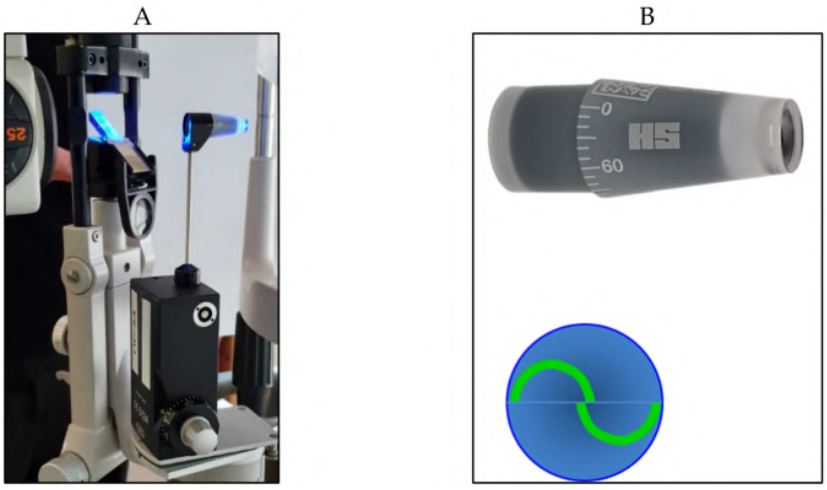


Figure 1.2: Goldmann applanation tonometer positioned on the slit lamp (A) with its cone prism (B) (on the top right); the two arcs appear correctly aligned (B) (on the bottom right) [1].

structures obscured by internal reflection. It identifies anatomical abnormalities such as pigment deposition or angle closure. Advanced imaging techniques—such as swept-source optical coherence tomography (OCT) and ultrasound biomicroscopy (UBM)—offer objective measurements and supplement traditional gonioscopy [28, 34].

Figure 1.3 provides a general view on the gonioscopy exam, while Figure 1.4 showcases the gonioprism, also known as gonioscopy prismatic lenses.

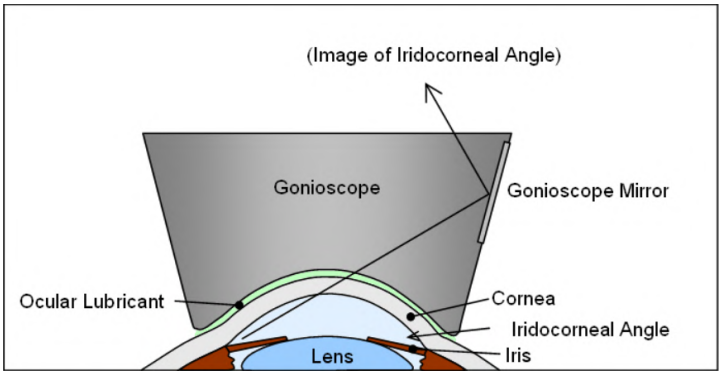


Figure 1.3: Illustration of gonioscopy [2].



Figure 1.4: Illustration of gonio prism [3].

1.2.3.5 Optic Nerve and Retinal Nerve Fiber Layer Assessment

Structural changes in the optic nerve head (ONH) and retinal nerve fiber layer (RNFL) are key indicators of glaucoma [28, 34, 35]. Stereoscopic slit-lamp biomicroscopy with high-plus lenses provides detailed views of ONH features:

- **Optic Disc Size :** Natural variability affects cup size; smaller discs have smaller cups.
- **Cup-to-Disc Ratio (CDR) :**Vertical CDR reflects RGC loss; inter-eye asymmetry 0.2 raises suspicion.
- **Neuroretinal Rim (NRR) :** The ISNT rule describes normal rim thickness ($inferior > superior > nasal > temporal$); thinning or notching indicates pathology.
- **RNFL Defects :** Axonal loss leads to visible RNFL thinning and increased visibility of retinal vessels.
- **Disc Hemorrhages** an**Parapapillary Atrophy** are also associated with glaucomatous progression [34, 35].

It is important to note that optic disc cupping is not pathognomonic for glaucoma and may occur in other optic neuropathies [34].

Conclusion

To establish a solid medical and clinical foundation for the automated diagnosis of glaucoma using fundus images, this chapter introduced the medical and anatomical background of glaucoma, We began by examining the anatomy of the human eye and the visual pathway, emphasising key structures such as the retina, optic nerve.

Next, we covered the glaucoma disease, types, risk factors, and current diagnostic techniques for understanding why early and accurate glaucoma detection is crucial. The next chapter will explore artificial intelligence and deep learning fundamentals and will focus on how these technologies can address the limitations of traditional methods and enhance the accuracy and efficiency of glaucoma screening and diagnosis.

Chapter 2

Artificial Intelligence and Ophthalmology

Introduction

Artificial intelligence (AI) has rapidly transformed many fields, and ophthalmology is no exception. This chapter introduces the main concepts of AI and explores how these technologies are reshaping eye care, especially in the diagnosis of glaucoma. First, we provide an overview of the medical imaging tools commonly used in ophthalmology, such as fundus photography and optical coherence tomography (OCT).

Next, we introduce the basics of deep learning, focusing on machine learning versus deep learning, and explain the significance of convolutional neural networks (CNNs), attention mechanisms, and vision transformers in medical image analysis. Finally, the chapter addresses the main challenges that must be overcome for a successful diagnosis of glaucoma.

2.1 Medical Imaging Technologies in Ophthalmology

Ophthalmology is the medical practice of screening, diagnosing, and treating ocular diseases. Advanced imaging technologies provide clinicians with high-resolution images that contain critical information on complex eye structures such as the optic nerve head [36]. Clinicians evaluate these images for any abnormalities that indicate ocular disorders.

As a milestone of modern ophthalmology, medical imaging helps clinicians diagnose ocular disorders accurately and encourages interdisciplinary collaboration between clini-

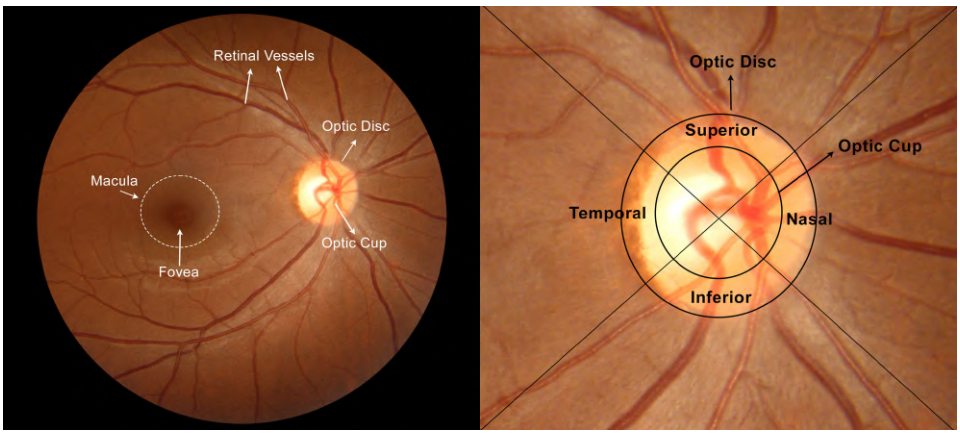


Figure 2.1: Fundus image highlighting glaucoma-relevant structures: optic disc (OD), cup (OC), and ISNT quadrants (inferior, superior, nasal, temporal)[4].

cians and computer scientists. Among the various medical imaging technologies used by ophthalmologists, fundus photography, in particular colored fundus photography, and Optical Coherence Tomography (OCT) stand out due to their frequent use in the detection of common eye diseases, such as diabetic retinopathy, cataracts, and glaucoma [36].

2.1.1 Fundus Photography

Fundus photography is used to capture images of the eye's interior surface opposite the lens, commonly referred to as the "Fundus". This imaging technology provides a 2D representation of the 3D half-transparent tissues of the retina, projected into a 2-dimensional imaging plane [37]. As illustrated in Figure 2.1, the fundus image aims to capture details on the optical disc, the optical cup, the macula, the fovea, and the blood vessels. Since it can capture the pathological changes and morphological features not visible in the standard bio-microscopic examinations, it plays a crucial role in the assessment of eye diseases.

Gloster and Garry are among the first to use fundus images for the evaluation of the cup to disc ratio, a phenomenon known as "Cupping"[38]. Cupping is the enlargement of the cup-to-disc ratio, which is a critical biomarker for glaucoma diagnosis, as it indicates damage to the optic nerve head [39, 40]. Thus, fundus photography provides an unparalleled en-face view crucial for qualitative and quantitative assessments. This is especially relevant to correlate these visible features with other clinical parameters [41].

2.1.2 Optical Coherence Tomography

Optical Coherence Tomography (OCT) is a non-invasive imaging technology that uses light waves to capture detailed cross-sectional or 3D images of tissue structures (see Figure 2.2). Unlike confocal microscopy, OCT achieves high axial (depth) resolution through the bandwidth of its light source, not the numerical aperture. This allows micrometre-scale resolution, even in eyes with optical imperfections. More specifically, OCT can image retinal layers with at least 100 times finer axial resolution in retinal imaging than traditional methods [5]. This capability makes OCT particularly valuable for visualising the retina's layered structure, which is critical in glaucoma diagnosis.

Time domain (TD)-OCT and spectral domain (SD)-OCT are the 2 main OCT methods [5]. Whereas TD-OCT uses a moving mirror, a slow but simple method to mechanically scan the depth, SD-OCT simultaneously captures depth data using a spectrometer or a tunable laser to analyse light spectra, enabling faster, higher-resolution imaging. SD-OCT is considered the standard for clinical use in modern OCT imaging due to its speed and sensitivity, while TD-OCT is less common today.

2.2 Role of Artificial Intelligence in Glaucoma Diagnosis

Early diagnosis of glaucoma remains one of the greatest challenges in modern ophthalmology due to the asymptomatic nature of the disease in the early stages [13]. This challenge, which remains even with the current advances in glaucoma treatment and research, is why glaucoma is generally referred to as the "silent thief of sight" [42]. contributing to its status as one of the leading causes of irreversible blindness worldwide [43].

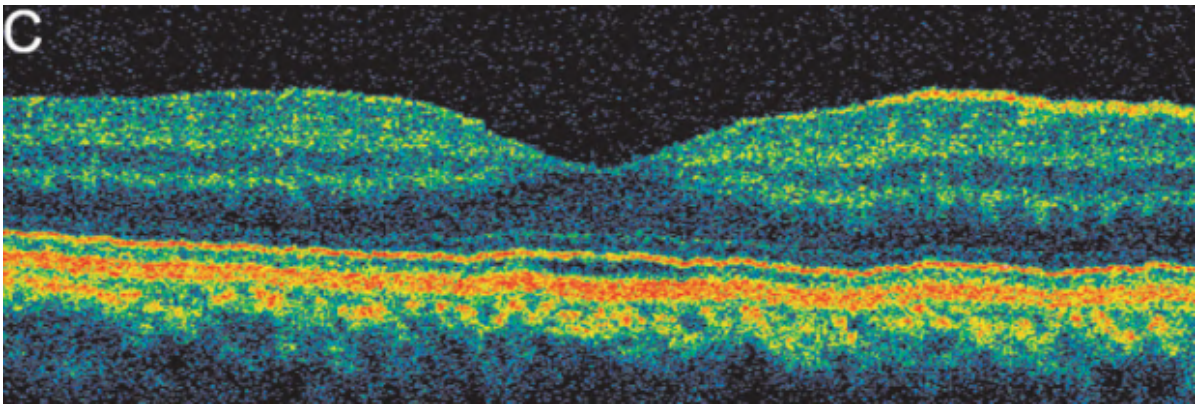


Figure 2.2: Ultrahigh-resolution ophthalmic Optical Coherence Tomography image [5].

Since there is no cure for the optic nerve damage caused by the disease, early detection is crucial to slow the progression of the disease and prevent total blindness. Traditional glaucoma diagnosis relies on tonometry, perimetry, gonioscopy, and ophthalmoscopy [44, 45]. However, these methods face various limitations, including time-consuming procedures, subjective image interpretation, and variable accuracy [13, 41].

In the last 2 decades, artificial intelligence (AI), specifically machine learning (ML) and deep learning (DL), (see Figure 2.3), has transformed medical imaging. ML relies on statistical models to enhance automation through experience [46], while DL employs a multi-layered neural network architectures to learn patterns directly from the provided data [47]. By automating feature extraction through the help of these techniques, clinicians can reduce reliance on subjective assessments and enhance diagnostic consistency [48, 49].

To diagnose glaucoma, Ophthalmologists use ocular fundus images features, such as the cup-to-disc ratio (CDR), the rim-to-disc ratio (RDR), and the ISNT (Inferior-Superior-Nasal-Temporal) parameters as illustrated in Figure 2.4 [7]. However, manually analyzing these features is time-consuming, requires specialised expertise to detect subtle changes, and is highly subjective [41]. Multiple studies have estimated that 90% of glaucoma cases globally remain undiagnosed, highlighting the need for new solutions [50, 51]. Among them, AI-driven methods have shown a remarkable improvement in diagnostic accuracy, diagnostic consistency, accelerated screening, and strengthened structure-function relationship modeling [43].

Initially, the application of artificial intelligence (AI) in glaucoma diagnosis relied on traditional image processing techniques, which depend on hand-crafted features to identify

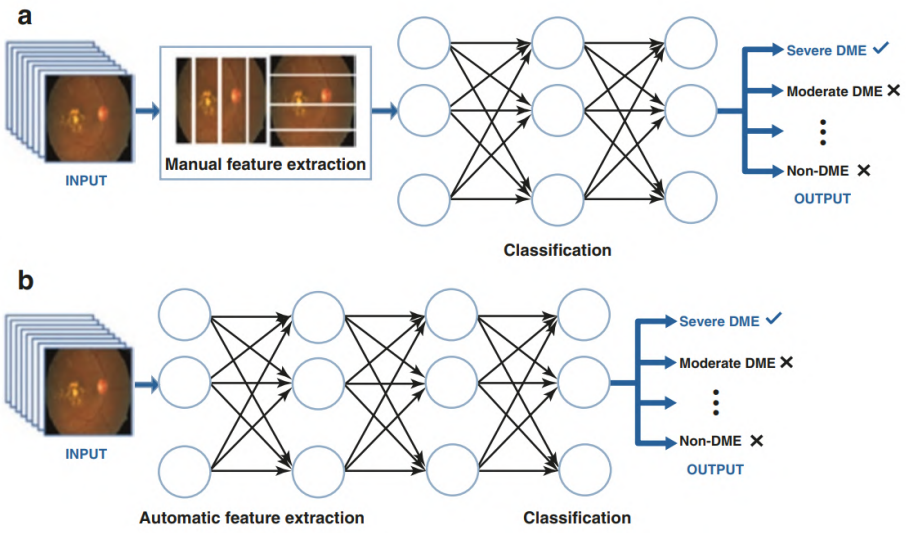
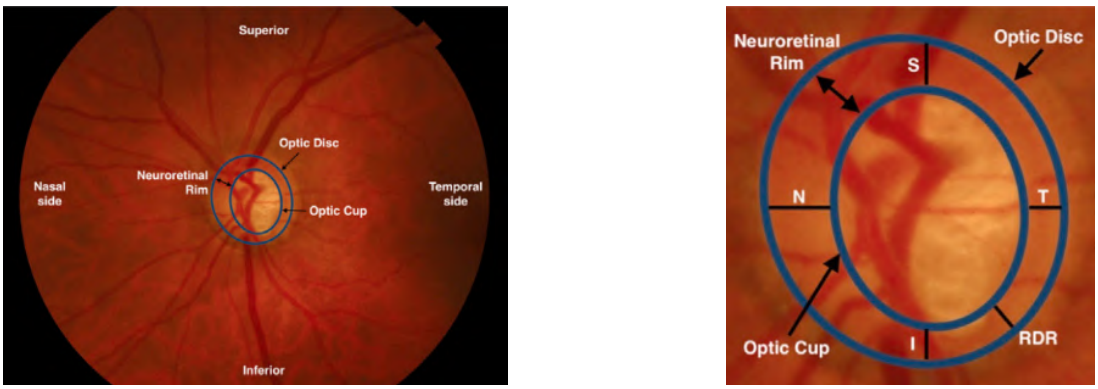


Figure 2.3: Illustration of fundus artificial intelligence-based diagnosis: (a) Machine Learning-based, (b) Deep Learning-based [6].



(a) (b)

Figure 2.4: Fundus image with the region of interest around the optic disc; (a) Ocular Fundus Image, (b) Region of interest extracted from an ocular fundus image. [7].

structural changes in fundus images [7, 42]. However, these methods are ultimately limited by their lack of generalizability due to their over-reliance on manually engineered features, which catalyzed a paradigm shift towards deep learning (DL) approaches. In contrast, DL automatically extracts hierarchical features, analyses spatial patterns and pixel relationships, and performs classification due to its multi-layered neural networks with non-linear transformations [52]. DL models can theoretically achieve greater generalizability and reduce domain-specific biases, as long as they are trained directly on large, diverse datasets [13, 53, 54].

2.3 Deep Learning Basics

2.3.1 Machine Learning vs Deep Learning

Deep Learning (DL) is a specialised field of machine learning (ML). It uses multi-layered models (artificial neural networks (ANNs) in most cases) to automatically learn hierarchical representations and features from raw data. That is what sets it apart from traditional ML, which usually requires human intervention in the form of feature engineering [55, 56]. A comparison of key differences between machine learning and deep learning techniques is summarised in Table 2.1.

While traditional ML models are often more effective and transparent when dealing with small datasets [55], DL requires large datasets and high computational power of GPUs/TPUs for training complex models [55, 56]. DL is preferable when dealing with complex tasks such as image and speech recognition, achieving higher accuracy due to automatically discovering intricate patterns and relevant features from vast amounts of data without explicit human guidance [55, 56]. DL's ability to process raw data and

Table 2.1: Comparison of conventional machine learning and deep learning.

Feature	**Machine Learning (Conventional)**	**Deep Learning**
Feature Engineering	Requires manual, human expert-driven feature extraction	Automatically learns features from raw data
Data Requirements	Can work with smaller datasets	Typically requires large datasets
Computational Power	Generally less computationally intensive	Requires significant resources
Accuracy (Complex Tasks)	May have limitations	Often achieves higher accuracy
Model Complexity	Simpler models	Complex, multi-layered neural networks
Transparency	More transparent, easier to interpret	Often a "black box," harder to interpret
Human Intervention	Significant in feature design and model tuning	Less during feature extraction, more in model architecture

excel in domains with rich, unstructured data makes it a powerful approach. However, combining DL and ML approaches can offer optimized solutions [55, 56].

2.3.2 Convolutional Neural Networks

Introduced back in 1979, CNNs are a type of neural networks dedicated to visual patterns recognition and object detection [57, 8, 6]. They are the most effective tool for image analysis since 1995 [6], and are currently being used in the study of ocular disorders [6].

Convolutional Neural Networks (CNNs) process input data as tensors, multidimensional matrices defined by image resolution [6]. Their architecture comprises hidden layers that apply convolutional, subsampling (pooling), and normalization operations [6]. As illustrated in Figure 2.5, the CNNs Key layers include:

- **Input Layer**: Resizes input images to the standard size for future processing.
- **Convolutional Layers**: Contains learnable kernels (Called "filters") that convolve across the input image to detect the spatial patterns. Early layers are responsible for detecting local patterns such as edges, while the deeper layers are responsible for hierarchically combining these features into complex structures.

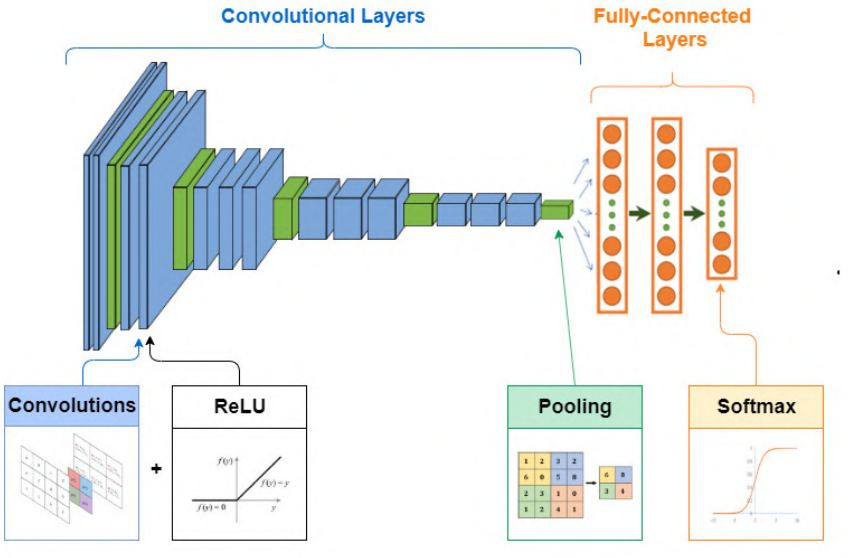


Figure 2.5: Convolutional Neural Network internal layers [8].

- **Pooling Layers**: Reduce spatial dimensions while retaining the most important information, usually through a max-pooling operation, which is called "down-sampling"
- **Rectified Linear Unit (ReLU) Layer**: Introduces non-linearity and mitigates vanishing gradients through the use of element-wise activation ($f(x) = \max(0, x)$)
- **Fully Connected Layer**: Maps high-level features to output predictions through dense connections, enabling classification or regression tasks.

2.3.3 Attention mechanism

Attention mechanisms mimic how humans focus on relevant information while ignoring the rest. They were first used in natural language processing tasks, where neural networks learned to highlight important parts of an input by assigning weights to each feature. One early and well-known method is Bahdanau attention, which computes a weighted sum of features in a matrix to emphasize what matters most for the task. This idea has since been applied in many areas, including vision tasks like medical imaging [58].

Self-attention builds on this idea by looking at how all parts of a sequence relate to each other, no matter their position. To compute how relevant one part is to another, it uses three components: queries (what you're looking for), keys (how you find it), and values (what you get). These help the model dynamically focus on the most important parts based on their relevance scores [58].

Multi-head self-attention improves this process by running several attention operations at the same time. Each ”head” looks at query-key-value relationships in its own way, picking up on different patterns. After all heads finish their work, their results are combined and transformed into a single, richer representation. This helps the model understand more complex structures than using just one attention head would allow [58].

2.3.4 Vision Transformers

In 2017, Vaswani et al [59], introduced the transformer, a new encoder-decoder architecture for natural language processing (NLP) based on self-attention. This mechanism allows models to assign dynamic weights to different input tokens—like words in a sentence—when building contextual meanings. Transformers quickly became dominant in NLP because they can process data in parallel and capture long-range dependencies more effectively than recurrent models [58].

Building on the success of transformers in natural language processing (NLP), researchers began exploring ways to integrate self-attention into convolutional neural networks (CNNs) for computer vision. However, a major breakthrough came in 2020 with the introduction of the Vision Transformer (ViT) by Dosovitskiy et al. [9]. ViT treats an image as a sequence of non-overlapping patches, linearly embeds each patch into a token, and processes these tokens using a standard transformer encoder. By applying self-attention across all patches, ViT effectively captures long-range spatial dependencies such as the relationships between the optic disc, cup, and surrounding vasculature in fundus images offering a key advantage over CNNs in holistic image analysis. An overview of the Vision Transformer architecture is shown in Figure 2.6.

In a standard Transformer architecture, the input is a one-dimensional sequence of token embeddings. However, since images are inherently two-dimensional, Vision Transformers (ViTs) begin with a preprocessing step known as **image patching** or **patch embedding** . In this step, the input image is divided into fixed-size square patches—commonly 16×16 pixels. Each patch is then flattened into a one-dimensional vector and linearly projected into a fixed-dimensional embedding space to form patch embeddings.

To retain spatial information, learnable **positional embeddings** are added to the sequence of patch embeddings. Additionally, a learnable **class token** is prepended to the sequence, which serves to aggregate global features across the entire image. The final input to the Transformer encoder thus consists of the combined patch embeddings, positional embeddings, and the class token.

This sequence is then processed through a standard Transformer encoder, which alternates between multi-head self-attention (MSA) and multilayer perceptron (MLP) blocks. Each of these blocks includes residual connections and layer normalization for

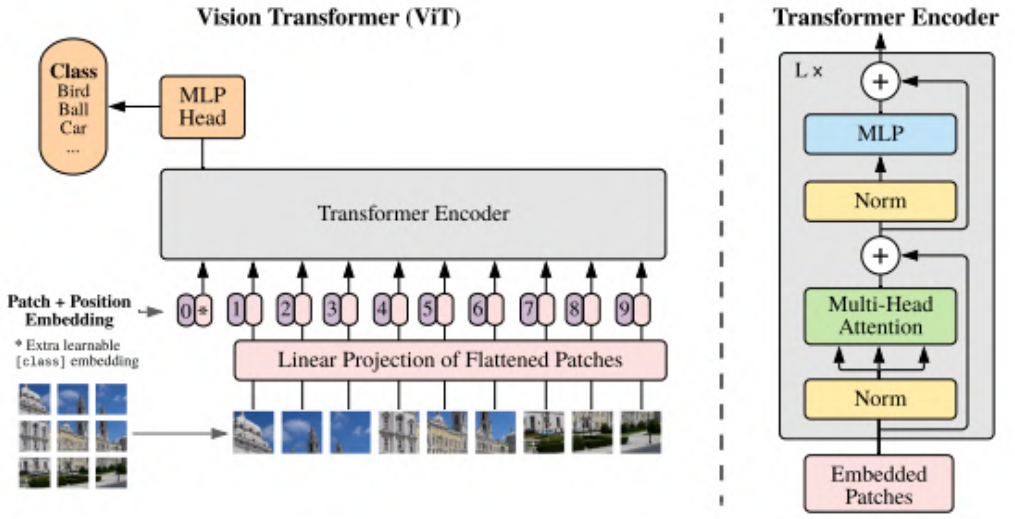


Figure 2.6: Overview of the Vision Transformer model [9].

stable training. Finally, the output corresponding to the class token is passed through a classification head—typically a small MLP—to produce the final prediction of the image label [9].

Vision Transformers (ViTs) differ significantly from convolutional neural networks (CNNs) in both architecture and inductive bias, see Table 2.2. CNNs embed spatial locality, 2D spatial hierarchies, and translation equivariance into every layer. In contrast, ViTs rely on self-attention mechanisms that do not have such built-in spatial constraints.

In ViTs, only the multilayer perceptron (MLP) layers enforce local and translationally equivariant processing. Meanwhile, the self-attention layers are responsible for capturing long-range dependencies across the image. The 2D structure of images is used only at two points: during the initial extraction of patches and when adapting resolution during fine-tuning. Moreover, position embeddings in ViTs are initialized without prior spatial knowledge. This means the model must learn the relationships between patches solely from the training data [9].

Table 2.2: Key Differences Between Convolutional Neural Networks and Vision Transformers.

Feature	**Convolutional Neural Networks (CNNs)**	**Vision Transformers (ViTs)**
Core Architecture	Convolutional layers with hierarchical feature extraction	Self-attention layers with patch-based processing
Inductive Biases	Built-in spatial locality and translation equivariance	Learns spatial relationships from data
Spatial Processing	Local receptive fields via convolution kernels	Global attention patterns with local MLPs
Position Information	Implicit through strided convolutions	Explicit learned position embeddings
2D Structure Usage	Exploited at every layer	Used only in patch extraction and adaptation
Context Handling	Local to global progression through layers	Immediate global context access
Data Efficiency	Effective with medium-sized datasets	Requires large-scale training data
Computational Demand	Optimized through parameter sharing	Quadratic complexity in sequence length
Interpretability	More transparent spatial processing	Black-box attention mechanisms

Hybrid models combine CNNs with Transformers by feeding feature maps from CNNs into Vision Transformers (ViTs). These feature maps are extracted from intermediate CNN layers and converted into patch embeddings through a linear projection. This design allows for flexibility; for instance, 1×1 convolutions can be used to collapse spatial dimensions before mapping features into the Transformer's embedding space. The use of a classification token and learnable positional embeddings follows the standard ViT architecture [9].

During fine-tuning at higher resolutions, Vision Transformers (ViTs) adapt pre-trained models for specific tasks by replacing the original classification head with a randomly initialized feedforward layer of size $D \times K$, where K is the number of target classes. When input resolution increases while patch size remains fixed, the sequence length also increases. To accommodate this, pre-trained position embeddings are resized using two-dimensional interpolation based on their initial spatial positions. This interpolation step, along with patch extraction, represents the only explicit use of 2D spatial bias in ViT inference [9].

2.3.5 Foundation Models

Foundation models (FMs) are large-scale deep learning systems trained on vast and diverse datasets, often using self-supervised learning methods. These models are designed to be flexible and adaptable. After their initial training, they can be fine-tuned for a wide range of specific tasks in healthcare, such as clinical language processing, medical imaging, omics data analysis, and graph-based learning. This flexibility helps address the challenge of applying AI tools to the complex and varied needs of medical practice [60].

In medical imaging, foundation models can serve as strong starting points for tasks like image segmentation and disease classification. Because they are trained on large and varied datasets, they learn general patterns that can transfer well to new problems, even when labeled data is limited. This transferability is especially valuable in healthcare, where collecting and annotating large datasets is often difficult and expensive [61].

Healthcare Foundation Models (HFMs) are a special class of FMs tailored for medical tasks. They are used in clinical language models, computer vision, and even in multi-modal applications that combine images, text, and other data types. HFMs support the development of more advanced and intelligent healthcare technologies by enabling better generalization and reducing the need for task-specific models [60, 61].

Segment Anything Model 2

The Segment Anything Model 2 (SAM2) is a recent example of a foundation model designed for promptable visual segmentation in both images and videos. It uses a transformer-based architecture with streaming memory, allowing it to process video in real-time. SAM2 can segment objects or regions based on various user prompts, such as clicks, bounding boxes, or text descriptions. This makes it highly interactive and adaptable to different segmentation needs [10].

Compared to the original SAM, SAM2 achieves higher accuracy and is six times faster at image segmentation. For video tasks, it performs better and requires three times fewer user interactions [10]. A key innovation in SAM2 is its data engine, which uses user feedback to continuously improve both the model and its training dataset. This feedback loop has enabled the creation of the largest video segmentation dataset to date. SAM2 also supports a wide range of annotation types, including masks, text, and user actions like clicks and scribbles.

The architecture of SAM2 is illustrated in Figure 2.7. Its promptable design and real-time capabilities make it especially suitable for clinical applications, where quick and accurate segmentation of medical images is crucial. By leveraging foundation models like SAM2, researchers and clinicians can achieve more precise and efficient image analysis, supporting tasks such as optic disc and cup segmentation for glaucoma diagnosis.

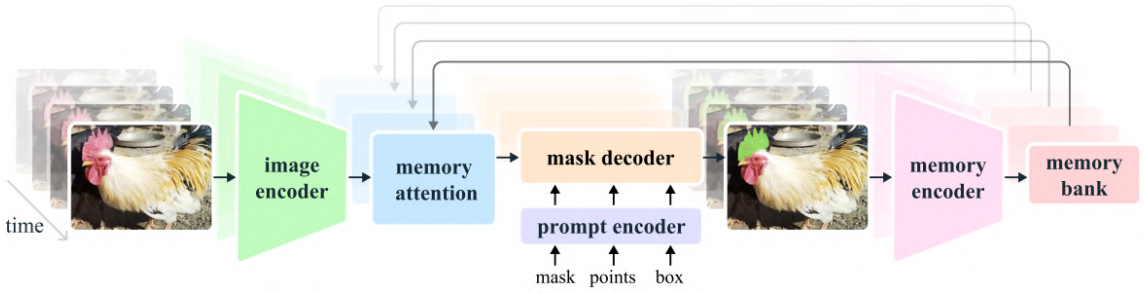


Figure 2.7: Illustration of the segment anything model 2 architecture [10].

2.4 Challenges in medical AI

Trust is a cornerstone of effective healthcare systems. While the integration of AI in medicine offers substantial benefits, it also introduces new risks and amplifies existing cybersecurity challenges. Biased AI models may produce incorrect medical assessments, and adversarial attacks can compromise model integrity unless countered by robust defenses. Failures in medical AI can erode public confidence in healthcare services. Therefore, strong governance and oversight are essential in the development and deployment of AI systems. Quinn et al. [11] from McKinsey & Company emphasize the importance of involving three key stakeholder groups: (1) algorithm developers, (2) independent validators, and (3) operational staff, to ensure accountability and reliability.

These three stakeholder groups are essential in addressing the primary challenges of AI adoption in healthcare, which can be categorized as (1) conceptual, (2) technical, and (3) humanistic challenges. Figure 2.8 provides an overview of these challenges.

- **conceptual challenges:** Despite its capabilities, AI remains a simulation of human intelligence and lacks true clinical reasoning or intuition. Successful implementation requires well-defined problem statements, appropriate data alignment, and careful mitigation of biases and errors during model training [11].
- **technical challenges:** Building effective AI models involves expert-driven design, rigorous data quality control, and hyperparameter tuning. These tasks are often complicated by the limited availability, complexity, and sensitivity of healthcare data [11].
- **humanistic challenges:** Ethical concerns remain central to AI deployment in medicine. Issues such as the ”black box” nature of AI models can undermine patient autonomy, while embedded biases may disproportionately harm marginalized populations. Over-reliance on AI risks reducing human empathy in care delivery, and unclear accountability for model failures presents further concern. Close collaboration among developers, validators, and clinicians is essential to ensure responsible and equitable use of AI in healthcare [11].

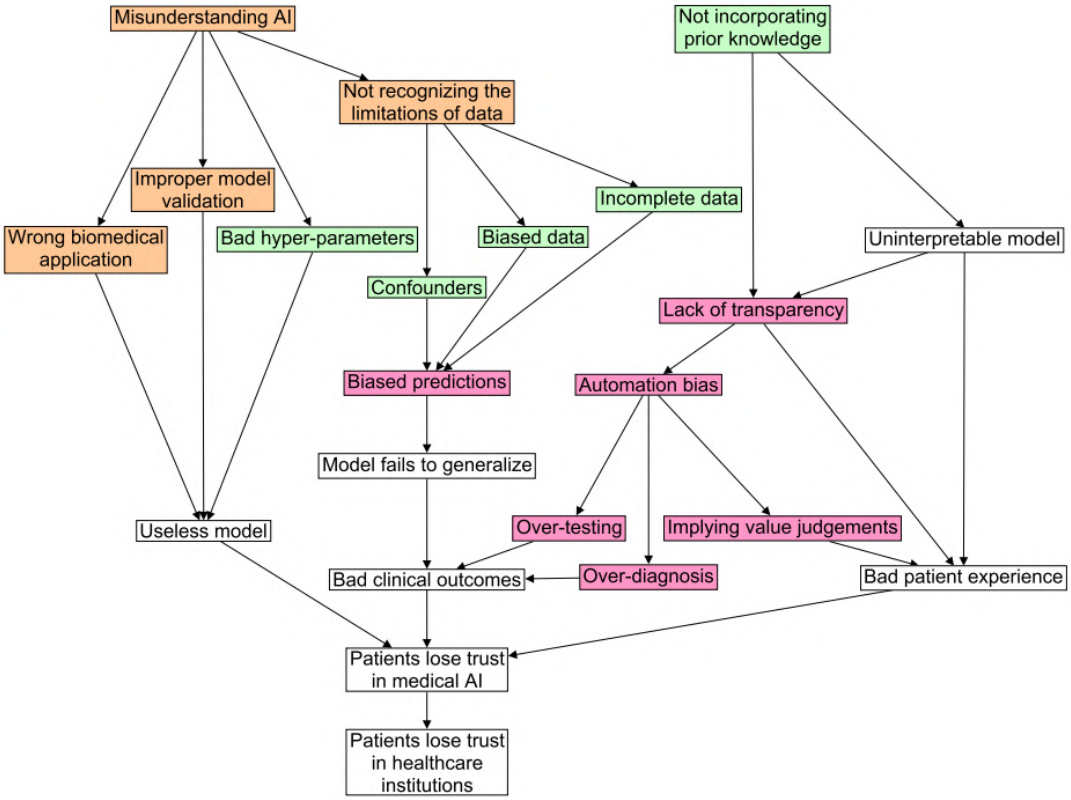


Figure 2.8: Key challenges in medical AI and how they relate to one another and clinical care. Node color represents the type of challenge: conceptual (orange), technical (green), or humanistic (pink)[11].

To effectively address these challenges, the three expert groups—Developers, Validators, and Operational Staff—each play a distinct and complementary role in ensuring the safe, ethical, and effective integration of AI in healthcare. Figure 2.9 illustrates how each group contributes to overcoming key challenges in medical AI.

- **Developer:** Responsible for designing and building AI systems, developers must combine technical expertise with medical knowledge, ethics, and an understanding of patient values. This requires transdisciplinary collaboration. Short-term strategies include forming cross-disciplinary teams, while long-term solutions involve integrating digital medicine and ethics training into formal education programs to better align technical and clinical perspectives [11].
- **Validators:** Tasked with ensuring AI systems meet clinical standards, validators evaluate performance through methods such as peer review and randomized clinical trials. They help bridge the ”AI chasm” by assessing how well models translate into real-world, equitable, and safe clinical practice. Long-term solutions include training experts in AI auditing and establishing regulatory frameworks, such as “Turing stamps,” to certify AI safety and fairness [11].

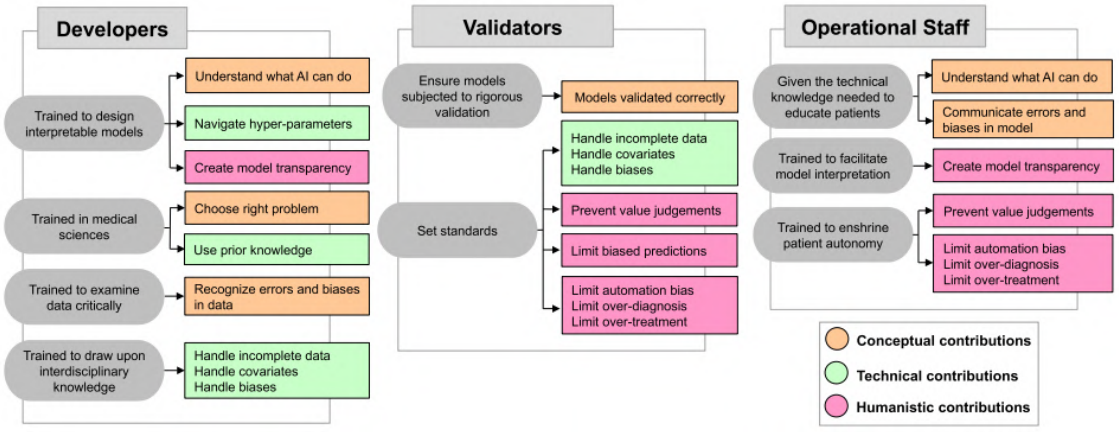


Figure 2.9: How accredited expert groups–developers, validators, and operational staff–can help overcome the key challenges in medical artificial intelligence [11].

- **Operational staff:** Charged with integrating AI into clinical workflows while maintaining high-quality patient care, operational staff work closely with clinicians and IT professionals. In the short term, improving AI literacy among healthcare workers can alleviate concerns. Long-term strategies include embedding AI education into medical training, and introducing specialized roles such as a "digital doctor" or "digital nurse" to oversee ethical and effective AI use at the point of care [11].

Conclusion

In this chapter, we provided a solid foundation for understanding how AI and deep learning are being used to advance automated glaucoma diagnosis from fundus images. We started by reviewing the main imaging technologies used in eye care, such as fundus photography and optical coherence tomography, which are essential for visualizing the eye's internal structures relevant to glaucoma diagnosis.

Then, we introduced the basics of deep learning, including the differences between classic machine learning and deep learning, and explained key models like convolutional neural networks, attention mechanisms, and vision transformers. Furthermore, we examined the emergence of foundation models, such as the Segment Anything Model 2, and highlighted their potential for flexible and efficient image segmentation in medical practice. Finally, we addressed the main challenges in applying AI to medicine, including conceptual, technical, and ethical concerns, and emphasized the need for collaboration among developers, validators, and clinical staff to ensure safe and trustworthy AI systems.

In the next chapter, we will build on this background by reviewing the state of

the art in fundus image segmentation and glaucoma classification. We will examine recent methods, compare their strengths and weaknesses, and identify current gaps and opportunities for future research.

Chapter 3

State of The Art

As detailed in the previous 2 chapters, glaucoma remains one of the leading causes of global blindness. As it currently has no known cure, early diagnosis remains the most crucial point in treating the disease and preventing total blindness. However, early diagnosis proved to be an incredibly hard task due to the asymptomatic nature of glaucoma, which led to ophthalmologists cooperating with computer scientists to develop automated systems to automate the process of diagnosing glaucoma and improve early detection rates.

This chapter provides a comprehensive overview of the recent retinal image segmentation and glaucoma classification studies. These studies showcase the need to improve the detection of changes to the optic disc, which is a biomarker in glaucoma detection.

The first section covers state-of-the-art retinal image segmentation methods, precisely, optic disc and optic cup segmentation. These methods, which includes U-nets and various variants and improvements, are the pillar of modern glaucoma detection systems as they provide detailed information on the optic nerve head.

The second section dives into recent glaucoma classification methods, highlighting different approaches that aim to improve the early detection of glaucoma. This overview aims to showcase the current state of the field, providing a framework to understand the automated glaucoma detection landscape, and identifying existing gaps for future research.

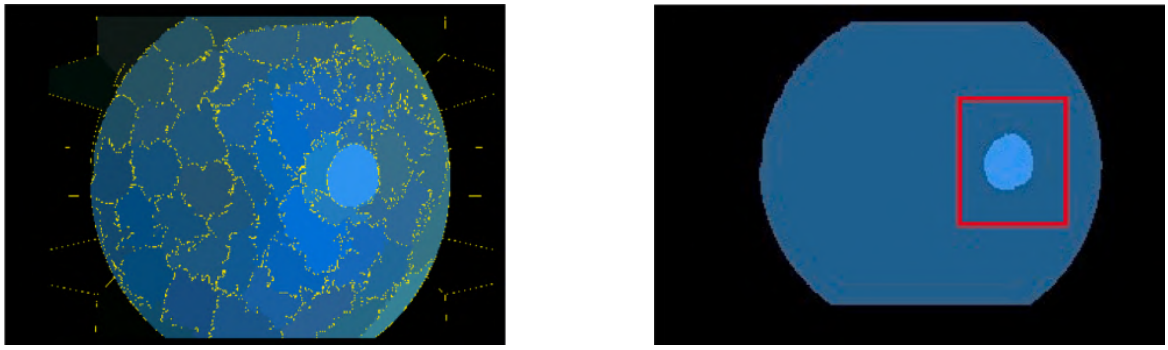
3.1 Fundus Image Segmentation

the optic nerve head (ONH), located within the optic disc. This imaging technique enables clinicians to detect and monitor structural and morphological changes in the ONH,

which are often indicative of glaucomatous damage. Therefore, accurate segmentation of the optic disc—and, in some cases, the optic cup—is essential for reliable glaucoma diagnosis. Over the years, numerous studies have proposed a variety of approaches to address the challenges associated with optic disc segmentation in fundus images.

Optic disc segmentation

Sanghavi and Kurhekar [12] proposed a lightweight framework for optic disc (OD) segmentation, specifically designed for environments with constrained computational resources. Rather than employing deep learning models, their approach combines traditional image processing techniques to achieve efficient and effective segmentation. The process begins with Simple Linear Iterative Clustering (SLIC), which segments the image into superpixels by clustering pixels in the Lab color space. This step increases sensitivity to minor color variations, thereby improving the detection of the optic disc's unique color characteristics.An example of the SLIC output is shown in Figure 3.1a. Next, a normalized graph cut algorithm is applied to separate the optic disc from the retinal background. This method represents superpixel clusters as nodes in a graph and optimizes segmentation by considering both color and spatial differences. An example of this stage is provided in Figure 3.1b. Collectively, these two stages exploit color coherence and spatial relationships to perform segmentation without requiring substantial computational power. The framework demonstrates accuracy comparable to conventional methods like CLAHE. Moreover, its use of Lab space transformation and adaptive clustering enables superior performance in low-contrast and poorly illuminated fundus images. Although its accuracy and precision are slightly lower than those of U-Net-based deep learning models, it offers significantly greater computational efficiency. Therefore, due to its simplicity, interpretability, and minimal resource requirements, this method presents a practical solution for optic disc segmentation in resource-limited settings, effectively balancing performance and efficiency.



(a) (b)

Figure 3.1: Example of SLIC image (a) and Nomalized graph cut image (b)[12].

In [62], the authors applied the Fuzzy C-Means (FCM) clustering algorithm to the V channel of fundus images after converting them into the HSV color space, where the optic disc exhibits enhanced contrast. Preprocessing involved morphological dilation using a disk-shaped structuring element with a radius of 20 pixels to fill gaps caused by blood vessels. Then, FCM was used to minimize an objective function and assign pixels to clusters based on membership degrees. For optic cup segmentation, which lies within the optic disc, the green channel was standardized to normalize intensity variation before applying Otsu's thresholding to binarize the bright central region of the optic cup. Post-processing included contour extraction to refine boundary delineation and compute clinical parameters such as the cup-to-disc ratio (CDR). The resulting segmentation pipeline was incorporated into a hybrid feature set that combined structural features (e.g., CDR, blood vessel ratio) and non-structural features (e.g., GLCM texture matrices). These features were fused with raw image data to train a ResNet-152 classifier. The method demonstrated high performance on the ORIGA and ACRIMA datasets.

An improved version of the Separable Linked Segmentation Network (SLS-Net) was introduced by Haider et al. [63], incorporating residual connections to enhance performance. This updated model, named SLSR-Net (Separable Linked Segmentation Residual Network), uses an encoder-decoder structure designed specifically for optic disc (OD) and optic cup (OC) segmentation in retinal images. The architecture integrates Separable Convolutional Link (SCL) units, which replace conventional convolutions with depthwise separable convolutions to lower computational cost while preserving accuracy. Moreover, by using fewer pooling layers, the network maintains larger feature maps 62×62 at the final encoding stage, thereby reducing spatial information loss and improving the delineation of ambiguous structures such as the OC boundary. Additionally, external residual skip connections—illustrated in Figure 3.2—help prevent feature degradation during downsampling and upsampling. These connections transfer fine-grained details from earlier encoder layers to corresponding decoder layers through element-wise addition, enhancing segmentation precision. The model was tested on four publicly available datasets: Drishti-GS, REFUGE, Rim-One-r3, and Drions-DB. It achieved a Dice coefficient of 0.915 for OC and 0.980 for OD on the Drishti-GS dataset, surpassing U-Net and SegNet in accuracy despite having only 4.67 million trainable parameters. Therefore, SLSR-Net offers a computationally efficient yet highly accurate solution, making it well-suited for clinical applications that rely on precise glaucoma biomarker segmentation.

Li et al. were the first to apply nonlinear spiking neural P systems (NSNP) to optic disc and optic cup segmentation [64]. They combined NSNP with a U-shaped deep learning architecture, addressing common issues in traditional CNNs and U-Net variants—such as loss of spatial detail during downsampling and poor modeling of long-range dependencies. Instead of standard convolutional layers, they used ConvSNP-DeSep units as the backbone. These units mimic the nonlinear activation behavior of spiking neurons through depth-separable operations, reducing information loss and improving feature extraction while keeping computation efficient. In addition, they introduced the

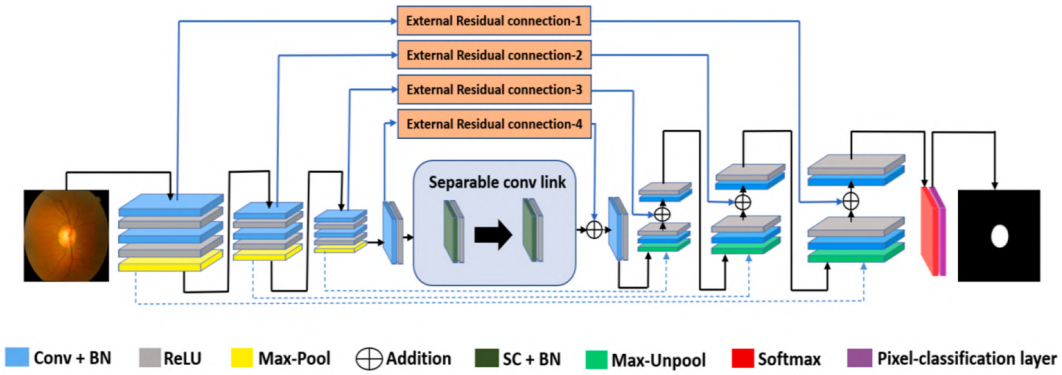


Figure 3.2: Overview of the proposed SLSR-Net model.

SRS-Operator, a new resampling mechanism that preserves multi-scale spatial details by reassembling features across larger receptive fields, thereby reducing blurring caused by pooling or strided convolutions. The hybrid design demonstrated strong performance, especially in capturing fine boundary changes between the optic cup and disc—an essential factor in early glaucoma detection—and achieved strong results.

To address common challenges in medical image segmentation, such as ambiguous anatomical boundaries and noise amplification, Venkataiah et al. (2025) proposed a novel variant of the U-Net architecture, termed TMUnet-NL. The model integrates three key components to improve segmentation accuracy, particularly for optic disc and optic cup structures. First, it combines the computational efficiency of MobileNet—a lightweight convolutional neural network (CNN) designed for mobile devices—with the well-established encoder-decoder framework of U-Net. This integration allows the model to benefit from U-Net's skip connections, which preserve spatial details during upsampling, while leveraging MobileNet's depthwise and pointwise convolutions to reduce computational cost. Second, a Transformer layer is introduced to enhance feature representation by capturing long-range dependencies and contextual relationships, capabilities that traditional CNNs often lack. Third, the authors propose a hybrid loss function that combines Mean Squared Error (MSE), to ensure pixel-level accuracy, and Mean Boundary Error (MBE), to penalize inaccuracies along object boundaries. This dual-component loss function promotes both global structural consistency and precise boundary delineation. As a result, TMUnet-NL achieves superior performance over baseline models such as U-Net, DeepLab v3, and MobileUNet, particularly in handling complex anatomical structures with unclear edges.

Singh et al. [65] introduced a novel optic disc segmentation model based on a modified level set algorithm that combines edge- and region-based strategies. This method uses edge gradients to detect boundaries, while also relying on regional intensity statistics to guide contour evolution in uniform areas. This hybrid design helps overcome common issues in classical level set methods, such as contour divergence and inaccuracies in

regions with weak edges or high noise. A key innovation is the dynamic adjustment of parameters, including the balloon force for "u," which adapts to local image gradient information through a Gaussian-filtered Laplacian operator. Moreover, the model replaces the standard Gaussian kernel with a bilateral kernel in the region-based term. This change helps preserve edge sharpness while smoothing homogeneous regions. As a result, the contour evolves more reliably across varying contrast and noise levels. This leads to accurate boundary detection of the optic cup, reduces contour instability, and lowers sensitivity to initial conditions. Compared to traditional level set approaches, this modified method demonstrates improved accuracy and robustness. It is therefore better suited for glaucoma diagnosis, particularly when assessing metrics such as the Cup-to-Disc Ratio (CDR).

Xiong et al.[66] proposed Multi-GlaucNet, a robust encoder-decoder architecture dedicated to optic disc (OD) segmentation and glaucoma diagnosis. The model makes full use of advanced deep learning components to achieve a high performance. For the encoder, the model employs bottleneck blocks, which are a parameter-efficient designs with stacked 1×1 and 3×3 convolutional layers with residual connections, for hierarchical features extraction with minimal computational overhead. When decoding, conventional transposed convolutions are replaced by pixel shuffle, which rearranges multi-channel features into spatially refined outputs to enhance resolution. A channel attention mechanism, which dynamically prioritizes salient features during mask reconstruction, is employed to further enhance the previous process. The model achieved exceptional optic disc segmentation accuracy, reaching a Dice coefficient of 96.7% on the ORIGA dataset, and outperforming single-task benchmarks like U-Net, while showcasing competitive performance compared to PKRT-Net on REFUGE. It should be noted that this framework also incorporates blood vessel segmentation to support multi-task learning. This performance highlights the benefits of incorporating bottleneck-based feature compression, resolution-aware decoding, and attention-driven refinement for OD localization, setting a new benchmark in automated glaucoma diagnostic pipelines.

Xiong et al. [66] introduced Multi-GlaucNet, an encoder-decoder model designed for optic disc segmentation and glaucoma diagnosis. The architecture integrates modern deep learning techniques to deliver high accuracy and efficiency. In the encoder, bottleneck blocks are used to extract hierarchical features while keeping computational cost low. These blocks consist of stacked 1×1 and 3×3 convolutional layers with residual connections, offering a compact yet powerful design. During decoding, pixel shuffle replaces traditional transposed convolutions, improving spatial resolution by rearranging multi-channel outputs. Additionally, a channel attention mechanism is applied to emphasize key features during mask reconstruction, further refining segmentation quality. Experimental results show that Multi-GlaucNet achieves a Dice coefficient of 96.7% on the ORIGA dataset, surpassing models such as U-Net and performing competitively against PKRT-Net on the REFUGE dataset. Notably, the framework also includes blood vessel segmentation, supporting a multi-task learning approach. Overall, the study demonstrates that combining feature compression, resolution-aware upsampling, and

attention-based refinement can significantly improve optic disc localization, advancing the state of automated glaucoma diagnosis.

Liu et al. [67] introduced ECSD-Net, a multi-task encoder-decoder model with domain adaptation, to tackle performance degradation in glaucoma-related optic disc (OD) and optic cup (OC) segmentation across different datasets. The model uses EfficientNet B7 for feature extraction and a Bidirectional Feature Pyramid Network (BiFPN) to combine multi-scale features, improving the detection of both local details—such as ambiguous OC boundaries—and global context. In the decoder, transposed convolutions with skip connections gradually upsample features, restoring spatial resolution while integrating high-level semantic information from the encoder. To address domain shift—such as differences in image quality and acquisition protocols—an adversarial discriminator is applied in the output space. This component aligns predictions from source and target domains through adversarial training, reducing distribution discrepancies without the need for labeled target data. As a result, the model shows strong generalization across domains. ECSD-Net was tested on three public datasets: REFUGE, Drishti-GS, and RIM-ONE-r3. It achieved Dice scores of 0.965 for OD and 0.876 for OC on the Drishti-GS dataset, outperforming state-of-the-art models like pOSAL and WGAN-DA. Moreover, the model demonstrated efficiency by achieving high accuracy with fewer parameters than traditional U-Net and its variants. This makes ECSD-Net a promising solution for clinical applications where computational efficiency and cross-dataset consistency are essential.

Shyamalee et al. [68] introduced an Attention U-Net architecture designed to balance segmentation precision and computational efficiency in optic disc (OD) and optic cup (OC) detection from fundus images. To improve localization accuracy, the model incorporates attention gates within skip connections, emphasizing clinically significant regions while suppressing irrelevant background features. The encoder component is adapted from pre-trained convolutional neural networks—namely Inception-v3, VGG19, and ResNet50—enabling the use of transfer learning to enhance feature extraction. Within this framework, the attention mechanism is formalized through an attention coefficient that dynamically adjusts the contribution of encoder outputs based on decoder-derived gating signals. Among the evaluated backbones, ResNet50 demonstrated superior performance on the RIM-ONE dataset, likely due to its residual blocks which alleviate gradient vanishing and support deeper feature learning. Prior to model input, images underwent preprocessing using CLAHE and median filtering to improve contrast and reduce noise. Data augmentation was also applied, including rotation, flipping, and brightness adjustments, to increase dataset diversity and reduce overfitting. The model achieved strong segmentation results, with a Dice coefficient of 99.40%, alongside high sensitivity—key metrics for effective glaucoma screening. Overall, this work highlights the value of combining attention mechanisms with transfer learning in advancing medical image segmentation tasks.

To address challenges such as low-contrast boundaries and pathological variations in

retinal fundus images, Rekha and Jayashree [69] introduced a balanced segmentation approach using the Pufferfish Optimization Algorithm (POA). This method aims to improve the accuracy of optic disc (OD) and optic cup (OC) segmentation, which are critical for glaucoma diagnosis. The POA is inspired by the defense mechanism of pufferfish and operates through a two-phase strategy: global exploration and local exploitation. During the global exploration phase, the algorithm simulates predator attacks to diversify the search space and locate global optima by adjusting population member positions. In the local exploitation phase, it mimics the pufferfish's inflation response to refine solutions through localized adjustments. Moreover, the fitness function ensures segmentation quality, while the termination criteria maintain computational efficiency. Experimental results on public datasets—specifically DRISHTI-GS, HRF, DRIONS-DB, and PSGIMS—showed that the POA-based method achieved a high performance, with Dice scores above 95% and Jaccard indices exceeding 90%. These results outperformed conventional techniques such as convolutional neural networks (CNNs) and VGG16. Importantly, the study highlights that accurate OD and OC segmentation leads to precise cup-to-disc ratio (CDR) calculations, which are essential for reliable glaucoma detection.

The methods reviewed in this work were specifically designed to address key challenges in optic disc and optic cup segmentation. Some aimed to strike a balance between computational efficiency and segmentation accuracy [63], while others focused on overcoming difficulties caused by low-contrast boundaries [69]. Nevertheless, each method comes with its own set of limitations. Several rely heavily on the availability of large, annotated datasets [64, 68], which can restrict their scalability and practical application. In addition, a number of these approaches show limited generalizability across different imaging conditions or populations [63, 66]. These issues highlight the need for further research to improve robustness and reduce dependency on labeled data. A summary of the main limitations of each method is provided in Table 3.1.

Table 3.1: Limitations of state-of-the-art classification methods.

Method	**Dataset(s)**	**Limitations**
GANs + pretrained MobileNetV2 [13]	ACRIMA; DRISHTI-GS; ORIGA	- GAN-generated images may include artifacts, reducing real-world reliability. - MobileNetV2's lightweight design may miss subtle glaucoma features. - GAN training and fine-tuning demand high computational power.
DeepEyeNet [70]	REFUGE; DRISHTI-GS	- High computational demands restrict accessibility for resource-limited settings. - Segmentation errors could propagate to feature extraction and classification. - Untested robustness on minority classes or noisy clinical images.
Graph convolution based multi-task model [71]	DRISHTI-GS; RIM-ONE r3	- Requires extensive annotated data for training. - Complex architecture hinders interpretability and deployment. - Classification depends on segmentation accuracy.
Deep Retinal Vision (DRV) model [14]	ACRIMA; RIM-ONE; ORIGA	- Dataset imbalance as small datasets tend to have bias towards one class. - Demands significant GPU resources, limiting real-time deployment in resource-constrained clinics.
Attention-driven Frame networks + ELM [72]	DRISHTI-GS; ORIGA	- Computational complexity due to integration of U-Net+, CapsNet, and ELMs. - Dataset imbalance and limited diversity despite augmentation.
Hybrid ResNet-GRU network [73]	RIM-ONE; DRIONS-DB	- Hybrid ResNet-GRU increases training time and complexity. - Classification accuracy depends on segmentation quality; errors propagate.
End-to-end simultaneous model [74]	REFUGE; DRISHTI-GS	- Performance depends on segmentation; errors in cup/disc segmentation propagate to classification. - Limited interpretability compared to biomarker-based approaches.
Hybrid framework [75]	ORIGA; DRIONS-DB	- Dependency on handcrafted features, which may fail to capture complex patterns. - Training and ensembling three distinct models requires significant resources, hindering deployment in limited settings.

3.2 Fundus Image Classification

Retinal image classification is a critical component in the development of automated systems for glaucoma detection and diagnosis. This process involves the analysis of fundus images to differentiate between healthy and glaucomatous eyes, primarily based on structural characteristics of the optic nerve head, such as the optic disc and cup. Key indicators of glaucoma include an increased cup-to-disc ratio, thinning of the neuroretinal rim, and morphological asymmetry in the optic nerve. Because these features are often subtle and subject to inter-observer variability, manual assessment can be inconsistent. Therefore, recent research has increasingly focused on deep learning models to enhance the accuracy, consistency, and efficiency of glaucoma screening through automated retinal image classification. These approaches aim to reduce diagnostic errors and support early intervention, which is essential for preventing vision loss associated with glaucoma.

Glaucoma Classification

In their 2025 study, Govindharaj et al. [13] proposed a classification model that leverages transfer learning using the pre-trained MobileNetV2 architecture. The primary objective was to classify fundus images into two categories: normal and glaucomatous. To tackle the issue of imbalanced datasets, the authors employed Generative Adversarial Networks (GANs) to generate high-quality synthetic fundus images. These images not only enriched the training set but also preserved realistic vessel structures and anatomical features. Prior to classification, Gaussian filtering was applied to reduce noise and improve input image quality. MobileNetV2 was selected for its computational efficiency and lightweight design. It was then fine-tuned on the augmented dataset to better capture glaucoma-specific features. As shown in Figure 3.3, the overall framework integrates data augmentation, preprocessing, and model adaptation. The model achieved an accuracy of 98.9%, surpassing contemporary methods such as Deep CNN and U-Net + EfficientNet. Moreover, high performance across additional metrics—including precision, recall, specificity, and sensitivity—demonstrates the model's reliability and effectiveness. This work highlights how combining GAN-based data augmentation with transfer learning can significantly improve classification performance, particularly in addressing class imbalance and enhancing generalization across varying imaging conditions. Overall, it presents a promising approach for automated glaucoma diagnosis in clinical settings.

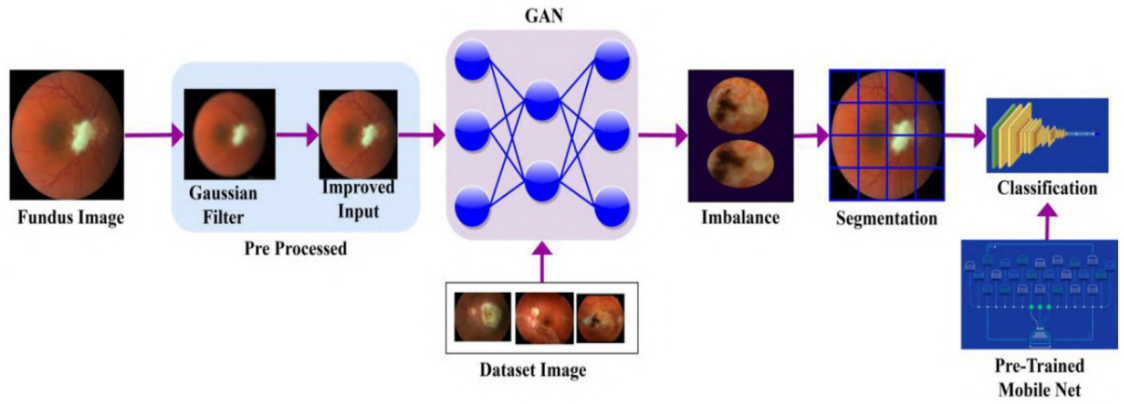


Figure 3.3: illustration of framework proposed by Govindharaj et al. [13].

The authors of [70] proposed DeepEyeNet, a hybrid diagnostic framework for glaucoma detection that integrates deep learning with manually engineered features. The architecture employs ConvNeXtTiny as its backbone, pre-trained on ImageNet to extract high-level hierarchical features from standardized fundus images. In addition to these deep features, domain-specific anatomical and texture-based features are manually extracted, including the cup-to-disc ratio (CDR), neuroretinal rim (NRR) area, and Gray-Level Co-occurrence Matrix (GLCM) descriptors. To capture both structural and textural biomarkers effectively, a late fusion strategy is applied, combining the deep features with the handcrafted ones. To improve classification performance, especially in handling class imbalance, the model incorporates a custom loss function that combines focal loss with a penalty for false negatives. A key innovation in the framework is the use of Adaptive Genetic Bayesian Optimization (AGBO), which merges the global search capability of genetic algorithms with the precision of Bayesian optimization for efficient hyperparameter tuning. The model was evaluated on the EyePACS-AIROGS-light-V2 dataset, achieving an accuracy of 95.84%, outperforming conventional models such as VGG16 and ResNet50. Ablation studies confirmed the importance of both AGBO and feature fusion, as removing either component led to a drop in accuracy by 1.42% and 2.72%, respectively. Moreover, the model demonstrated strong discriminative ability with an AUC-ROC of 0.9848, highlighting its potential as a reliable tool for automated glaucoma screening in clinical settings.

Lenka et al. [71] introduced a multi-task learning model for glaucoma detection that integrates structural and image-based features. The architecture is based on a modified U-Net, where the encoder employs MobileNetV2 to extract hierarchical visual features from retinal images. To enhance feature localization, an attention module is incorporated into the skip connections, directing focus toward the optic disc and cup regions—areas critical for glaucoma diagnosis. The model processes image-based and structural features along separate pathways: image features are passed through a fully connected layer for classification, while structural features—primarily the vertical cup-to-disc ratio (CDR), derived from segmented optic structures—are analyzed using a linear classifier. A multi-

task loss function combines focal loss with standard classification loss to mitigate class imbalance and improve model performance on difficult cases. Experimental results across publicly available datasets such as DRISTI-GS, and REFUGE. Showing that the model achieves an accuracy of 97.43% and an AUC-ROC of 0.985. These results indicate strong discriminative capability in distinguishing between healthy and glaucomatous retinas, establishing the model as a new benchmark in automated glaucoma detection. Figure 3.4 depicts the proposed method.

In their 2025 study, Naidana et al. [14] proposed the Deep Retinal Vision (DRV) model, a hybrid deep learning architecture that integrates convolutional neural network (CNN) bottlenecks with transformer-based self-attention mechanisms for glaucoma classification from retinal fundus images. The model follows a multi-stage pipeline beginning with image pre-processing, where input images are resized to 224×224 pixels and enhanced using histogram equalization to improve contrast. Next, feature extraction is performed using modified MobileNetV2-inspired bottleneck layers, which employ expansion, depthwise convolution, and projection blocks to efficiently capture spatial features while minimizing computational load. These CNN-derived features are then fed into transformer self-attention layers, enabling the model to learn long-range dependencies and emphasize diagnostically significant regions such as the optic disc and cup. The self-attention mechanism dynamically computes weighted interactions between global and local features, allowing the model to focus on subtle pathological cues often indicative of glaucoma. Finally, classification is carried out using fully connected layers that map the refined feature representations into one of two output classes: healthy or glaucomatous. Experimental results demonstrate that the DRV model achieves a high performance, attaining classification accuracy of 93% on the DRISHTI-GS dataset—outperforming both conventional CNNs and lightweight deep learning models.

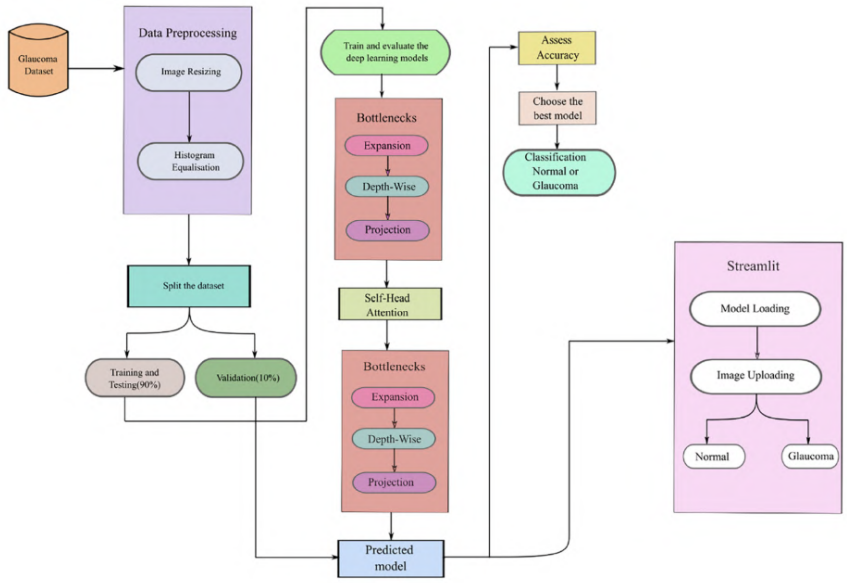


Figure 3.4: Illustration of the method proposed by Naidana et al. [14]

A hybrid classification architecture that integrates attention mechanisms with Extreme Learning Machines (ELM) was introduced by the authors in [72]. The model begins with a U-Net framework to segment the optic disc, followed by a Capsule Network (CapsNet), which captures hierarchical spatial relationships through dynamic routing between capsules, enabling more effective feature extraction. To improve modeling of long-range dependencies, self-attention layers are used in place of traditional convolutional layers; this allows the network to focus selectively on diagnostically significant regions. Next, an Extreme Learning Machine (ELM)—a fast and computationally efficient single-hidden-layer neural network—is employed to classify the extracted features. Consequently, the model achieves over 90% accuracy across three publicly available datasets: DRIONS-DB, DRISHTI-GS, and HRF. This performance is attributed to the synergy of accurate segmentation via U-Net, robust feature representation using CapsNet, and swift classification with ELM. Moreover, the model surpasses existing methods such as E-NETS and E-DLM in both accuracy and generalization. This is further enhanced through data augmentation and median filtering, which reduce noise and improve consistency across datasets, underscoring the approach's efficiency and practical viability.

Kumar et al. [73] propose a hybrid deep learning framework that integrates Residual Neural Networks (ResNet) and Gated Recurrent Units (GRU) for glaucoma classification. In this architecture, ResNet extracts spatial features from segmented optic disc and cup images, while GRU captures temporal relationships between these features. To further enhance model performance, the authors introduce a Modified Density Factor-based Honey Badger Algorithm (MDF-HBA). This algorithm optimizes the number of hidden neurons in both ResNet and GRU layers to maximize classification accuracy. The classification process is conducted in two stages: first, disc-segmented and cup-segmented images are processed independently using the combined ResNet-GRU structure; second, the output scores from each pathway are averaged to generate the final prediction. Model evaluation was based on key metrics such as accuracy, F1-score, and Matthews Correlation Coefficient (MCC). Results show that the proposed method outperforms conventional models—including DCNNs, CNNs, and ensemble techniques—by notable margins. For example, he model achieves a 24.1% improvement in F1-score over DCNNs and a 34.5% improvement over ensemble methods, with an overall accuracy of 98.7% on the tested dataset. Additionally, ROC-AUC analysis confirms the model's robustness, revealing high true-positive rates and low false-positive rates. These findings suggest that the hybrid architecture, together with MDF-HBA optimization, offers a promising approach for accurate glaucoma detection.

Hervella et al. [74] introduced an end-to-end multi-task deep learning model designed to perform glaucoma classification and optic disc/cup segmentation concurrently. The architecture builds upon a U-Net-based encoder-decoder structure, enhanced with a novel classification head that utilizes skip connections to aggregate multi-scale features from the decoder. Consequently, this design enables shared feature representations between the two tasks, promoting more efficient learning. The classification output

is generated through a mini-encoder that integrates both global context and localized anatomical details. Moreover, the proposed multi-adaptive optimization strategy serves as a key innovation; it dynamically balances task-specific gradients during training without requiring manual tuning of loss weights. As a result, both tasks contribute equally to the model's parameter updates. The model was validated on two publicly available datasets—REFUGE and DRISHTI-GS—achieving AUC-ROC scores of 0.976 and 0.9474, respectively. These results surpass those of both single-task models and baseline multi-task approaches. Additionally, the integration of segmentation into the model improved classification performance by capturing morphological cues such as the cup-to-disc ratio, while also utilizing non-morphological features directly from raw fundus images.

Aljohani and Aburasain [75] introduced a hybrid framework that combines traditional machine learning (ML) and deep learning (DL) models for glaucoma classification. The proposed approach uses an ensemble strategy involving three models: a Random Forest (RF) classifier trained on texture features extracted using the Gray-Level Co-Occurrence Matrix (GLCM), and two convolutional neural networks (CNNs), ResNet-50 and VGG-16, trained on grayscale retinal fundus images. Prior to model training, RGB images were converted to grayscale to enhance texture analysis and reduce noise. GLCM-derived features—such as contrast, correlation, and energy—were fed into the RF classifier, while the CNNs processed the grayscale images directly. Final predictions were generated by applying a simple voting rule: if at least two models classified an image as glaucoma, the final output was labeled as glaucoma; otherwise, it was labeled as normal. The ensemble method significantly outperformed individual models, achieving an accuracy of 95.41%, precision of 99.37%, recall of 88.33%, and an F1-score of 93.52%. This study demonstrates the value of integrating handcrafted features with deep learning, showing improved diagnostic performance over previous works that relied on single-model architectures or smaller datasets.

The methods reviewed in this work demonstrate strong performance and achieve high accuracy. However, each method also faces notable challenges that future research must address. Common limitations include a heavy reliance on preprocessing steps [74], high computational demands [13, 72], and performance that is highly dependent on accurate segmentation [70]. These issues may restrict their practical use and generalizability. Table 3.2 provides a detailed summary of the key limitations associated with each method.

Table 3.2: Limitations of state-of-the-art classification methods.

Method	**Dataset(s)**	**Limitations**
GANs + pretrained MobileNetV2 [13]	ORIGA	- GAN-generated images may include artifacts, reducing real-world reliability. - MobileNetV2's lightweight design may miss subtle glaucoma features. - GAN training and fine-tuning demand high computational power.
DeepEyeNet [70]	EyePACS-AIROGS-light-V2	- High computational demands restrict accessibility for resource-limited settings. - Segmentation errors could propagate to feature extraction and classification. - Untested robustness on minority classes or noisy clinical images.
Graph convolution based multi-task model [71]	DRISHTI-GS; ORIGA; REFUGE	- Requires extensive annotated data for training. - Complex architecture hinders interpretability and deployment. - Classification depends on segmentation accuracy.
Deep Retinal Vision (DRV) model [14]	ACRIMA; DRISHTI-GS; ORIGA	- Dataset imbalance as small datasets tend to have bias towards one class. - Demands significant GPU resources, limiting real-time deployment in resource-constrained clinics.
Attention-driven Frame networks + ELM [72]	DRISHTI-GS; HRF; DRIONS-DB	- Computational complexity due to integration of U-Net+, CapsNet, and ELMs. - Dataset imbalance and limited diversity despite augmentation.
Hybrid ResNet-GRU network [73]	ACRIMA; Fundus-Train-Val-Data(FTVD); ORIGA	- Hybrid ResNet-GRU increases training time and complexity. - Classification accuracy depends on segmentation quality; errors propagate.
End-to-end simultaneous model [74]	REFUGE; DRISHTI-GS	- Performance depends on segmentation; errors in cup/disc segmentation propagate to classification. - Limited interpretability compared to biomarker-based approaches.
Hybrid framework [75]	ORIGA; ACRIMA; G1020; REFUGE	- Dependency on handcrafted features, which may fail to capture complex patterns. - Training and ensembling three distinct models requires significant resources, hindering deployment in limited settings.

Conclusion

This chapter provided a comprehensive review of the latest methods for automated glaucoma diagnosis from fundus images, emphasizing both segmentation and classification approaches based on deep learning. We discussed the main segmentation techniques for the optic disc and cup, highlighting the critical role of anatomical accuracy in extracting key glaucoma biomarkers, such as the cup-to-disc ratio and neuroretinal rim area. Recent advancements, especially those using convolutional neural networks (CNNs) and hybrid models, have led to significant improvements in segmenting these structures, although challenges remain due to image variability and subtle anatomical boundaries in complex cases.

For classification, we showed that deep learning models including CNNs, DenseNets, and Vision Transformers achieve strong results in distinguishing glaucomatous from healthy eyes, particularly when focusing on clinically relevant regions like the neuroretinal rim. However, the effectiveness and generalization of these models depend on the quality and diversity of annotated datasets, as well as the application of regularization and data augmentation techniques to prevent overfitting.

Additionally, we highlighted the growing importance of model interpretability for clinical adoption. Recent approaches that incorporate attention mechanisms or visualization tools help clinicians better understand automated decisions, thereby increasing trust in these systems.

Despite these advances, several challenges persist, such as ensuring robustness to data variability, balancing sensitivity and specificity, and integrating automated tools into clinical workflows. Future research should focus on enhancing model generalization, expanding annotated datasets, and developing hybrid solutions that combine the strengths of different network architectures for more reliable glaucoma detection.

As we move to the next chapter, we shift from reviewing existing work to presenting our own proposed framework for automated glaucoma diagnosis. Chapter 4 details the experimental design and evaluation of our method, which integrates advanced image enhancement, a two-step optic disk and optic cup segmentation strategy, and glaucoma classification. We describe the datasets, evaluation metrics, and provide a thorough analysis of the results, demonstrating how our approach addresses many of the challenges identified in this chapter and advances the field of automated glaucoma screening.

Chapter 4

Contribution and Experimental Study

Introduction

Glaucoma remains a major cause of irreversible blindness worldwide, and early detection is essential for preventing vision loss. Automated diagnosis using deep learning offers promising solutions, especially in settings where specialist expertise is limited. In this chapter, we present a comprehensive experimental study of our proposed framework for automated glaucoma diagnosis from fundus images. The focus is on evaluating the effectiveness of our segmentation and classification models using standardized datasets and robust metrics.

The chapter is organized into several key sections to ensure clarity and logical flow. First, we outline the overall structure of our glaucoma diagnosis framework, detailing the pre-processing steps applied to the input images, the segmentation of optic disc and optic cup using a fine-tuned Segment Anything Model 2 (SAM2), and the subsequent classification phase employing a Vision Transformer (ViT) model. Next, we describe the experimental setup, including the system configuration, the datasets used (namely ORIGA and REFUGE) and the evaluation metrics selected for both segmentation and classification tasks.

Following the methodological overview, we present the experimental results. This includes a detailed analysis of segmentation performance, comparing our approach to state-of-the-art methods using metrics such as Dice coefficient and Intersection over Union (IoU). We also report on the classification results, highlighting the strengths and limitations of our ViT-based model in distinguishing between glaucomatous and healthy eyes. Finally, the chapter concludes with a summary of the main findings, a discussion of observed limitations.

4.1 Proposed Glaucoma Diagnosis Framework

For automated glaucoma diagnosis, this study proposes a classification framework based on recent deep learning models. We aim to classify an input retinal fundus image as either glaucomatous (indicating the presence of glaucoma) or non-glaucomatous (healthy).

As illustrated in Figure 4.1, the framework consists of three main phases. In the preprocessing phase , the input retinal fundus image is enhanced to improve quality and consistency before segmentation. In the optic disc/cup segmentation phase , a SAM2 model is first fine-tuned on the enhanced image to segment the optic disc region. Then, a second SAM2 model is fine-tuned on the segmented optic disc images to isolate the optic cup. The optic cup is then subtracted from the optic disc to generate an image representing the optic rim area. Finally, in the glaucoma classification phase , a Vision Transformer (ViT) model is fine-tuned on the optic rim images to determine whether the eye is glaucomatous or healthy.

The details of each phase are described in the following subsections.

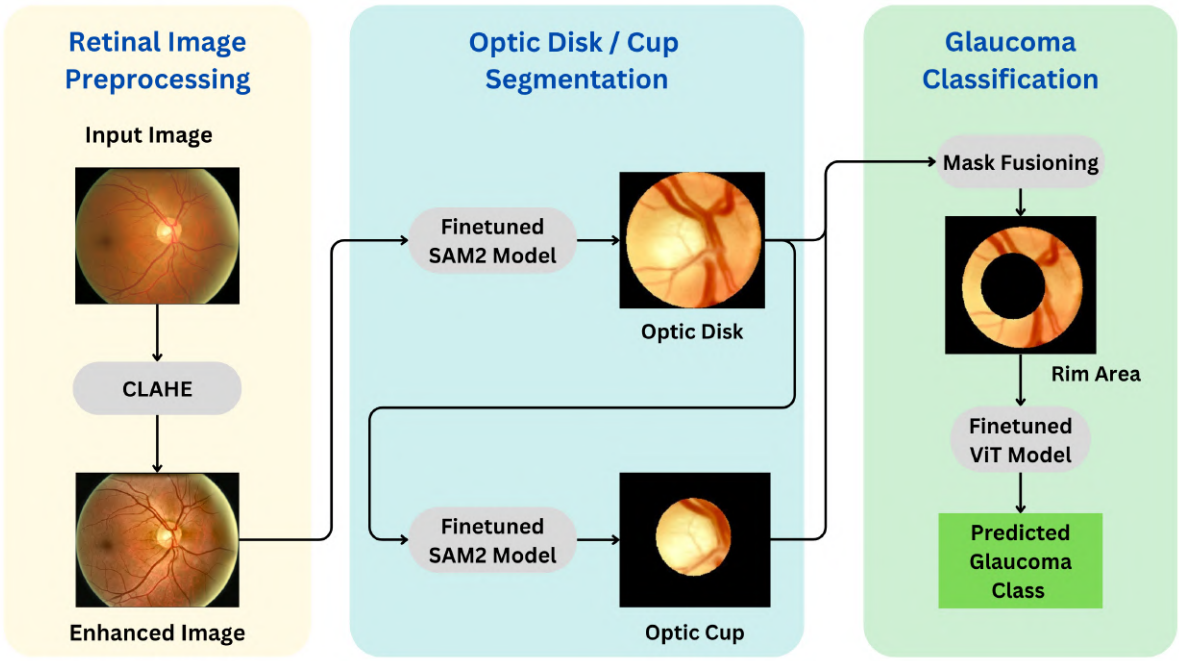
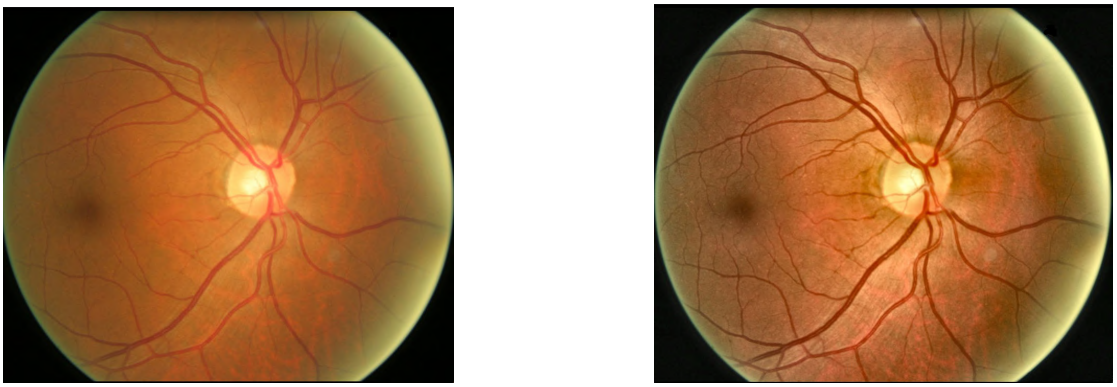


Figure 4.1: Illustration of the proposed glaucoma classification framework.

4.1.1 Phase 1: Pre-Processing

Contrast Limited Adaptive Histogram Equalization (CLAHE) is a commonly used image enhancement technique, especially in medical imaging such as retinal fundus photography.



(a) (b)

Figure 4.2: Illustration of contrast enhancement using Contrast Limited Adaptive Histogram Equalization technique on ORIGA dataset fundus image, normal (a) and enhanced (b)

It improves local contrast, which helps make subtle anatomical features more visible—such as the optic disc and optic cup. These structures are vital for glaucoma detection, as changes in their shape and size often signal disease progression. The method works by dividing an image into small regions called tiles. Then, histogram equalization is applied independently to each tile. This localized approach enhances contrast where needed while limiting noise amplification in smoother areas.

To apply CLAHE effectively, the RGB fundus image is first converted into the LAB color space. In this space, the ”L” channel encodes lightness, while ”A” and ”B” encode color information. CLAHE is applied only to the ”L” channel, using a clip limit of 2 and a tile grid size of 8×8. These settings help balance contrast improvement and noise control. After enhancement, the image is converted back to the RGB color space for further processing. Visual comparison shows that CLAHE significantly improves image clarity. As shown in Figure 4.2, the enhanced image (4.2b) displays sharper details than the original (4.2a), supporting more accurate feature extraction and diagnosis.

Why Use Contrast Limited Adaptive Histogram Equalization

Fundus images for glaucoma detection often suffer from low contrast, uneven illumination, and noise, which obscure critical retinal features like blood vessels, optic cup boundaries, and neuroretinal rim details. These issues hinder accurate segmentation and classification by deep learning models, as subtle indicators of glaucoma may remain undetectable. Contrast Limited Adaptive Histogram Equalization (CLAHE) is widely preferred over global or naive methods because it enhances local contrast without over-amplifying noise, thanks to its tile-based processing and clip limit. This localized approach improves the visibility of fine structures while preserving anatomical gradients, making it especially

effective for fundus images with spatially varying illumination. Studies show that CLAHE preprocessing significantly boosts segmentation accuracy and enhances feature extraction for classification tasks [76, 77, 78]. Additionally, CLAHE enriches training data when used in data augmentation, indirectly improving model robustness. Parameters like tile size and clip limit are carefully tuned to balance contrast enhancement and noise control, ensuring reliable detection of glaucoma biomarkers like cup-to-disc ratios and disc hemorrhages. Thus, CLAHE plays a vital role in enhancing image quality and advancing glaucoma screening performance.

4.1.2 Phase 2: Optic Disc And Optic Cup Segmentation

To generate a discriminative fundus image representation, we focus on extracting the neuroretinal rim area, which is a key biomarker in glaucoma diagnosis. For this reason, we adopted a two-step procedure for the accurate segmentation of the optic disc and optic cup.

The initial segmentation aimed to accurately localize the larger anatomical structure (the optic disc), which serves as the region of interest for subsequent processing. In the second step, we aim to isolate the smaller and more variable optic cup area. By applying this sequential segmentation approach, we were able to reduce background interference and focus on clinically relevant structures with greater precision.

4.1.2.1 Fine-tuning Segment Anything Model 2 for Optic Disc Segmentation

Fine-tuning is a widely used transfer learning technique. It adapts a pre-trained model to a new task by updating only a subset of its parameters. This approach retains the knowledge the model acquired during its initial training. The advantages are clear: reduced computational cost, faster convergence, and improved performance, especially when working with limited data.

In this study, we fine-tuned the SAM2 model for optic disc segmentation. Specifically, we updated only the mask decoder and prompt encoder. The image encoder was kept fixed. This preserved its strong feature extraction capabilities, which were learned from the large-scale Segment Anything 1 Billion (SA-1B) dataset. We used the Lion optimizer, assigning different learning rates to different parts of the model. Additionally, we applied the OneCycleLR learning rate scheduler to ensure stable training. During training, we selected prompts based on anatomical landmarks. Each optic disc mask received 3–7 carefully chosen prompts, rather than random points.

The total loss function was a weighted sum of four components as follows:

$$\mathcal{L}_{\text{total}} = 0.5 \cdot \mathcal{L}_{\text{BCE}} + 0.5 \cdot \mathcal{L}_{\text{Dice}} + 0.2 \cdot \mathcal{L}_{\text{Focal}} + 0.1 \cdot \mathcal{L}_{\text{Score}}$$



(a) (b)

Figure 4.3: Optic disc segmentation result from finetuned Segment Anything Model 2 (SAM2), (a) input image, (b) output optic disk region.

Were, $\mathcal{L}_{\text{BCE}}$ refers to Binary Cross Entropy loss. $\mathcal{L}_{\text{Dice}}$ improves overlap between predicted and ground truth masks. $\mathcal{L}_{\text{Focal}}$, with $\gamma = 2.0$, addresses class imbalance. Lastly, $\mathcal{L}_{\text{Score}}$ aligns the model's confidence scores with segmentation accuracy. We further enhanced training efficiency using mixed precision and gradient clipping. Figure 4.3 presents example output from finetuned SAM2 model. It highlights the segmented optic disc region in 4.3b, from an enhanced retinal fundus image input 4.3a

4.1.2.2 Fine-tuning Segment Anything Model 2 For Optic Cup Segmentation

Optic cup segmentation is more complex task than optic disc segmentation. This is because the optic cup lies within the boundaries of the optic disc. Our proposed hierarchical approach helps preserve anatomical consistency and makes better use of spatial relationships between structures. In our proposed method, optic cup segmentation is limited to the previously segmented optic disc region, rather than the full image. This has several advantages: it removes irrelevant background content, focuses the model's attention on the region of interest, and improves computational efficiency by working with smaller image patches. Figure 4.4 illustrates an optic cup segmentation output (4.4b) of the finetuned SAM2 model segmented optic disc region 4.4a.

The finetuning process for optic cup segmentation follows the same general strategy as for the optic disc. However, we train the model for more epochs—200 instead of 100 to allow it to better capture the subtle differences between the optic cup and surrounding tissues. We used the same SAM2 model hyperparameters for both tasks, including the loss function and point selection method. This ensures consistency across stages and supports a smooth transition from one task to the next. This sequential design works well with SAM2's prompt-based architecture. By focusing on the cropped optic disc area, the model can better detect the optic cup boundary using localized context.



(a) (b)

Figure 4.4: Optic cup segmentation result from finetuned Segment Anything Model 2 (SAM2), (a) input optic disk image, (b) output optic cup region.

4.1.3 Phase 3: Glaucoma classification

After segmenting the optic disc and optic cup, the framework proceeds to the glaucoma classification stage, which includes a feature fusion and analysis step. In this phase, the segmented optic disc and optic cup regions are combined through mask fusion. Specifically, the optic cup region is subtracted from the optic disc mask to isolate the neuroretinal rim which is an important biomarker for glaucoma detection. Figure 4.5 illustrates the resulting rim area image.



Figure 4.5: Illustration of rim area image.

The resulting rim area image is resized to 224×224 pixels to ensure consistency in input dimensions for the model. These images are then used to fine-tune a pretrained Vision Transformer (ViT) model. ViT was selected due to its strong performance in capturing long-range dependencies and global contextual information, characteristics that are especially important when analyzing subtle structural variations in the neuroretinal rim [79]. Its self-attention mechanism allows the model to dynamically focus on discriminative

regions, such as areas with abnormal thinning, which are critical for glaucoma detection.

Fine-tuning was chosen over training from scratch to leverage the model's pre-trained weights, which encode general visual features learned from large-scale datasets like ImageNet [80]. This approach reduces the need for large amounts of medical imaging data and accelerates convergence while maintaining high feature adaptability to the specific task. The ViT generates a probability distribution over the two classes: glaucomatous or non-glaucomatous.

By focusing exclusively on the clinically relevant rim region, the method enhances the discriminative power of the extracted features while reducing noise interference. As a result, the proposed approach achieves higher classification accuracy compared to methods based on whole-image analysis.

4.2 Experimental Study

To evaluate the performance of the proposed glaucoma classification framework, we conducted a series of experiments. This section outlines the experimental setup, covering the system configuration, used datasets, and evaluation metrics. Furthermore, we present the achieved results and compare them with state-of-the-art methods.

4.2.1 System Configuration

The experiments were conducted on a high-performance computer featuring a 12th Gen Intel® Core™ i7-12700F CPU with 12 cores, 64 GB of RAM, and an NVIDIA GeForce RTX 3080 Ti GPU with 12 GB of VRAM. The system operated on Ubuntu 22.04.4 LTS. CUDA 12.4 was used to enable GPU acceleration, and PyTorch 3.10.14 was employed for efficient parallel training of deep learning models.

4.2.2 Datasets

Two publicly available retinal fundus image datasets are used in our experimental study, the Retinal Fundus Glaucoma challenge (REFUGE) [81] and the Online Retinal fundus Image database for Glaucoma Analysis (ORIGA)[82] for the both optic disk/cup Segmentation and glaucoma classification tasks.

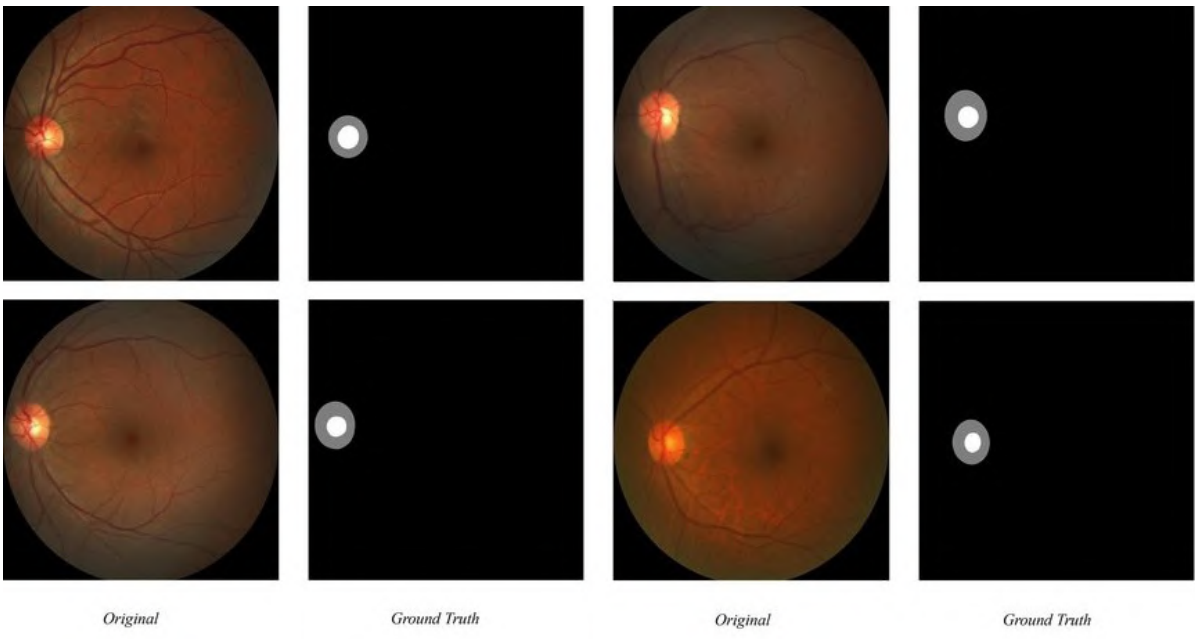


Figure 4.6: Sample images with optic disk/cup groundtruth masks from REFUGE fundus images dataset.

4.2.2.1 REFUGE Dataset

The Retinal Fundus Glaucoma (REFUGE) dataset was developed for a glaucoma detection challenge hosted at the Ophthalmic Medical Image Analysis (OMIA) workshop, held in conjunction with the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). It contains 1200 annotated retinal fundus images, each accompanied by optic disc and cup segmentation masks (see Figure 4.6. The dataset is categorized into two classes: glaucoma (120 images) and healthy (1080 images). Its primary purpose is to enable standardized comparison of automated methods for glaucoma classification and optic disc/cup segmentation [81].

4.2.2.2 ORIGA Dataset:

The Online Retinal Fundus Image Database for Glaucoma Analysis (ORIGA) is a widely used benchmark for evaluating automated glaucoma diagnosis models. It consists of 650 retinal fundus images collected during a population-based study conducted by the Singapore Eye Research Institute. Of these, 168 images are from glaucomatous eyes, while 482 were randomly selected from non-glaucomatous eyes[82]. Sample images from ORIGA fundus images dataset are shown in Figure 4.7. Table 4.1 Summaries the main features of ORIGA and REFUGE Datasets.

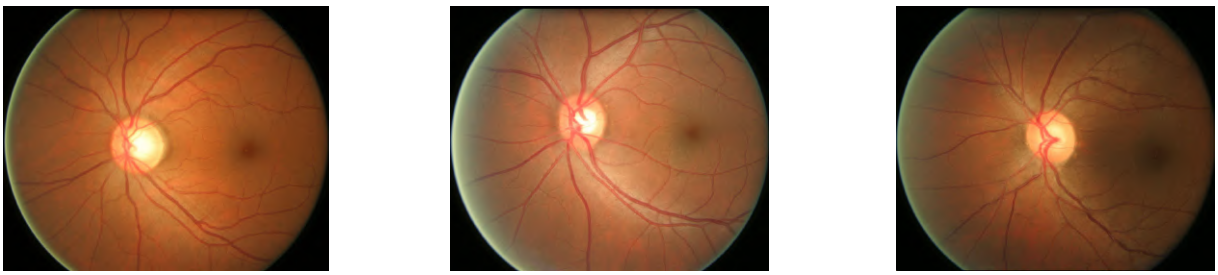


Figure 4.7: Random samples from the ORIGA fundus image dataset for glaucoma analysis.

Table 4.1: Summary of ORIGA and REFUGE Datasets

Attribute	**ORIGA [82]**	**REFUGE [81]**
Number of Images	650	1200
Image Resolution	Varied	2124 × 2056 and 1634 × 1634
Ground Truth Annotations	Optic Disc & Cup	Optic Disc & Cup
Glaucoma Labels	Yes (binary)	Yes (binary) (800 images)
Class Distribution	168 glaucomatous, 482 normal	80 glaucomatous, 720 normal

4.2.3 Evaluation Metrics

4.2.3.1 Segmentation Metrics

- **Intersection over Union (IoU)**: is a common metric for evaluating image segmentation models. It measures how much the predicted segment overlaps with the groundtruth segment. A higher IoU indicates better alignment between the predicted and actual segments. The IoU value is calculated by dividing the size of the overlapping region by the total number of pixels in either segment as follow:

$$\text{IoU} = \frac{|\text{Prediction} \cap \text{Ground Truth}|}{|\text{Prediction} \cup \text{Ground Truth}|}$$

- **Dice Coefficient**: is another measure used to assess the similarity between two segments. As detailed in the next equation, the Dice Coefficient compares the size of the overlap to the average size of the two segments. This metric gives more

balanced results when the sizes of the predicted and ground truth segments differ. Higher values show better agreement between the predicted and true regions.

$$\text{Dice} = \frac{2 \cdot |\text{Prediction} \cap \text{Ground Truth}|}{|\text{Prediction}| + |\text{Ground Truth}|}$$

4.2.3.2 Classification Metrics

- **Accuracy**: measures how often a classification model makes correct predictions. While accuracy gives a general idea of performance, it can be misleading when classes are imbalanced. For instance, if most data belong to one class, a model might appear accurate simply by predicting that class every time, even if it fails to detect other important cases. The Accuracy can be calculated by dividing the number of correct predictions by the total number of predictions made, such as:

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

where TP means the number of True Positives TN the true negatives, FP and FN are the number of false positives and false negatives, respectively.

- **Precision**: shows how many of the predicted positive cases were correct. as exprimed in the following equation, the precision metric is calculated by dividing true positive predictions (TP) by the sum of true positives and false positives (FP). Precision is important when false positives are costly. For example, in medical testing, high precision means fewer healthy people are incorrectly diagnosed as sick.

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

- **Recall (Sensitivity or True Positive Rate)**: measures how well a model identifies actual positive cases. Recall is crucial when missing positive cases is dangerous or unacceptable. In disease detection, for example, high recall ensures that most sick patients are correctly identified. It is calculated by dividing true positives by the sum of true positives (TP) and false negatives (FN) as follows:

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

- **F1-Score**: this metric is the average of precision and recall, giving equal weight to both. It is especially useful when dealing with imbalanced datasets, where both false positives (FP) and false negatives (FN) matter. A high F1-score means the model performs well in both identifying real positive cases and avoiding incorrect positive predictions. The F1-Score is mathematically formulated by the next equation:

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

- **Specificity (True Negative Rate)**: measures how well a model identifies actual negative cases. As shown in the next formula, it is calculated by dividing true negatives (TP) by the sum of true negatives (TN) and false positives (FP). High specificity means the model rarely mislabels negative cases as positive. This is important in health-related tasks, where it's essential not to label healthy individuals as sick.

$$\text{Specificity} = \frac{TN}{TN + FP}$$

- **Area Under the ROC Curve (AUC-ROC)**: evaluates a model's ability to distinguish between classes at different decision thresholds. It plots the true positive rate against the false positive rate. A higher AUC value indicates better overall performance across all possible thresholds. This metric is especially helpful when classes are imbalanced or when evaluating models without fixing a specific threshold.

$$\text{AUC-ROC} = \int_0^1 \text{TPR}(FPR)\, d(\text{FPR})$$

4.2.4 Experimental Results

To evaluate the performance of the proposed glaucoma classification framework, we conducted a series of experiments focusing on optic cup and optic disc segmentation, as well as glaucoma classification.

4.2.4.1 Optic Disc and Optic Cup Segmentation

This study employed two publicly available datasets: ORIGA and REFUGE. Both were already split into training and testing subsets using an 80:20 ratio. The ORIGA dataset includes 520 CLAHE-enhanced images with corresponding segmentation masks for training, and 130 images for testing. The REFUGE dataset contains 960 training images and 240 test images.

We fine-tuned the SAM2 model throughout 200 epochs and assessed its performance using commonly used evaluation metrics, including the Dice score and the Intersection over Union (IoU). A summary of the hyperparameters used to fine-tune SAM2 is provided in Table 4.2.

a) ORIGA Dataset Results

Figure 4.8 illustrates the training dynamics of SAM2 model on the ORIGA dataset for optic disc segmentation. The training loss drops quickly and stabilizes below 0.1,

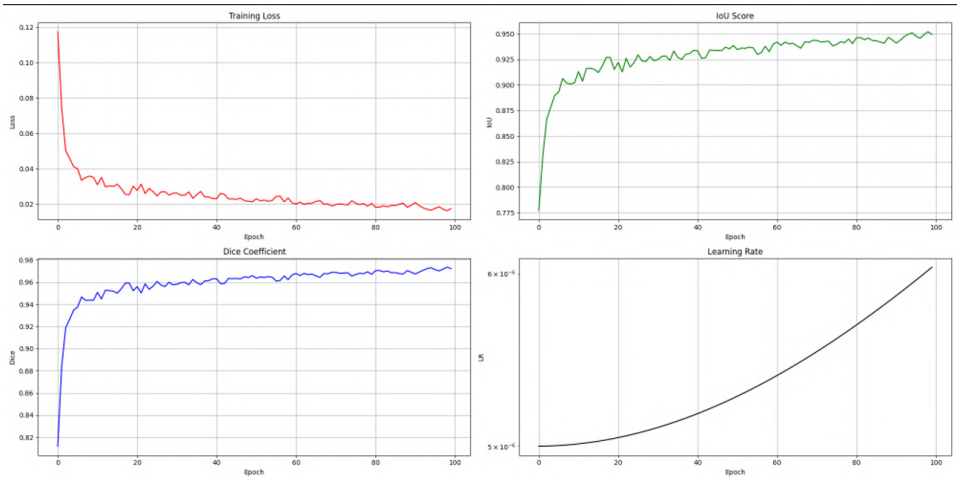


Figure 4.8: Fine-tuning performances of Segment Anything Model 2 on the ORIGA dataset for optic disc segmentation.

suggesting effective learning. Both IoU and Dice coefficients rise rapidly, surpassing 0.9 and eventually reaching close to 0.98, indicating strong segmentation performance. The learning rate increases gradually during training, consistent with a warm-up schedule, and no signs of overfitting or underfitting are observed.

For optic cup segmentation, Figure 4.9 shows similar trends. The loss continues to decline steadily, and both IoU and Dice scores stabilize at high values early in training. This suggests consistent and reliable convergence. The learning rate rises smoothly throughout the 200 epochs, supporting stable optimization.

Table 4.3 presents a comparison between our fine-tuned SAM2 model and several

Table 4.2: Fine-tuning Hyperparameters for SAM2 Optic Disc and Cup Segmentation

Parameter	**Optic Disc**	**Optic Cup**
Epochs	100	200
Iterations per Epoch	100	100
Base Learning Rate	1×10^{-4}	1×10^{-4}
Mask Decoder LR	2×10^{-4}	2×10^{-4}
Prompt Encoder LR	1×10^{-4}	1×10^{-4}
Optimizer	Lion	Lion
Weight Decay	0.025	0.025
Scheduler	OneCycleLR	OneCycleLR
Warmup Phase	15%	15%
Early Stopping Patience	25	25
Points per Mask	3–7	3–7
Gradient Clipping	1.0	1.0

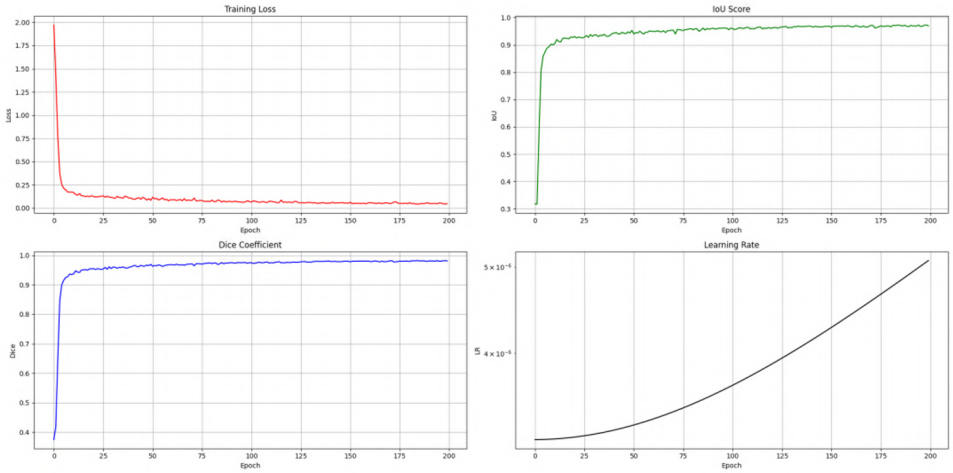


Figure 4.9: Fine-tuning performances of Segment Anything Model 2 on the ORIGA dataset for optic cup segmentation.

Table 4.3: Segmentation performance comparison on ORIGA dataset

Method	ORIGA optic disc		ORIGA optic cup	
	DICE	IoU	DICE	IoU
U-Net [83]	91.80	87.13	82.03	69.94
EnsDiff [83]	95.76	91.88	85.02	73.94
UTNet [83]	94.57	89.70	84.49	73.15
Res_UTNet [83]	93.60	87.98	83.45	71.60
AGNet [83]	94.91	90.56	83.73	73.29
Att U-Net [83]	96.01	92.33	85.05	73.99
GlaucoNet [84]	93.81	–	88.92	–
U-Net+ edge detect. [85]	96.2	92.8	87.1	77.3
JointRCNN [85]	96.2	92.8	87.1	77.3
FC DenseNet	**96.53**	**93.34**	86.59	76.88
fine-tuned SAM2	**95.84**	**92.21**	**95.89**	**92.24**

state-of-the-art methods for optic disc and optic cup segmentation on the ORIGA dataset. These metrics include Dice coefficient (Dice), Intersection over Union (IoU), which are standard measures of segmentation accuracy.

For optic disc segmentation, our method achieved a Dice score of 95.84% and an IoU of 92.21%. Although slightly lower than FC-DenseNet, which obtained the highest Dice (96.53%) and IoU (93.34%), our results remain competitive. Notably, both U-Net+ edge detection and JointRCNN reported similar performance, suggesting that incorporating edge-aware features may improve optic disc segmentation.

In contrast, for optic cup segmentation, our fine-tuned SAM2 model significantly outperformed all other methods. It achieved the highest Dice score of 95.89% and an IoU of 92.24%, surpassing even FC-DenseNet by more than 9 percentage points in Dice. This indicates a stronger ability to accurately capture the smaller and more variable optic cup region.

Furthermore, the sensitivity reached 95.44%, reflecting high recall and showing that the model effectively identifies true positive pixels in optic cup segmentation. This is particularly important in medical imaging, where missing critical regions can lead to misdiagnosis.

These achieved results suggest that while the model performs well on optic disc segmentation, its main strength lies in optic cup segmentation—outperforming existing methods by a significant margin. This improvement may be attributed to the robust feature learning capabilities of SAM2, which appear especially beneficial when segmenting complex or less-defined anatomical structures such as the optic cup.

b) REFUGE Dataset Results

The fine tuning performance of the SAM2 model on the optic disc segmentation task on REFUGE dataset is illustrated in Figure 4.10. The training loss decreases rapidly and stabilizes around 0.03, suggesting efficient learning. The IoU score surpasses 0.9 early in training and continues to improve gradually. Meanwhile, the Dice coefficient reaches and remains near 0.96, indicating high segmentation accuracy. The learning rate increases steadily, reflecting a controlled schedule that supports stable training.

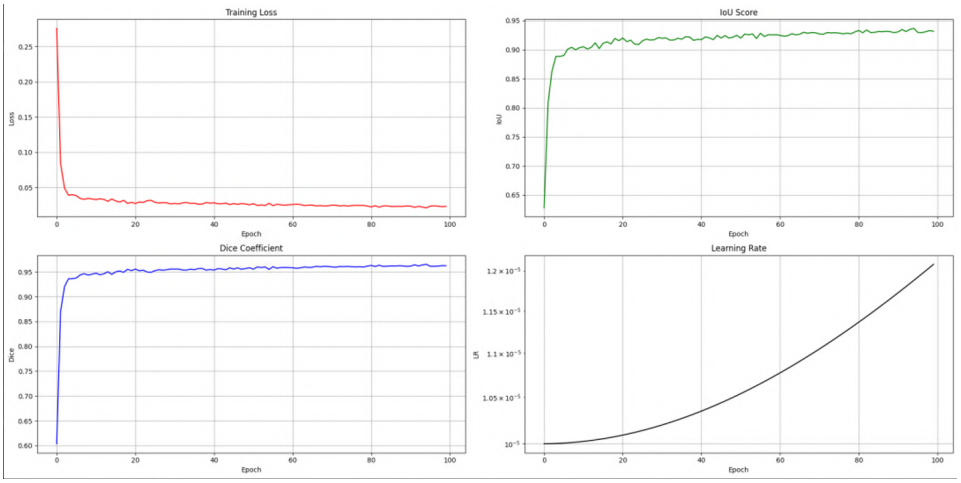


Figure 4.10: Fine-tuning performances of Segment Anything Model 2 on the REFUGE dataset for optic disc segmentation.

For optic cup segmentation, as shown in Figure 4.11, the observed trends are similar. Initially, the training loss spikes but quickly converges to near zero within the first 25 epochs and remains stable afterward, which suggests effective model adaptation. The IoU score rises from approximately 0.3 to 0.9 early in training and then plateaus, demonstrating strong segmentation performance. The Dice coefficient follows a comparable trend, approaching nearly 1.0 and maintaining high consistency. The learning rate increases gradually over time, improving training performance. Notably, this steady improvement indicates that the model effectively captures fine details of the optic cup. Furthermore, the consistent alignment between the IoU and Dice metrics reinforces the robustness of the segmentation process.

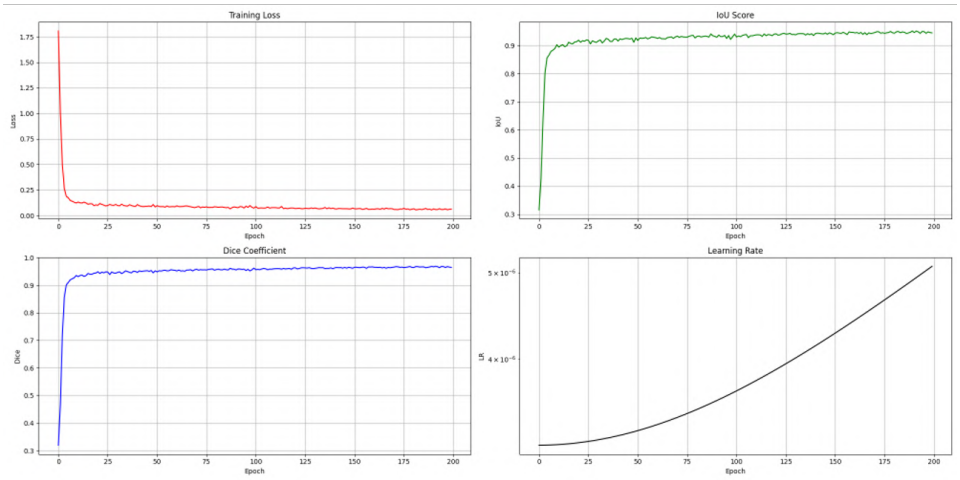


Figure 4.11: Fine-tuning performances of Segment Anything Model 2 on the REFUGE dataset for optic cup segmentation.

The segmentation performance comparison presented in Table 4.4 provides valuable insights into the strengths and limitations of various approaches for optic disc and optic cup segmentation. Specifically, the results highlight how our fine-tuned SAM2 model performs relative to state-of-the-art methods.

For optic disc segmentation, our model achieves a Dice score of 90.63 and an IoU of 83.39. These metrics are lower than those reported by top-performing methods such as OCDS-NSNP (Dice: 97.19, IoU: 95.45) and SLRS-Net (Dice: 96.5, IoU: 93.3). This discrepancy suggests that our model struggles with capturing precise optic disc boundaries. However, it is important to note that many high-performing methods rely on domain-specific optimizations or additional training data. For instance, OCDS-NSNP and SLRS-Net likely benefit from tailored architectures or extensive preprocessing. In contrast, our model operates without such task-specific adjustments, which enhances its generalizability but limits its specialization for this task. On the other hand, our model excels in optic cup segmentation, achieving a Dice score of 93.35 and an IoU of 87.77. These results surpass those of several competing methods, including EARDS

Table 4.4: Segmentation performance comparison on REFUGE dataset

Method	REFUGE optic disc		REFUGE optic cup	
	DICE	IoU	DICE	IoU
OCDS-NSNP [64]	97.19	95.45	93.87	86.72
SLRS-Net [63]	96.5	93.3	89.5	81.5
Multi-Glauc [66]	95.9	–	–	–
ECSD [67]	95.87	88.71	–	–
M_Ada [74]	95.85	–	88.25	–
EARDS [86]	95.49	91.47	88.72	80.17
pOSAL[87]	94.6	–	87.5	–
fine-tuned SAM2	**90.63**	**83.39**	**93.35**	**87.77**

(Dice: 88.72, IoU: 80.17) and pOSAL (Dice: 87.5). This indicates that our model is particularly effective at capturing the smaller and more complex structure of the optic cup. Interestingly, some methods, such as Multi-Glauc and ECSD, do not report complete metrics for both structures, limiting their comparability. Furthermore, methods like M_Ada and pOSAL omit IoU scores entirely, making it difficult to assess their overall performance. In contrast, our model provides full transparency by reporting both Dice and IoU metrics for both tasks.

A notable advantage of our model is its consistency across both segmentation tasks. While some methods achieve near-perfect results for one structure, they often fail to provide comparable performance for the other. For example, OCDS-NSNP achieves outstanding results for the optic disc but does not report any metrics for the optic cup. Similarly, Multi-Glauc reports only optic disc performance, leaving its capability for optic cup segmentation unknown. In this context, our model stands out for its balanced performance, delivering reliable results for both structures without requiring task-specific adjustments. This generalizability is particularly valuable in clinical settings, where models must handle diverse anatomical features without extensive customization. Despite its strengths, our model faces challenges in optic disc segmentation. The lower Dice and IoU scores suggest room for improvement in capturing the larger and more uniform structure of the optic disc. One potential avenue for future work is the integration of adaptive learning strategies. By dynamically adjusting to the specific characteristics of the optic disc, we could enhance segmentation accuracy while maintaining the model's generalizability. Additionally, further investigation into the impact of training data size and quality is warranted. Many high-performing methods may achieve superior results due to access to larger or more curated datasets. Exploring data augmentation techniques or unsupervised learning approaches could help bridge this gap.

In summary, the fine-tuned SAM2 model demonstrates competitive performance in optic cup segmentation while offering consistent results across both structures. Although it lags behind top-performing methods in optic disc segmentation, its generalizability

and transparency make it a strong candidate for real-world applications. Future research should focus on addressing its limitations through adaptive learning and improved data utilization strategies.

4.2.4.2 Glaucoma Classification

For the glaucoma classification, we employ the REFUGE dataset with 400 images for training and 400 images for validation which are used in our experiment as a test set, whereas the 400 images of REFUGE testing set are ignored in our experiments because they have not annotated since they were used to evaluate the winners during the competition. Table 4.5 summarises the hyperparameters used to fine-tune the Vision Transformer model.

During the ViT model finetuning, the REFUGE dataset demonstrates robust learning dynamics, as visualised in Figure 4.12. Notably, these results were achieved without data augmentation, highlighting the model's inherent design efficacy and regularization strategies. Training and validation loss curves exhibit smooth, stable convergence to low values (0.1–0.15), indicating minimal overfitting and strong generalization. This alignment reinforces the model's capacity to balance complexity and predictive reliability.

Performance metrics further validate this stability. Training and validation accuracy plateau at approximately 95%, reflecting effective knowledge transfer to unseen data. The F1 score stabilizes near 0.95, signaling balanced precision and recall, while recall itself rises steadily to $\tilde{0}.95$, underscoring reliable identification of positive cases. The ROC AUC metric peaks at 0.85, with the ROC curve (AUC 0.83) illustrating high true positive rates and low false positive rates across thresholds. Collectively, these results underscore the model's robustness in discrimination and predictive consistency.

The model exhibits robust training behavior and solid generalization. Training could have been stopped earlier, around epoch 40–50, as performance had already plateaued. However, continuing training caused no harm due to the absence of overfitting. These findings suggest that the model is well-regularized and appropriately complex for the given

Table 4.5: Fine-tuning hyperparameters for Vision Transformer model

Parameter	**value**
Epochs	100
Batch size	8
Learning Rate	1×10^{-4}
Optimizer	AdamW
Weight Decay	0.01
Scheduler	ReduceLRonPlateau

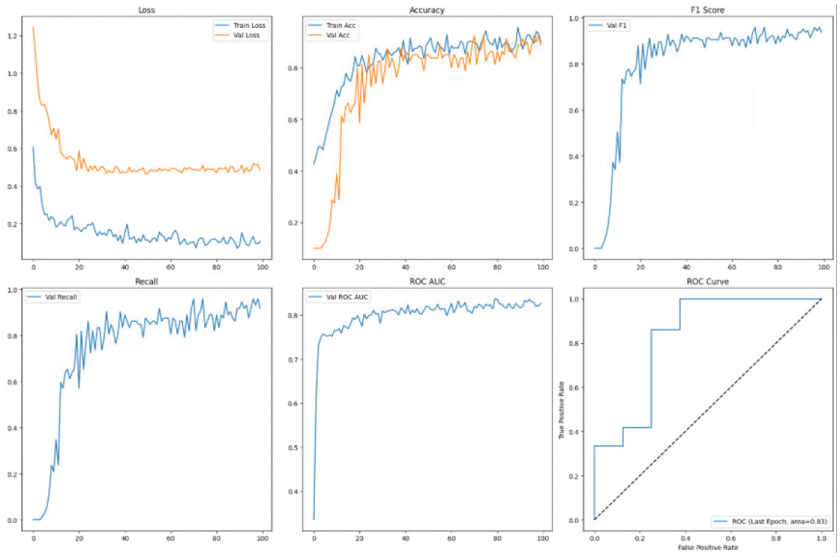
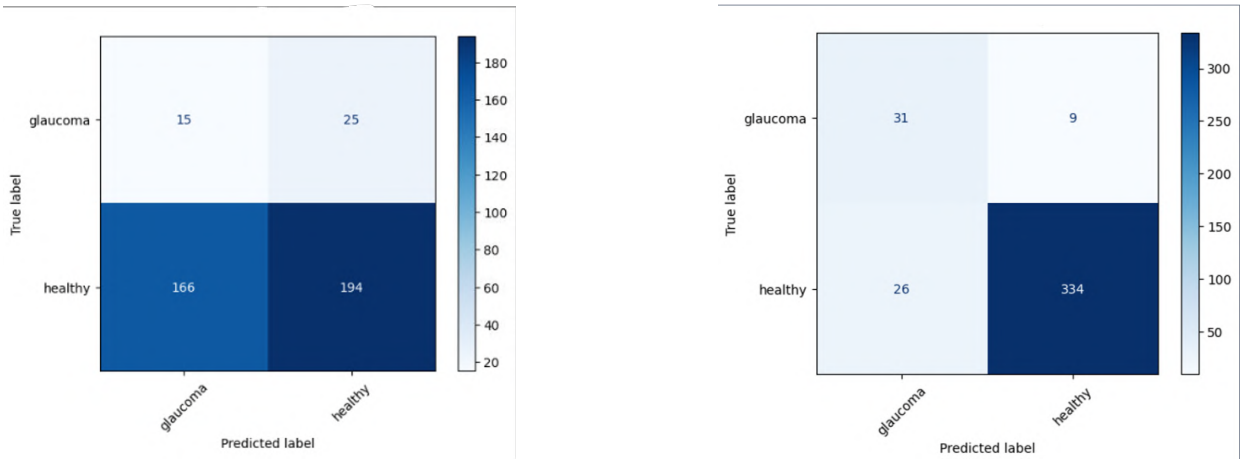


Figure 4.12: Finetuning performances of Vision Transformer model on REFUGE dataset.

dataset, Figure 4.13a illustrates the confusion matrix of the test set without OD/OC segmentation, and Figure 4.13b illustrates the confusion matrix of the test set with OD/OC segmentation.

The experimental results in Table 4.6 demonstrate the critical role of optic disc/optic cup (OD/OC) segmentation in enhancing the Vision Transformer (ViT) for glaucoma classification. Without segmentation, the fine-tuned ViT achieves low accuracy (52.25%), sensitivity (53.89%), and specificity (37.50%), despite high precision (88.58%). This indicates that, without structural guidance from OD/OC regions, the model struggles to



(a) (b)

Figure 4.13: Confusion Matrix of the REFUGE dataset test set: (a) without optic disc/cup segmentation and (b) with optic disc/cup segmentation

generalize effectively, misclassifying many positive and negative samples. The inflated precision suggests the model becomes overly conservative, predicting fewer positive cases with high certainty but failing to balance recall adequately.

In contrast, incorporating OD/OC segmentation significantly improves performance, achieving 91.25% accuracy, 97.39% precision, 92.78% sensitivity, and 77.5% specificity, alongside an F1-score of 95.02%. This highlights the importance of anatomical priors in guiding ViT's attention and enabling more accurate feature extraction. By focusing on clinically relevant regions, the model better distinguishes subtle pathological features from background noise. Compared to baseline CNN architectures like AlexNet and VGG16, the segmentation-based ViT outperforms in precision, sensitivity, and F1-score, demonstrating its strength in correctly identifying glaucoma cases and producing balanced predictions. However, it still lags behind the Hybrid ResNet50-Inception V3, which achieves 99.4% accuracy and superior specificity (99.1%), underscoring the robustness of ensemble CNN approaches in capturing both local and global image features.

This performance gap reveals a key trade-off: while transformers excel in modeling long-range dependencies and producing confident predictions, they are sensitive to training conditions and benefit greatly from task-specific cues like OD/OC segmentation. The poor performance of the non-segmentation ViT also underscores the limitations of global self-attention in medical imaging, where disease-relevant regions are highly localized. To address these challenges, future work should explore integrating data augmentation techniques (e.g., MixUp, elastic deformation) to improve generalization and reduce overfitting. Additionally, hybrid CNN-Transformer architectures could combine the spatial hierarchy modeling of CNNs with ViT's global attention. Self-supervised or contrastive learning strategies may further strengthen feature learning in limited data scenarios. These approaches could bridge the performance gap while leveraging ViT's strengths in precision and attention modeling, aligning with advancements in transformer-

Table 4.6: Classification results on the REFUGE dataset

	REFUGE				
Method	Accuracy	Precision	Sensitivity	Specificity	F1-score
AlexNet [88]	91.2	–	88.4	90.4	91.4
VGG16 [88]	93.2	–	91.3	93.2	92.4
ResNet50 [88]	95.2	–	93.6	96.1	95.4
Hybrid ResNet50-InceptionV3 [88]	99.4	–	98.1	99.1	99.2
fine-tuned Vit (no segmentation)	52.25	88.58	53.89	37.50	67.01
fine-tuned ViT (with segmentation)	**91.25**	**97.39**	**92.78**	**77.5**	**95.02**

based medical imaging frameworks.

4.3 Conclusion

In this chapter, we presented a full experimental evaluation of our automated glaucoma diagnosis framework, which combines image enhancement, advanced segmentation, and transformer-based classification. Our approach begins with CLAHE for contrast enhancement, ensuring that subtle anatomical features in fundus images are more visible and thus easier to analyze. Next, we employed a two-step segmentation using a fine-tuned Segment Anything Model 2 (SAM2), first isolating the optic disc, then the optic cup. This hierarchical strategy improved segmentation accuracy, especially for the optic cup, where our model achieved superior Dice and IoU scores compared to many recent methods on both the ORIGA and REFUGE datasets.

For the classification stage, we used a Vision Transformer (ViT) model trained on the neuroretinal rim region, which is a key biomarker for glaucoma. The ViT demonstrated strong performance, particularly in precision and F1-score, outperforming several traditional CNN-based models on the REFUGE dataset. However, while the model showed high sensitivity, its specificity was relatively lower, indicating challenges in correctly identifying healthy cases. This limitation is likely due to the absence of data augmentation, which could help the model generalize better to negative cases.

Overall, the results confirm the viability and efficiency of our framework for real-time glaucoma detection. The integration of SAM2 and ViT enables rapid and accurate analysis, making the method suitable for clinical settings where both speed and reliability are crucial. Nevertheless, there is room for improvement, especially in enhancing specificity. Future work should focus on incorporating data augmentation and exploring hybrid model architectures to further boost performance and address current limitations.

General Conclusion

This thesis addressed the urgent need for fast and reliable glaucoma diagnosis by designing a deep learning framework that automates the analysis of retinal fundus images. Glaucoma is a leading cause of irreversible blindness, and early detection is critical for preventing vision loss in millions of people worldwide. Manual screening is slow and requires expert knowledge, which is often unavailable in many regions. Therefore, automated solutions based on artificial intelligence are essential for scalable and accessible eye care.

The proposed framework integrates three main components. First, it uses Contrast-Limited Adaptive Histogram Equalization (CLAHE) to enhance image contrast, making subtle features like the optic disc and cup more visible. Next, a two-step segmentation process with a fine-tuned Segment Anything Model 2 (SAM2) isolates the optic disc and then the optic cup. This hierarchical approach ensures anatomical accuracy and reduces background noise, which is vital for reliable feature extraction. Finally, a Vision Transformer (ViT) model analyzes the neuroretinal rim (the region between the disc and cup) to classify images as glaucomatous or healthy. This focus on the rim area leverages clinical knowledge, as rim thinning is a key biomarker for glaucoma.

Experiments were conducted on two benchmark datasets, ORIGA and REFUGE, to validate the framework. The SAM2 model achieved Dice scores above 95% for both optic disc and cup segmentation on ORIGA, and over 90% on REFUGE, outperforming or matching many state-of-the-art methods. For classification, the ViT model reached an F1-score of 95.02% and a precision of 97.39% on the REFUGE dataset, surpassing traditional CNNs like AlexNet and VGG16 in these metrics. However, the specificity was about (77.5%), mainly due to the absence of data augmentation, which limited the model's ability to generalize to negative cases.

The strengths of this work are its efficiency, interpretability, and clinical relevance. The real-time inference capability of SAM2, combined with the ViT's global feature modeling, makes the system suitable for practical deployment in clinics where speed and accuracy are both crucial. The use of a two-step segmentation process ensures that the diagnostic features are anatomically meaningful and interpretable for clinicians. Moreover, by avoiding data augmentation, the results reflect the model's true capacity on real-world data, which also highlights the need for further improvements in specificity.

Despite these achievements, there are areas for future enhancement. Incorporating targeted data augmentation strategies, such as elastic transformations or MixUp, could help address class imbalance and improve specificity without compromising the integrity of the data. Exploring hybrid models that combine the strengths of CNNs and transformers may also close the performance gap with top ensemble methods. Additionally, expanding the system to handle multi-class diagnosis, such as distinguishing between different types of glaucoma, would increase its clinical utility.

Bibliography

[1] Paolo Brusini, Maria Letizia Salvetat, and Marco Zeppieri. How to measure intraocular pressure: An updated review of various tonometers. *Journal of Clinical Medicine*, 10(17):3860, August 2021.

[2] Adam Debrowski. Gonioscopy: Eye test details & results. `https://www.allaboutvision.com/eye-care/eye-tests/gonioscopy/`„ 2025. Image retrieved from All about vision.

[3] Ophthalmic Instruments Pte Ltd. Gonio prism / gonio. `https://www.ophthalmic.com.sg/product/gonio-prism/`, 2025. Image retrieved from Ophthalmic Instruments Singapore.

[4] Lauren J. Coan, Bryan M. Williams, Venkatesh Krishna Adithya, Swati Upadhyaya, Ala Alkafri, Silvester Czanner, Rengaraj Venkatesh, Colin E. Willoughby, Srinivasan Kavitha, and Gabriela Czanner. Automatic detection of glaucoma via fundus imaging and artificial intelligence: A review. *Survey of Ophthalmology*, 68(1):17–41, January 2023.

[5] W DREXLER and J FUJIMOTO. State-of-the-art retinal optical coherence tomography. *Progress in Retinal and Eye Research*, 27(1):45–88, January 2008.

[6] Sidong Liu, Yuyi You, and Stuart L. Graham. *Artificial Intelligence and Glaucoma*, page 75–89. Springer Singapore, 2021.

[7] Oleksandr Kovalyk-Borodyak, Juan Morales-Sánchez, Rafael Verdú-Monedero, and José-Luis Sancho-Gómez. Glaucoma detection: Binocular approach and clinical data in machine learning. *Artificial Intelligence in Medicine*, 160:103050, February 2025.

[8] Neha Sharma, Vibhor Jain, and Anju Mishra. An analysis of convolutional neural networks for image classification. *Procedia Computer Science*, 132:377–384, 2018.

[9] 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. *CoRR*, abs/2010.11929, 2020.

[10] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollár, and Christoph Feichtenhofer. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*, 2024.

[11] Thomas P Quinn, Manisha Senadeera, Stephan Jacobs, Simon Coghlan, and Vuong Le. Trust and medical ai: the challenges we face and the expertise needed to overcome them. *Journal of the American Medical Informatics Association*, 28(4):890–894, December 2020.

[12] Jignyasa Sanghavi and Manish Kurhekar. An efficient framework for optic disk segmentation and classification of glaucoma on fundus images. *Biomedical Signal Processing and Control*, 89:105770, March 2024.

[13] I. Govindharaj, D. Santhakumar, K. Pugazharasi, S. Ravichandran, R. Vijaya Prabhu, and J. Raja. Enhancing glaucoma diagnosis: Generative adversarial networks in synthesized imagery and classification with pretrained mobilenetv2. *MethodsX*, 14:103116, June 2025.

[14] Krishna Santosh Naidana, Madhu Hasitha Manne, and Hema Yalavarthi. A deep retinal vision network for glaucoma classification. *Discover Applied Sciences*, 7(3), February 2025.

[15] Anand V. Mantravadi and Neil Vadhar. Glaucoma. *Primary Care: Clinics in Office Practice*, 42(3):437–449, 2015. Primary Care Ophthalmology.

[16] H A Quigley and A T Broman. The number of people with glaucoma worldwide in 2010 and 2020. *British Journal of Ophthalmology*, 90(3):262–267, 2006.

[17] Seth R. Flaxman, Rupert R. A. Bourne, Serge Resnikoff, Peter Ackland, Tasanee Braithwaite, Maria Vittoria Cicinelli, Aditi Das, Jost B. Jonas, John Keeffe, John H. Kempen, Janet Leasher, Hans Limburg, Kovin Naidoo, Konrad Pesudovs, Alex Silvester, Gretchen A. Stevens, Nina Tahhan, Tien Y. Wong, Hugh R. Taylor, and Vision Loss Expert Group of the Global Burden of Disease Study. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. *The Lancet Global Health*, 5(12):e1221–e1234, 2017.

[18] Stuart J McKinnon, Lawrence D Goldberg, Patti Peeples, John G Walt, and Thomas J Bramley. Current management of glaucoma and the need for complete therapy. *Am J Manag Care*, 14(1 Suppl):S20–S27, 2008.

[19] Sahu Mahendra Kumar. A review on glaucoma: causes, symptoms, pathogenesis &; treatment. *Journal of Clinical Research and Ophthalmology*, 11(1):001–004, March 2024.

[20] Helga Kolb. How the retina works: Much of the construction of an image takes place in the retina itself through the use of specialized neural circuits. *American scientist*, 91(1):28–35, 2003.

[21] Kishore Cholkar, Supriya Reddy Dasari, Dhananjay Pal, and Ashim K Mitra. Eye: anatomy, physiology and barriers to drug delivery. In *Ocular transporters and receptors*, pages 1–36. Elsevier, 2013.

[22] Ron Ofri. Intraocular pressure and glaucoma. *Veterinary Clinics of North America: Exotic Animal Practice*, 5(2):391–406, May 2002.

[23] Richard H Masland. The fundamental plan of the retina. *Nature neuroscience*, 4(9):877–886, 2001.

[24] Hans E Grossniklaus, Eldon E Geisert, and John M Nickerson. Introduction to the retina. *Progress in molecular biology and translational science*, 134:383–396, 2015.

[25] Juan J. Salazar, Ana I. Ramírez, Rosa De Hoz, Elena Salobrar-Garcia, Pilar Rojas, José A. Fernández-Albarral, Inés López-Cuenca, Blanca Rojas, Alberto Triviño, and José M. Ramírez. *Anatomy of the Human Optic Nerve: Structure and Function*. IntechOpen, February 2019.

[26] Suneeta Dubey and Monica Gandhi. Evaluation of the optic nerve head in glaucoma. *Journal of Current Glaucoma Practice*, 7(3):106–114, September 2013.

[27] M. Cristina Leske. Open-angle glaucoma—an epidemiologic overview. *Ophthalmic Epidemiology*, 14(4):166–172, January 2007.

[28] Isabella V. Wagner, Michael W. Stewart, and Syril K. Dorairaj. Updates on the diagnosis and management of glaucoma. *Mayo Clinic Proceedings: Innovations, Quality &; Outcomes*, 6(6):618–635, December 2022.

[29] Pratap Challa. Glaucoma genetics: Advancing new understandings of glaucoma pathogenesis. *International Ophthalmology Clinics*, 44(2):167–185, 2004.

[30] Pratap Challa. Glaucoma genetics. *International Ophthalmology Clinics*, 48(4):73–94, 2008.

[31] RajulS Parikh, ShefaliR Parikh, Shoba Navin, Ellen Arun, and Ravi Thomas. Practical approach to medical management of glaucoma. *Indian Journal of Ophthalmology*, 56(3):223, 2008.

[32] Feng Gao, Jiajian Wang, Junyi Chen, Xiaolei Wang, Yuhong Chen, and Xinghuai Sun. Etiologies and clinical characteristics of young patients with angle-closure glaucoma: a 15-year single-center retrospective study. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 259(8):2379–2387, April 2021.

[33] Nadeem Hafeez Butt, Muhammad Hammad Ayub, and Muhammad Hassaan Ali. Challenges in the management of glaucoma in developing countries. *Taiwan Journal of Ophthalmology*, 6(3):119–122, September 2016.

[34] L. P. Cohen and L. R. Pasquale. Clinical characteristics and current treatment of glaucoma. *Cold Spring Harbor Perspectives in Medicine*, 4(6):a017236–a017236, June 2014.

[35] Shibal Bhartiya, Ritu Gadia, Harinder S Sethi, and Anita Panda. Clinical evaluation of optic nerve head in glaucoma. *Current Journal of Glaucoma Practice with DVD*, 4:115–132, September 2010.

[36] Malaya Kumar Nath and Samarendra Dandapat. Techniques of glaucoma detection from color fundus images: A review. *International Journal of Image, Graphics and Signal Processing*, 4(9):44–51, September 2012.

[37] Michael D. Abramoff, Mona K. Garvin, and Milan Sonka. Retinal imaging and image analysis. *IEEE Reviews in Biomedical Engineering*, 3:169–208, 2010.

[38] J Gloster and D G Parry. Use of photographs for measuring cupping in the optic disc. *British Journal of Ophthalmology*, 58(10):850–862, October 1974.

[39] Jagadish Nayak, Rajendra Acharya U., P. Subbanna Bhat, Nakul Shetty, and Teik-Cheng Lim. Automated diagnosis of glaucoma using digital fundus images. *Journal of Medical Systems*, 33(5):337–346, August 2008.

[40] Yuki Hagiwara, Joel En Wei Koh, Jen Hong Tan, Sulatha V. Bhandary, Augustinus Laude, Edward J. Ciaccio, Louis Tong, and U. Rajendra Acharya. Computer-aided diagnosis of glaucoma using fundus images: A review. *Computer Methods and Programs in Biomedicine*, 165:1–12, October 2018.

[41] Aljoscha S. Neubauer, Tina R. Krieglstein, Christos Chryssafis, Martin Thiel, and Anselm Kampik. Comparison of optical coherence tomography and fundus photography for measuring the optic disc size. *Ophthalmic and Physiological Optics*, 26(1):13–18, December 2005.

[42] Niharika Thakur and Mamta Juneja. Classification of glaucoma using hybrid features with machine learning approaches. *Biomedical Signal Processing and Control*, 62:102137, September 2020.

[43] Fei Li, Deming Wang, Zefeng Yang, Yinhang Zhang, Jiaxuan Jiang, Xiaoyi Liu, Kangjie Kong, Fengqi Zhou, Clement C. Tham, Felipe Medeiros, Ying Han, Andrzej Grzybowski, Linda M. Zangwill, Dennis S.C. Lam, and Xiulan Zhang. The ai revolution in glaucoma: Bridging challenges with opportunities. *Progress in Retinal and Eye Research*, 103:101291, November 2024.

[44] Alexander K. Schuster, Carl Erb, Esther M. Hoffmann, Thomas Dietlein, and Norbert Pfeiffer. The diagnosis and treatment of glaucoma. *Deutsches Ärzteblatt international*, March 2020.

[45] Debjit Bhowmik, KP Sampath Kumar, Lokesh Deb, Shravan Paswan, and AS Dutta. Glaucoma-a eye disorder its causes, risk factor, prevention and medication. *The Pharma Innovation*, 1(1, Part A):66, 2012.

[46] A. L. Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3):210–229, 1959.

[47] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015.

[48] Ramandeep Singh, Weiwen Wu, Ge Wang, and Mannudeep K. Kalra. Artificial intelligence in image reconstruction: The change is here. *Physica Medica*, 79:113–125, November 2020.

[49] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, December 2017.

[50] Mohamed Dirani, Jonathan G Crowston, Penny S Taylor, Peter T Moore, Sophie Rogers, M Lynne Pezzullo, Jill E Keeffe, and Hugh R Taylor. Economic impact of primary open-angle glaucoma in australia: Economic impact of glaucoma. *Clinical & Experimental Ophthalmology*, 39(7):623–632, June 2011.

[51] Paul J. Foster. The prevalence of glaucoma in chinese residents of singapore: A cross-sectional population survey of the tanjong pagar district. *Archives of Ophthalmology*, 118(8):1105, August 2000.

[52] An Ran Ran, Clement C. Tham, Poemen P. Chan, Ching-Yu Cheng, Yih-Chung Tham, Tyler Hyungtaek Rim, and Carol Y. Cheung. Deep learning in glaucoma with optical coherence tomography: a review. *Eye*, 35(1):188–201, October 2020.

[53] Yidong Chai, Hongyan Liu, and Jie Xu. Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models. *Knowledge-Based Systems*, 161:147–156, December 2018.

[54] Atalie C. Thompson, Alessandro A. Jammal, and Felipe A. Medeiros. A review of deep learning for screening, diagnosis, and detection of glaucoma progression. *Translational Vision Science & Technology*, 9(2):42, July 2020.

[55] Niall O'Mahony, Sean Campbell, Anderson Carvalho, Suman Harapanahalli, Gustavo Velasco Hernandez, Lenka Krpalkova, Daniel Riordan, and Joseph Walsh. *Deep Learning vs. Traditional Computer Vision*, page 128–144. Springer International Publishing, April 2019.

[56] Ajay Shrestha and Ausif Mahmood. Review of deep learning algorithms and architectures. *IEEE Access*, 7:53040–53065, 2019.

[57] K. FUKUSHIMA. A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol, Cybern*, 36:193/202, 1980.

[58] Emerald U. Henry, Onyeka Emebob, and Conrad Asotie Omonhinmin. Vision transformers in medical imaging: A review, 2022.

[59] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[60] Yuting He, Fuxiang Huang, Xinrui Jiang, Yuxiang Nie, Minghao Wang, Jiguang Wang, and Hao Chen. Foundation model for advancing healthcare: Challenges, opportunities, and future directions, 2024.

[61] Wasif Khan, Seowung Leem, Kyle B. See, Joshua K. Wong, Shaoting Zhang, and Ruogu Fang. A comprehensive survey of foundation models in medicine, 2025.

[62] Sangeetha J, Rekha D, and Priyanka M. A residual network integrated with multimodal fundus features for automatic glaucoma classification. *Computers and Electrical Engineering*, 122:109880, March 2025.

[63] Adnan Haider, Muhammad Arsalan, Min Beom Lee, Muhammad Owais, Tahir Mahmood, Haseeb Sultan, and Kang Ryoung Park. Artificial intelligence-based computer-aided diagnosis of glaucoma using retinal fundus images. *Expert Systems with Applications*, 207:117968, November 2022.

[64] Wang Li, Meichen Xia, Hong Peng, Zhicai Liu, and Jun Guo. Odcs-nsnp: Optic disc and cup segmentation using deep networks enhanced by nonlinear spiking neural p systems. *Biomedical Signal Processing and Control*, 108:107935, October 2025.

[65] Piyush Bhushan Singh, Pawan Singh, and Harsh Dev. Optimized convolutional neural network for glaucoma detection with improved optic-cup segmentation. *Advances in Engineering Software*, 175:103328, January 2023.

[66] Haoren Xiong, Fei Long, Mohammad S. Alam, and Jun Sang. Multi-glaucnet: A multi-task model for optic disc segmentation, blood vessel segmentation and glaucoma detection. *Biomedical Signal Processing and Control*, 99:106850, January 2025.

[67] Bingyan Liu, Daru Pan, Zhenbin Shuai, and Hui Song. Ecsd-net: A joint optic disc and cup segmentation and glaucoma classification network based on unsupervised domain adaptation. *Computer Methods and Programs in Biomedicine*, 213:106530, January 2022.

[68] Thisara Shyamalee and Dulani Meedeniya. Attention u-net for glaucoma identification using fundus image segmentation. In *2022 International Conference on Decision Aid Sciences and Applications (DASA)*, page 6–10. IEEE, March 2022.

[69] C. Rekha and K. Jayashree. Optic disc and cup segmentation methods for glaucoma detection using twin- inception transformer hinge attention network with cycle consistent convolutional neural network. *Biomedical Signal Processing and Control*, 107:107844, September 2025.

[70] Angshuman Roy, Anuvab Sen, Soumyajit Gupta, Soham Haldar, Subhrajit Deb, Taraka Nithin Vankala, and Arkapravo Das. Deepeyenet: Adaptive genetic bayesian algorithm based hybrid convnexttiny framework for multi-feature glaucoma eye diagnosis, 2025.

[71] Satyabrata Lenka, Zefree Lazarus Mayaluri, and Ganapati Panda. Glaucoma detection from retinal fundus images using graph convolution based multi-task model. *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, 11:100931, March 2025.

[72] Pradeep Kumar, Pramod Rangaiah, and Robin Augustine. Enhanced glaucoma detection using u-net and u-net+ architectures using deep learning techniques, 2025. SSRN preprint.

[73] Vutukuru Venkata Naga Satish Kumar, G. Harinath Reddy, and M.N. GiriPrasad. A novel glaucoma detection model using unet++-based segmentation and resnet with gru-based optimized deep learning. *Biomedical Signal Processing and Control*, 86:105069, September 2023.

[74] Álvaro S. Hervella, José Rouco, Jorge Novo, and Marcos Ortega. End-to-end multi-task learning for simultaneous optic disc and cup segmentation and glaucoma classification in eye fundus images. *Applied Soft Computing*, 116:108347, February 2022.

[75] Abeer Aljohani and Rua Y. Aburasain. A hybrid framework for glaucoma detection through federated machine learning and deep learning models. *BMC Medical Informatics and Decision Making*, 24(1), May 2024.

[76] Douglas Abreu da Rocha, Ana Beatriz Lordeiro Barbosa, Daniela Silva Guimarães, Leonardo Miranda Gregório, Luiz Henrique Nacife Gomes, Luciana da Silva Amorim, and Zélia Myriam Assis Peixoto. An unsupervised approach to improve contrast and segmentation of blood vessels in retinal images using clahe, 2d gabor wavelet, and morphological operations. *Research on Biomedical Engineering*, 36(1):67–75, January 2020.

[77] Jucelino Cardoso Marciano dos Santos, Gilberto Arantes Carrijo, Cristiane de Fátima dos Santos Cardoso, Júlio César Ferreira, Pedro Moises Sousa, and

Ana Cláudia Patrocínio. Fundus image quality enhancement for blood vessel detection via a neural network using clahe and wiener filter. *Research on Biomedical Engineering*, 36(2):107–119, February 2020.

[78] Heng Li, Haojin Li, Mingyang Ou, Xiangyang Yu, Xiaoqing Zhang, Ke Niu, Huazhu Fu, and Jiang Liu. Fundus image quality assessment and enhancement: a systematic review. *arXiv preprint arXiv:2501.11520*, 2025.

[79] 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 (ICLR)*, 2021.

[80] Maithra Raghu, Chiyuan Zhang, Robert Goroshin, and Yann Lecun. Do vision transformers see like convolutional neural networks? *Advances in Neural Information Processing Systems*, 34:12137–12148, 2021.

[81] Huazhu Fu, Fei Li, José Ignacio Orlando, Hrvoje Bogunović, Xu Sun, Jingan Liao, Yanwu Xu, Shaochong Zhang, and Xiulan Zhang. Refuge: Retinal fundus glaucoma challenge, 2019.

[82] Zhuo Zhang, Feng Shou Yin, Jiang Liu, Wing Kee Wong, Ngan Meng Tan, Beng Hai Lee, Jun Cheng, and Tien Yin Wong. Origa-light: An online retinal fundus image database for glaucoma analysis and research. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE, August 2010.

[83] Zhouchen Lin, Ming-Ming Cheng, Ran He, Kurban Ubul, Wushouer Silamu, Hongbin Zha, Jie Zhou, and Cheng-Lin Liu, editors. *Pattern Recognition and Computer Vision: 7th Chinese Conference, PRCV 2024, Urumqi, China, October 18–20, 2024, Proceedings, Part XV*. Lecture Notes in Computer Science. Springer Nature Singapore, 2025.

[84] Rashmi Panda, N. B. Puhan, Bappaditya Mandal, and Ganapati Panda. Glauconet: Patch-based residual deep learning network for optic disc and cup segmentation towards glaucoma assessment. *SN Computer Science*, 2(2), February 2021.

[85] Srikanth Tadisetty, Ranjith Chodavarapu, Ruoming Jin, Robert J. Clements, and Minzhong Yu. Identifying the edges of the optic cup and the optic disc in glaucoma patients by segmentation. *Sensors*, 23(10):4668, May 2023.

[86] Wei Zhou, Jianhang Ji, Yan Jiang, Jing Wang, Qi Qi, and Yugen Yi. Eards: Efficientnet and attention-based residual depth-wise separable convolution for joint od and oc segmentation. *Frontiers in Neuroscience*, 17, March 2023.

[87] Shujun Wang, Lequan Yu, Xin Yang, Chi-Wing Fu, and Pheng-Ann Heng. Patch-based output space adversarial learning for joint optic disc and cup segmentation. *IEEE Transactions on Medical Imaging*, 38(11):2485–2495, 2019.

[88] Murali Govindan, Vinod Kumar Dhakshnamurthy, Kannan Sreerangan, Manikanda Devarajan Nagarajan, and Suresh Kumar Rajamanickam. A framework for early detection of glaucoma in retinal fundus images using deep learning. In *CC 2023*, CC 2023, page 3. MDPI, February 2024.