mmmmmmmmmmmPEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
Ministry of Higher Education and Scientific Research

UNIVERSITY OF ECHAHID HAMMA LAKHDAR - EL OUED

FACULTY OF EXACT SCIENCES
Computer Science Department

Thesis
submitted in partial fulfilment of the requirements for the degree of

ACADEMIC MASTER

Field: **Mathematics and Computer Science**
Option: **Computer Science**
Specialty: **Artificial Intelligence & Data Science**

Presented by :

- **Douna Malek**
- **Ladgham Douaa**

Title

An automatic and accurate recognition approach of glaucoma within fundus images based on machine learning and deep Learning.

Examination Commitee :

M.	Kertiou Ismail	MCA	President
M.	Settou Tarablesse	MCB	Examiner
M.	Hamoud Meriem	MCB	Supervisor
M.	Kholledi Nadjoua Houda	MAA	Co-Supervisor

Academic year: 2024-2025

شكر و تقدير

بسم الله الرحمن الرحيم

الحمد لله الذي بنعمته تتم الصالحات، وبفضله تتيسر الأمور وتُنال الغايات .

تتقدم بخالص الشكر و التقدير لكل من كان له دور في إنجاز هذا العمل .

ونخص بالشكر مشرفتنا الأستاذة حمود مريم ، على إشرافها العلمي وتوجيهها المنهجي خلال مشروع التخرج، وعلى حرصها المستمر على جودة العمل ومضمونه .

كما نوجه شكرا خاصا للدكتور علي سعدون على ما قدمه لنا من دعم مباشر و مساعدة فعالة خلال مراحل انجاز هذا المشروع .

أيضا لا يفوتنا ان نشكر جميع اساتذة قسم الاعلام الالي لما كان لهم الفضل في تكويننا و تزويدنا بالمعارف و المهارات التي مهدت لنا طريق النجاح.

إهداء

إلى والديَّ العزيزين، نبض القلب ونور الطريق،

إلى إخوتي وأخواتي، شركاء الروح والدرب،

إلى صديقاتي حبيباتي، رفيقات القلب، من خففن عني ثقل الأيام،

إلى أساتذتي الأفاضل، وكل من علّمني حرفًا، ومنحني من علمه، فارتقى بفكري ووسّع آفاقي،

إلى أستاذتي الغالية، الدكتورة **مريم حمود**، التي آمنت بي ودفعتني للأفضل، وإلى رفيقتي **هبة الله** التي رافقتني في

دراستي و تقاسمت معي الجهد و اللحظات الصعبة،

أهدي لكم ثمرة تعبي، ونتاج سنواتي،

أهدي لكم هذا النجاح. . . بفضل الله ثم دعمكم اكتمل .

وإلى روح من كانت النور في بداياتي،

إلى معلمتي الراحلة **نور الهدى**، التي غرست في قلبي حب العلم والأمل،

أهديك هذا الإنجاز، دعائي يسبق كلماتي،

رحمك الله وجعل الجنة مستقرك ومثواك.

ملاك

إهداء

إلى أبي الغالي، الذي علّمني أن التعب لا يضيع، وأن الصبر يصنع الفرق،

إلى أمي الحنونة، التي كانت دعواتها سر راحتي ونجاحي،

إلى إخوتي وأخواتي، أنتم العائلة التي أفتخر بها ، وجودكم حولي أعطاني القوة والدعم،

إلى صديقتي الغالية **أنفال**، رفيقة الدرب، وشريكة الأيام الجميلة والصعبة،

إلى شريكتي **ملاك**، شكرًا من القلب على دعمك ومواقفك الجميلة،

إلى أستاذتي **حمود مريم** ، شكرًا على دعمك وتوجيهاتك التي كانت مفتاحًا لنجاحي،

وإلى كل الأساتذة الذين علّموني ووقفوا معي، لكم جزيل الشكر والاحترام،

لكل من كان له أثر في هذا الإنجاز، أهدي هذا النجاح عربون شكر وامتنان.

دعاء

Abstract

Glaucoma is a progressive neurodegenerative disease and remains one of the leading causes of irreversible, yet largely preventable, blindness worldwide. Vision loss in glaucoma is frequently attributed to delayed diagnosis and inadequate therapeutic intervention. Consequently, early screening is crucial for initiating timely treatment, thereby preserving visual function and improving patients' quality of life. Optical Coherence Tomography (OCT) facilitates the assessment of retinal nerve fiber layer (RNFL) thickness by exploiting the optical reflectivity properties of retinal tissues. RNFL thickness, typically measured from volumetric OCT scans acquired in a circular configuration around the optic disc, serves as a key biomarker for the early detection of individuals at risk of developing glaucoma.

In this study, we propose a deep learning-based dual-branch glaucoma grading system that integrates fundus images and OCT volumes, mimicking the clinical diagnostic workflow employed by ophthalmologists. deep learning technologies is actually considered as a cutting-edge approach for medical image analysis., offering promising potential for automated disease diagnosis. The proposed system comprises four interdependent stages. First, fundus images are segmented using the SegFormer model to efficiently delineate key anatomical structures, including blood vessels, the optic disc, and the optic cup. Second, Optical Coherence Tomography (OCT) images are processed using Gabor filters to extract texture features across multiple orientations, thereby capturing the intrinsic textural characteristics of the retinal layers. Third, clinically significant features such as vessel thickness and the cup-to-disc ratio (CDR) are computed. Finally, these extracted features are input into a classification module based on the SAINT model, which leverages attention mechanisms to produce diagnostic labels aligned with clinical standards, categorizing cases as normal, early-stage, or progressive glaucoma.

Experimental results demonstrated that the model achieved high classification accuracy of 94.27%, effectively distinguishing between different stages of glaucoma. These findings underscore the potential of the system to assist clinicians in making timely and informed decisions, particularly in the early detection and management of glaucoma.

key words

Glaucoma, Deep Learning,OCT, Fundus Image, SegFormer, SAINT.

Résumé

Le glaucome est une maladie neurodégénérative progressive et demeure l'une des principales causes de cécité irréversible, mais largement évitable, dans le monde. La perte de vision dans le glaucome est souvent attribuée à un diagnostic tardif et à une intervention thérapeutique inadéquate. Par conséquent, le dépistage précoce est crucial pour initier un traitement en temps opportun, préservant ainsi la fonction visuelle et améliorant la qualité de vie des patients. La Tomographie par Cohérence Optique (OCT) facilite l'évaluation de l'épaisseur de la couche des fibres nerveuses rétiniennes (RNFL) en exploitant les propriétés de réflectivité optique des tissus rétiniens. L'épaisseur de la couche des fibres nerveuses de la rétine (RNFL), généralement mesurée à partir de scans OCT volumétriques acquis en configuration circulaire autour du disque optique, sert de biomarqueur clé pour la détection précoce des individus à risque de developer un glaucome.

Dans cette étude, nous proposons un système de classification du glaucome à double branche basé sur l'apprentissage profond qui intègre des images du fond d'œil et des volumes OCT, imitant le flux de travail diagnostique clinique utilisé par les ophtalmologistes. les technologies d'apprentissage profond sont en réalité considérées comme une approche de pointe pour l'analyse des images médicales, offrant un potentiel prometteur pour le diagnostic automatisé des maladies. Le système proposé comprend quatre étapes interdépendantes. Tout d'abord, les images du fond d'œil sont segmentées à l'aide du modèle SegFormer pour délimiter efficacement les structures anatomiques clés, y compris les vaisseaux sanguins, le disque optique et la cupule optique. Deuxièmement, les images de Tomographie par Cohérence Optique (OCT) sont traitées à l'aide de filtres de Gabor pour extraire des caractéristiques texturales dans plusieurs orientations, capturant ainsi les caractéristiques texturales intrinsèques des couches rétiniennes. Troisièmement, des caractéristiques cliniquement significatives telles que l'épaisseur des vaisseaux et le rapport cup-to-disc (CDR) sont calculées. Enfin, ces caractéristiques extraites sont introduites dans un module de classification basé sur le modèle SAINT, qui utilise des mécanismes d'attention pour produire des étiquettes diagnostiques conformes aux normes cliniques, catégorisant les cas en glaucome normal, précoce ou progressif.

Les résultats expérimentaux ont démontré que le modèle a atteint une grande précision de classification de 94.27% , distinguant efficacement les différents stades du glaucome. Ces résultats soulignent le potentiel du système à aider les cliniciens à prendre des décisions éclairées et opportunes, en particulier dans la détection précoce et la gestion du glaucome.

Mots clés

Glaucome, Apprentissage Profond,OCT, Image du Fond d'Œil, SegFormer,SAINT.

ملخص

الجلوكوما هو مرض تنكسي عصبي تقدمي وتظل واحدة من الأسباب الرئيسية للعمى غير القابل للرجوع، والذي يمكن الوقاية منه إلى حد كبير، في جميع أنحاء العالم. فقدان البصر في الجلوكوما يرجع غالبًا إلى التشخيص المتأخر والتدخل العلاجي غير الكافي. وبالتالي، فإن الفحص المبكر ضروري لبدء العلاج في الوقت المناسب، مما يحافظ على الوظيفة البصرية ويحسن جودة حياة المرضى. يساعد التصوير المقطعي البصري (OCT) في تقييم سمك طبقة الألياف العصبية للشبكية (RNFL) من خلال استغلال الخصائص الانعكاسية البصرية لأنسجة الشبكية. سماكة طبقة الألياف العصبية للشبكية تُقاس عادةً من مسح OCT الحجمي الذي يُكتسب بتكوين دائري حول القرص البصري، حيث تُعتبر علامة حيوية رئيسية للكشف المبكر عن الأفراد المعرضين لخطر الإصابة بالجلوكوما.

في هذه الدراسة، نقترح نظام تصنيف مزدوج الفرع قائم على التعلم العميق لتصنيف الجلوكوما يدمج صور قاع العين وحجوم التصوير المقطعي البصري، مقلدًا سير العمل التشخيصي السريري الذي يستخدمه أطباء العيون. تُعتبر تقنيات التعلم العميق في الواقع نهجًا متقدمًا لتحليل الصور الطبية، مما يوفر إمكانيات واعدة لتشخيص الأمراض بشكل آلي. يتكون النظام المقترح من أربع مراحل مترابطة. أولاً، يتم تقسيم صور قاع العين باستخدام نموذج SegFormer لتحديد الهياكل التشريحية الرئيسية بكفاءة، بما في ذلك الأوعية الدموية، القرص البصري، والكأس البصري. ثانيًا، يتم معالجة صور التصوير المقطعي البصري (OCT) باستخدام مرشحات Gabor لاستخراج ميزات النسيج عبر اتجاهات متعددة، مما يتيح التقاط الخصائص النسيجية الجوهرية لطبقات الشبكية. ثالثًا، يتم حساب الميزات السريرية الهامة مثل سمك الأوعية ونسبة الكأس إلى القرص (CDR). أخيرًا، يتم إدخال هذه الميزات المستخرجة في وحدة تصنيف تعتمد على نموذج SAINT الذي يستخدم آليات الانتباه لإنتاج تسميات تشخيصية تتماشى مع المعايير السريرية، مصنفة الحالات كطبيعية أو في مرحلة مبكرة أو زرق متقدم.

أظهرت النتائج التجريبية أن النموذج حقق دقة تصنيف عالية تقدر ب 94.27 بالمئة, تؤكد هذه النتائج على إمكانيات النظام في مساعدة الأطباء في اتخاذ قرارات سريعة ومستنيرة, لا سيما في الكشف المبكر وإدارة مرض الجلوكوما.

الكلمات المفتاحية

الجلوكوما, التعلم العميق, التصوير المقطعي البصري, صورة قاع العين, SegFormer , SAINT.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
OCT	Optical Coherence Tomography
RNFL	Retinal Nerve Fiber Layer
CDR	Cup-to-Disc Ratio
SAINT	Self-Attention and Intersample Attention Transformer
RCGs	Retinal Ganglion Cells
WHO	World Health Organization
CAD	Computer-Aided Diagnosis
CADe	Computer-Aided Detection
CADx	Computer-Aided Diagnosis
ML	Machine Learning
LR	Logistic Regression
KNN	K Nearest Neighbor
SVM	Support Vector Machine
DL	Deep Learning
ANNs	Artificial Neural Networks

ReLU	Rectified Linear Unit
DNNs	Deep Neural Networks
CNN	Convolutional Neural Network
PCA	Principal Component Analysis
RNNs	Recurenrt Neural Networks
NLP	Natural Language Processing
ViT	Vision Transformer
MLP	Multi Layer Perceptron
CoTr	Convolutional Transformer
HiFormer	Hierarchical Multi-Scale Representations
IOP	Intraocular Pressure
NTG	Normal tension Glaucoma
NRR	Neuroretinal Rim
GCC	Ganglion Cell Complexity
HRT	Heidelberg Retinal Tomography
vCDR	vertical Cup-to-Disc Ratio
OD	Optic Disc
CLAHE	Contrast-Limited Adaptive Histogram Equalisation
NLM	Non Local Means
TP	True Positive
FP	false Positive
TN	True Negative
FN	False Negative
IoU	Intersection Over Union

GENERAL INTRODUCTION

Glaucoma, also known as the silent thief of sight, is a progressive optic neuropathy characterized by irreversible damage to the optic nerve and the gradual structural loss of retinal ganglion cells (RGCs). This degeneration leads to a progressive decline in the visual field, typically beginning in the mid-peripheral region and advancing until only a central or peripheral island of preserved vision remains [14]. According to the World Health Organization (WHO), glaucoma is one of the leading causes of irreversible blindness worldwide, ranking second after cataracts [15]. In 2013, the global prevalence of glaucoma among individuals aged 40 to 80 years was estimated at 64.3 million, with projections increasing to 76.0 million by 2020 and 111.8 million by 2040 [16]. This alarming rise underscores the need for increased surveillance, early detection strategies, and targeted interventions to reduce its global burden.

Furthermore, significant disparities in glaucoma prevalence exist across geographic regions and age groups, suggesting the influence of genetic, environmental, and socioeconomic factors on disease progression[17] . Glaucoma often progresses asymptomatically, remaining undiagnosed until reaching an advanced stage, which implies that the actual number of affected individuals is likely much higher than reported cases [18]. In addition to visual field loss, patients may experience functional impairments such as reduced contrast sensitivity, altered color perception, and difficulty reading, further impacting their quality of life [14].

Given these challenges, artificial intelligence (AI) has emerged as a promising tool in the early and accurate detection of glaucoma. The ultimate goal is to develop advanced deep learning and machine learning models utilizing both fundus images and optical coherence tomography (OCT) volumes to enhance diagnostic accuracy and enable early intervention. Building upon studies that highlight the influence of ethnicity and geographic factors on the prevalence and progression of glaucoma[17], our research focuses on improving glaucoma diagnosis in the Maghreb region by incorporating ethnicity-specific data. To achieve this, we utilize a dataset collected from the Cooperative Hospital Foundation of Cuba -El Oued, facilitating the development of AI-driven diagnostic models tailored to regional populations. This approach aims to bridge healthcare gaps, optimize screening processes, and ultimately reduce the risk of irreversible blindness caused by late-stage glaucoma detection.

To address the challenges associated with early-stage glaucoma diagnosis, we developed a comprehensive methodology based on artificial intelligence techniques, particularly deep learning, through a multi-stage diagnostic pipeline.Initially, we collected a database consisting of OCT reports and based on fundus images and OCT scans extracted from OCT reports, we performed segmentation of the optic disc, cup, and blood vessels to extract the most reliable clinical features for medical diagnosis, CDR and the thickness of the blood vessels. As for the OCT scan

image, it was passed through a Gabor filter to extract the texture features. Additionally, we relied on accompanying clinical information which are the thickness of the retinal nerve fiber layer (RNFL), age, and gender. All these characteristics were integrated as inputs for the SAINT classification model, which was trained to provide an accurate and reliable diagnosis of the eye condition. The model ultimately classifies the input cases into one of three categories: normal, early-stage, or progressive glaucoma, contributing to improved diagnostic accuracy, reduced human intervention, and supporting doctors in making early and effective treatment decisions.

As a result of this work, we achieved highly competitive performance, with an overall accuracy of 94.27%. This result outperformed several state-of-the-art approaches in the task of multi-class glaucoma classification, where the eye condition is categorized into one of three clinically recognized classes: normal eye, glaucoma suspect, and glaucomatous eye. This classification aligns with real-world diagnostic practices, thereby increasing the practical value of our system.

The robustness and reliability of the proposed approach were further confirmed through clinical validation. Experts at the Ophthalmology Hospital of Cuba - El Oued reviewed the system's performance and approved its relevance and potential for integration into actual clinical workflows. Their endorsement highlights the practical feasibility of our model and its capacity to support ophthalmologists in making early, accurate, and objective glaucoma diagnoses, potentially contributing to the prevention of irreversible vision loss.

The structure of this thesis is as follow:

chapter I :*Image Processing and Artificial Intelligence in Healthcare: Concepts and Applications*

Throughout the first chapter, we will explore the fundamental concepts of image analysis and computer vision, extending towards machine learning, deep learning, and their applications in medical image processing.

chapter II :*Artificial Intelligence Paradigms for Automated Glaucoma Diagnosis: Challenges and Literature Review*

In this chapter, we will clarify the disease of glaucoma by studying its causes, types, and how it develops within the eye. We will also provide a detailed explanation of the OCT device and its pivotal role in the early detection of changes in the optic nerve and retina. Then we will move on to present the scientific problem that drives this work concluding it with a comprehensive review of previous research in the field of artificial intelligence and glaucoma diagnosis.

chapter III : *The proposed approach for early glaucoma detection integrating both OCT vol-*

umes and fundus images

This chapter will present the diagnostic line followed in this study, which will rely on a set of intelligent processes for analyzing medical images. The method for segmenting fundus images using the SegFormer model to extract the cup, disc, and blood vessels will be explained, with the aim of calculating CDR and vessel thickness. It will also be explained how to use a Gabor filter to extract textural features from OCT images. Moreover, clinical information such as RNFL, age, and gender will be integrated with these features to feed the SAINT model, which will classify the cases. The model will be trained according to various improvements to increase its diagnostic accuracy.

chapter IV :*Experimental Results and Performance Evaluation*

We will present the results of the proposed system based on experimental performance evaluations. Performance measurement will be conducted using metrics such as accuracy, sensitivity, and specificity, and we will compare the results with current methods in the field. The results will be displayed using graphs and tables to compare the model's performance on different data sets.

CHAPTER I

IMAGE PROCESSING AND ARTIFICIAL INTELLIGENCE IN HEALTHCARE: CONCEPTS AND APPLICATIONS

I.1 Introduction

In the last several years, the world has witnessed unprecedented progress because of artificial intelligence (AI) that has touched sectors as diverse as the economy, industry and medicine. AI plays a significant role in medical image analysis and disease prediction at an early level, which assists in the improved treatment of patients.

Thanks to developments in artificial intelligence and more particularly deep learning, image classification and segmentation models can now successfully analyze scans such as optical coherence tomography (OCT) and visual field tests. These models are emerging as effective at detecting early signs of glaucoma, sometimes beating human assessment in terms of accuracy and consistency.

In this chapter, different models of AI applied in the analysis of ophthalmic images will be reviewed. Particular focus will be put on feature extraction approaches and their utilization in glaucoma classification, highlighting the importance of these advanced approaches.

I.2 Artificial Intelligence in Healthcare

Artificial intelligence (AI) is currently one of the fastest-growing fields in the world. The goal is to develop systems that can mimic human abilities such as learning, logical reasoning, and decision making. AI in medicine has a very long history from early rule-based expert systems to advanced machine learning algorithms that can read elaborate medical data. This innovation has highly improved the accuracy of diagnosis as well as patient care level. While earlier clinical AI instruments generally used rules based on hard coding, new development in machine and deep learning technology allows the application of AI to handle very large medical datasets. This is important in facilitating disease detection in early stages as well as devising personalized treatment procedures [19].

I.2.1 AI-driven decision support systems

Other than diagnosis, AI-based decision support systems play a pivotal role in healthcare today. They help practitioners identify patterns, predict disease progression, and create optimized treatment protocols from big data. For example AI-based image analysis has achieved physician-level diagnostic performance for disease detection such as diabetic retinopathy and cancer and improved early diagnosis and treatment outcomes[20].These capabilities do not only improve patient care but they also improve clinical effectiveness in that they do away with errors

in diagnosis while speeding up decision-making.

I.2.2 Computer-Aided Diagnosis (CAD)

Computer-Aided Diagnosis (CAD) refers to the use of computer-generated outputs to assist clinicians in making diagnostic decisions. Unlike fully automated diagnostic systems—where algorithms make the final decision—CAD serves as a supportive tool that enhances the efficiency and accuracy of medical professionals.

As one of the earliest applications of artificial intelligence in medicine, CAD has been extensively used in radiology. Initially, these systems depended on manually engineered features derived from expert domain knowledge. However, recent advances in machine learning have enabled automatic feature learning, allowing systems to identify significant patterns in medical images and enhance diagnostic accuracy [21].

CAD systems generally fall into two main categories:

- **Computer-Aided Detection (CADe):** Identifies regions in medical images that may indicate abnormalities, thereby reducing the likelihood of missed pathologies.
- **Computer-Aided Diagnosis (CADx):** Assists clinicians in analyzing and classifying detected abnormalities, leading to more accurate and informed diagnoses [21].

With the continuous evolution of artificial intelligence and deep learning, CAD systems are becoming increasingly sophisticated. They are contributing significantly to the early detection of diseases, improving diagnostic workflows, and ultimately enhancing patient outcomes across various medical fields [21].

I.2.3 Challenges and Ethical Considerations in AI Adoption

Despite the growth in AI capabilities in medicine, its performance will still be dependent on the variety and quality of the training data. Hospital to hospital discrepancies and biases may compromise model generalizability. Data acquisition standardization and incorporation of new methodologies such as the 'silver standard' or domain-specific optimization methods for the data are therefore central to ensuring safe and reliable AI systems in clinical use [19]. integration of AI in clinical workflows has major challenges. Some of them include concerns around data privacy restrictions on sharing data algorithmic opacity and the necessity for standardized frameworks to allow interoperability between several healthcare systems. Ethical concerns such as patient safety, AI model bias, and unintended consequences must be dealt with to achieve

trust in AI-based medical applications. Regulatory policies have an important influence on AI adoption, with significant differences between regions like the United States, Europe and China. Variations in AI governance, data protection regulations and clinical validation strategies determine the rate of AI integration and its potential influence on healthcare. Overcoming these issues is important to realize the full potential of AI to transform medical research improve treatment practices, and ultimately patient care [22].

I.3 Machine Learning and Deep Learning in Medical Imaging

Two of the most revolutionary medical imaging modalities are machine learning and deep learning that have significantly enhanced diagnostic sensitivity and aided medical decision-making by physicians. They enable the automatic evaluation of medical images via pattern and subtle abnormality detection that is hard to detect using the naked eye. They then enable early diagnosis and improved treatment planning of the majority of conditions, like glaucoma[23].

I.3.1 Fundamentals of Machine Learning (ML)

Machine Learning (ML) is a branch of Artificial Intelligence (AI) that enables systems to automatically learn from data and improve their performance without being explicitly programmed. Unlike traditional AI techniques, ML relies on statistical models to identify patterns and make predictions, allowing for more adaptive and data-driven decision-making [24].
ML has revolutionized the medical field by enhancing disease diagnosis, predicting medical emergencies, and personalizing treatment plans. It encompasses three primary learning paradigms:

- Supervised Learning: This approach involves training models on labeled datasets, making it particularly useful for tasks such as disease classification and medical image segmentation.
- Unsupervised Learning: This method identifies hidden structures within unlabeled data, facilitating anomaly detection and the grouping of similar medical conditions.
- Reinforcement Learning: Through a trial-and-error process, this paradigm enables systems to optimize decision-making, with notable applications in robotic surgery and adaptive treatment strategies. [24]

Several ML algorithms have demonstrated effectiveness in medical applications, particularly in image analysis:

- **Logistic Regression (LR)**

Logistic Regression (LR) is a supervised learning algorithm of machine learning that is primarily used for classification tasks. It estimates the probability that a given instance will belong to a given class by studying relationships between input features, Despite its name LR is actually a classification algorithm rather than a regression algorithm because it upgrades linear regression with a sigmoid function to bound outputs between 0 and 1 [25].

In medical imaging, LR is important in:

- Diagnosis of diseases through the estimation of the probability of their occurrence
- Correlating image derived biomarkers and patient outcomes [25].

In binary classification, LR applies a decision threshold often 0.5 where values above it are labeled one class (e.g :disease exists) and below it as the other (e.g: disease does not exist). Though LR is enjoyed for its interpretability and ease, it is not necessarily challenged by complex patterns like more advanced models like Random Forest or Deep Learning approaches[25].

- **Decision Tree**

Decision tree is a supervised learning nonparametric algorithm for regression and classification. It is a hierarchical tree-like structure consisting of a root node, internal nodes, leaf nodes and branches. Some advantages of decision trees are as follows:

- Interpretability and Explainability: Being hierarchical, decision trees allow full transparency in the decision-making process, and therefore are easy to interpret for users.
- Handling Mixed Data Types: They can efficiently work with both numerical and categorical data without any complex preprocessing. [26].

In fact, despite decision trees being many advantages, they also have some disadvantages. Some of the major disadvantages attributed to decision trees are:

- Complexity: Decision trees increase in complexity when they grow larger in size and become hard to interpret as well as control. Growth of branches at an increased rate is associated with the augmented computational expense and reduced efficiency.

- Increased expense: Because decision trees use a greedy search method of learning, they may be more costly to train than other algorithms.
- Overfitting problem: Decision trees overfit the training data and therefore perform poorly on new data. Pruning techniques reduce this by removing excessive complexity [26].

Despite such limitations, decision trees are an extremely helpful approach if well-optimized and utilized in combination with other approaches.

- **The K nearest neighbors (K-NN: K Nearest Neighbors)**

K-Nearest Neighbors (KNN) is a non-parametric supervised learning model utilized for regression, as well as classification. KNN makes predictions of data points as the class with the greatest number of counts among their K nearest neighbors within feature space. [27] KNN assumes nearest points are near each other, hence it is strong in non-linear decision boundaries. It does not have any training step, but it is computationally costly in high-dimensional data and big data. Some advantages the K-Nearest Neighbours algorithm is credited with include:

- Easy to Implement: As it is easy and accurate K-NN is one of the first classifiers that data science beginners learn.
- Robustness to New Data: The algorithm automatically adjusts when new training samples are included because it keeps all training data in memory.
- No Assumptions Regarding Data Distribution: K-NN never relies on any assumptions regarding data distribution as opposed to the linear models.
- Flexibility: It can model complicated decision boundaries like nonlinear and arbitrarily shaped boundaries.
- Minimal Hyperparameters: K-NN only needs to specify the value of k and a distance function, which is less complicated compared to other machine learning algorithms [27].

Although K-NN has many advantages, it also has some disadvantages. Its inability to scale makes it memory and computationally expensive, especially when handling large data, making processing time and storage to grow. In addition, it is also plagued by the curse of dimensionality where performance is degraded as the number of features grows, working to reduce classification accuracy, particularly in cases of small sample size. It also makes the model more susceptible to overfitting, particularly in the case of high dimensional data while highly generalized models have a tendency to cause underfitting resulting in less precise predictions.

- **Support Vector Machines (SVM)**

Support Vector Machine (SVM) is a machine learning algorithm that learns from examples to classify objects.

In the medical field, SVMs are effectively used to classify gene expression patterns obtained from tumor samples or body fluids, aiming to make diagnoses or predict disease progression. They are also used to classify DNA sequences, protein sequences, and mass spectrometry data [1]. The figure I.1 illustrates an example of how Support Vector Machines (SVM) are used to classify gene expression data for leukemia patients. The points in the diagram represent samples affected by acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML), based on the two-dimensional gene expression of two genes: MARCKSL1 and ZYX.

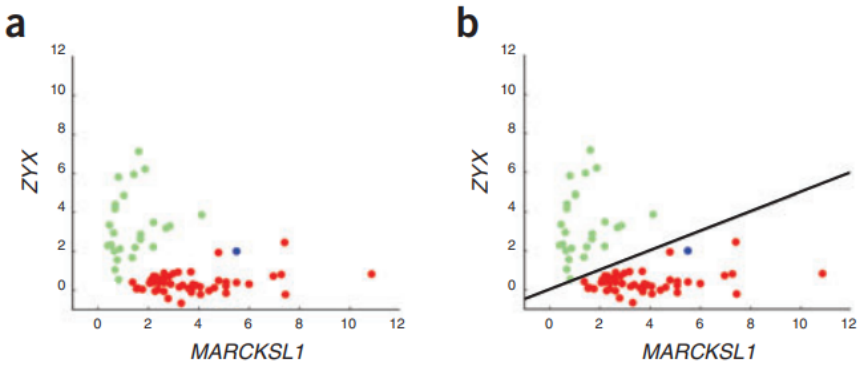

Figure I.1: (a) Gene expression data of two leukemia types. (b) SVM finds a line that separates the two groups and classifies the unknown sample[1]

.

Support Vector Machines (SVMs) attempt to find the best hyperplane that has the largest margin among different classes of data. Since the Banach space theory gives a strong mathematical framework whenever the theory is employed on high- or infinite-dimensional data, it can easily deal with the resulting hard optimization problems [28].

In such a setting, SVM's decision function that spits out the class labels based on the position of a point relative to the hyperplane can be analyzed rigorously. Banach space properties better understand this decision better, particularly with the capacity for generalization and interpretability of solutions in infinite-dimensional spaces [28].

Also, even though SVMs are linear classifiers in nature, kernel functions applied make them capable of handling non-linear problems. Banach space theory is a broader theoretical framework within which these kernel-based mappings can be explored [28].

Last but not least, SVMs can be trained by optimizing a convex optimization problem in order

to maximize the margin and minimize the classification error. Optimization can be formulated as a Banach space minimax problem, enabling us to benefit from advanced techniques like gradient-based methods and quadratic programming for an efficient discovery of optimal solutions [28].

SVMs possess several benefits that render them very well adapted to classification and regression problems:

- Very effective for regression and classification.
- Regularizing by nature, decreasing the likelihood of overfitting.
- Extremely extensible to multi-class classification through different approaches.
- Kernel functions are versatile and facilitate nonlinear classification.
- Model sparseness, making decisions using only support vectors, leading to it being useful for deployment [**?**].

I.3.2 Introduction to Deep Learning (DL)

Deep learning is a method of artificial intelligence that enables computers to process data in the same way as the human brain. Deep learning algorithms can identify complicated patterns in pictures, text, audio, and other types of data to make accurate predictions and insights. Deep learning algorithms can be used to automate tasks that typically require human intelligence, such as explaining pictures or transcribing audio recordings into text [29].
Deep learning technology powers the majority of the AI software utilized in everyday products. It is a branch of machine learning that employs artificial neural networks with multiple processing layers to learn data representations in a hierarchical fashion [29].

- **Neural Networks: Structure and Significance**

At deep learning are artificial neural networks (ANNs) which are inspired by the structure and function of the human brain. A typical neural network consists of multiple layers:

- Input Layer: Receives raw data as input.
- Hidden Layers: Perform feature extraction through a series of transformations with the assistance of activation functions such as ReLU (Rectified Linear Unit) and Sigmoid.
- Output Layer: Produces the final prediction based on the learned representations [30].

Deep Neural Networks (DNNs) consist of more than one hidden layer, enabling them to extract and learn abstract representations from complex and unstructured data, and their generalisation capabilities are more powerful [30].

- **Applications of Deep Learning in Medical Imaging**

Medical imaging has been greatly enhanced by deep learning improved accuracy in tasks such as disease diagnosis, image segmentation and classification. Its ability to automatically extract informative features from medical images has made it a very powerful tool for use in the clinic. Some of its notable applications include:

- Medical Image Classification: Convolutional Neural Networks (CNNs) are used extensively for medical image classification, separating healthy and pathological conditions with high precision.
- Image Segmentation: Methods such as U-Net and Mask R-CNN are used extensively to segment anatomical structures and detect abnormalities, facilitating treatment planning as well as medical research.
- Disease Prediction and Anomaly Detection: Through the detection of intricate patterns within medical data, deep learning models assist in identifying diseases at early stages, increasing diagnostic accuracy, and patient outcomes[23].

As the number of large annotated medical datasets increases and computational resources continue to improve, deep learning is breaking new frontiers in medical image analysis towards more accurate and automated diagnostic systems [23].

I.4 Computer Vision and Image Processing in Medical Applications

Computer vision and image processing have come a long way, becoming an essential part of contemporary medical applications. These technologies improve diagnostic accuracy and efficiency in healthcare by extracting vital information from medical images. Medical imaging processing is a critical field of research, and advanced techniques like multimodal information fusion improve qualitative assessment and quantitative diagnosis. Computer-aided diagnosis (CAD) systems play a key role in disease detection through high-quality feature extraction and multi-stage classification, particularly in conditions such as breast cancer. Besides, computer

vision enables predictive analytics and treatment planning by creating sophisticated algorithms to identify pathological patterns and analyze clinical data to enhance patient outcomes. With the growth of machine learning, such solutions are increasingly being used in clinical practice, leading to more intelligent and efficient diagnostic systems [31].

I.4.1 Definition and Scope of Computer Vision

Computer vision, a field-based branch of artificial intelligence, enables computers to read and process visible information in the same way that human eyes do. It includes image and multidimensional data set acquisition, processing, and inference of useful information[32].

In medical diagnosis, computer vision has transformed image detection, classification, and segmentation to a great extent, increasing accuracy and efficiency significantly. Computer vision is distinct from traditional methods grounded in human expertise, as it employs machine learning, image processing, and pattern recognition to detect intricate abnormalities in medical images. These operations are also extended to primary applications such as lesion detection, anatomical segmentation and 3D reconstruction, hence becoming an indispensable tool in modern medical imaging [32].

Benefits of Computer Vision in the Medical Field

Computer vision enhances the treatment of patients and medical diagnoses significantly by simplifying complex processes and streamlining them. Its biggest positives are:

- Early and Accurate Disease Diagnosis: Computer vision enables accurate medical image analysis with high precision, which results in early diagnosis of diseases such as cancer and glaucoma. This allows for early treatment and better results from treatments.
- Minimization of Diagnoses' Errors: Compared to human interpretation prone to subjectivity or tiredness computer vision can offer repeated and objective image evaluation, minimizing the likelihood of misdiagnosis.
- Automating Routine Medical Imaging Processes: Such routine medical imaging processes such as tumor size measurement and cell count can be automated, allowing health experts to be relieved of excessive drudgery and boosting productivity[33].

Key Differences of Image Analysis and Traditional Medical Examination

Traditional medical examination depends on clinical experience, human expertise, and manual interpretation of medical images. Computer vision enhances all these operations by including automation and precision:

- Automation of Repetitive Operations: Processes such as tumor size calculation and detection of abnormalities can be automated, reducing the human workload and enabling physicians to focus on critical decision-making.
- Quantitative and Objective Analysis: Unlike subjective judgments by humans, computer vision provides precise measurements and statistical data to support data-based clinical decisions.
- Prevention of Human Bias and Error: Computer vision circumvents inconsistencies of image interpretation to provide better diagnosis and prevent chances of misdiagnosis [34].

I.4.2 Image Analysis and Its Various Levels

Visual data analysis is an essential technique for cracking the visual language presented through multimedia, whether in printed, electronic, or audiovisual form. Minute examination of the visual data in color, shape, composition, and symbols for their underlying significance and emotional meaning is involved here. Unlike reflective superficial beauty, analysis also encompasses understanding of the subject and message of the picture. With the inquiry of whence, why, and what and what it seeks to inform us of, it is possible to foster a critical position regarding how images depict all our lives. [35].

Other than its role in communication and media, image analysis is also a significant foundation in scientific and medical sciences. Of particular interest, uses of digital image processing technologies have revolutionized the diagnostic process in medicine and offered more precise and automated interpretations of medical images. For this reason, computer-aided disease diagnosis (CAD) has been a significant research area since it is in the middle of medical imaging and diagnostic procedures[36]. In the domain of image analysis, the processing of images is generally categorized into three levels:

I.4.3 Low-Level Processing

This phase focuses on the basic processes enhancing the quality of an image and preprocessing of the visual information, including :

- Image Preprocessing: Increases the clarity of the image by discarding noise and distortion. item Noise Reduction: Applies filters like Gaussian and Median in order to remove unnecessary distortions.
- Contrast Enhancement: Adjusts brightness and contrast to highlight significant detail.
- Sharpening of Images: Sharpen edges and textures to improve visual definition.
- Features: Input and output are both images [37].

I.4.4 Mid-Level Analysis

This refers to the removal of meaningful structures from an image to be processed computationally, including:

- Segmentation: Divide an image into meaningful objects or areas for exclusive analysis.
- Feature extraction: Finds and measures features such as edges, contours, and object boundaries.
- Object Classification: Classifies and identifies objects from extracted features.
- Characteristics: Inputs are images, and outputs are features that have been extracted [38].

I.4.5 High-Level Image Analysis

High-level image analysis relies on advanced artificial intelligence and deep learning algorithms to interpret image content, extract complex information and make important decisions

- Scene Analysis: Examines spatial interactions and relationships between objects within an image.
- Pattern Recognition: Recognizes objects based on their distinctive features and learned models.
- Artificial Intelligence Methods: Utilizes machine learning and deep learning for decision-making automatically.
- Features: Deals with understanding and making conclusions on known objects [39].

I.5 Medical AI Image Processing Pipeline

Image processing is a fundamental aspect of medical AI rooted in different methods to enhance image quality feature extraction and region of interest detection.

I.5.1 Preprocessing Methods

Noise Reduction

Medical images are generally subject to noise resulting from acquisition constraints. To counter this, noise reduction algorithms such as Gaussian and median filtering eliminate noise without losing significant details. The Wiener filter also adapts to local variation, further making the image clear as well as enhancing analysis by AI systems [40].

Normalization

Image normalization is a critical preprocessing procedure in medical imaging, aligning images to a common statistical distribution based on spatial properties and pixel intensity. Spatial normalization rescales images to a standard size and alignment through scaling, rotation, or deformation so that comparisons become simpler. Intensity normalization corrects pixel values across images to reduce variability under imaging conditions. Techniques like scan bias correction help to correct intensity inhomogeneity, especially in MRI scans. The data normalization methods increase datasets' consistency, which leads to improved performance of AI models in the analysis of medical images [41].

Data Augmentation

Deep learning models in medical imaging require large datasets to provide accurate and unbiased predictions. Medical image shortage due to limited patient data and ethics, however, primarily leads to biased models and overfitting. Data augmentation remedies this shortcoming by artificially enriching dataset variety through operations such as rotation, flipping, scaling, and elastic deformations approximating anatomical variations. Contrast enhancement techniques like histogram equalization also improve the clarity of images to more sharply define subtle detail. Selecting appropriate augmentation tactics based on image type and modality is critical when optimizing AI-driven medical diagnosis [42].

I.5.2 Segmentation Techniques

Medical image segmentation is essential in distinguishing regions of interest (ROI), disease diagnosis, treatment planning, and quantitative analysis [43].

Thresholding

Thresholding is a segmentation technique that converts a grayscale image to a binary image by classifying pixels based on intensity.

- Global thresholding: Uses a constant threshold for the entire image Otsu method maximizes it by maximizing inter class variance.
- Adaptive thresholding: Adapts the threshold locally for different regions improving segmentation in non uniform illumination.
- Multi-thresholding: Used multiple thresholds to separate an image into more than one region and is used in the detection of more than one object [44].

Region-Based Segmentation

Region-based segmentation is a method that groups neighboring pixels based on pre-defined spatial relationships and similarity measures to produce homogenous meaningful regions. These methods apply similarity measures, such that the resulting regions become homogenous with topological preservation and without over-segmentation.

- Top-down method: Starts with a whole image or a large initial region and continues to subdivide it into smaller and smaller subregions based on homogeneity constraints. This method typically utilizes predetermined seed pixels, randomly selected or systematically selected, and keeps on recursing until segmentation needs are met.
- Bottom-up approach: Begins with seed points within objects of interest and gradually extends these regions progressively by merging neighboring pixels satisfying the similarity condition, supporting more refined segmentation [9].

U-Net Architecture

U-Net is a biomedical image segmentation-specific convolutional neural network architecture that was first presented in 2015. U-Net applies pixel-wise classification to allow for precise definition of anatomical structures in medical images. Figure I.2 shows the U-Net architecture,

which consists of three main parts: the contracting path (encoder), the bottleneck, and the expanding path (decoder). The encoder captures high-level contextual features by applying repeated convolution and max-pooling operations, reducing spatial dimensions while increasing depth. The bottleneck connects both paths and represents the deepest features. The decoder then upsamples the feature maps and combines them with high-resolution features from the encoder via skip connections, enabling precise localization. The final output is a segmentation map generated through a 1×1 convolution followed by a Softmax activation function for pixel-wise classification.

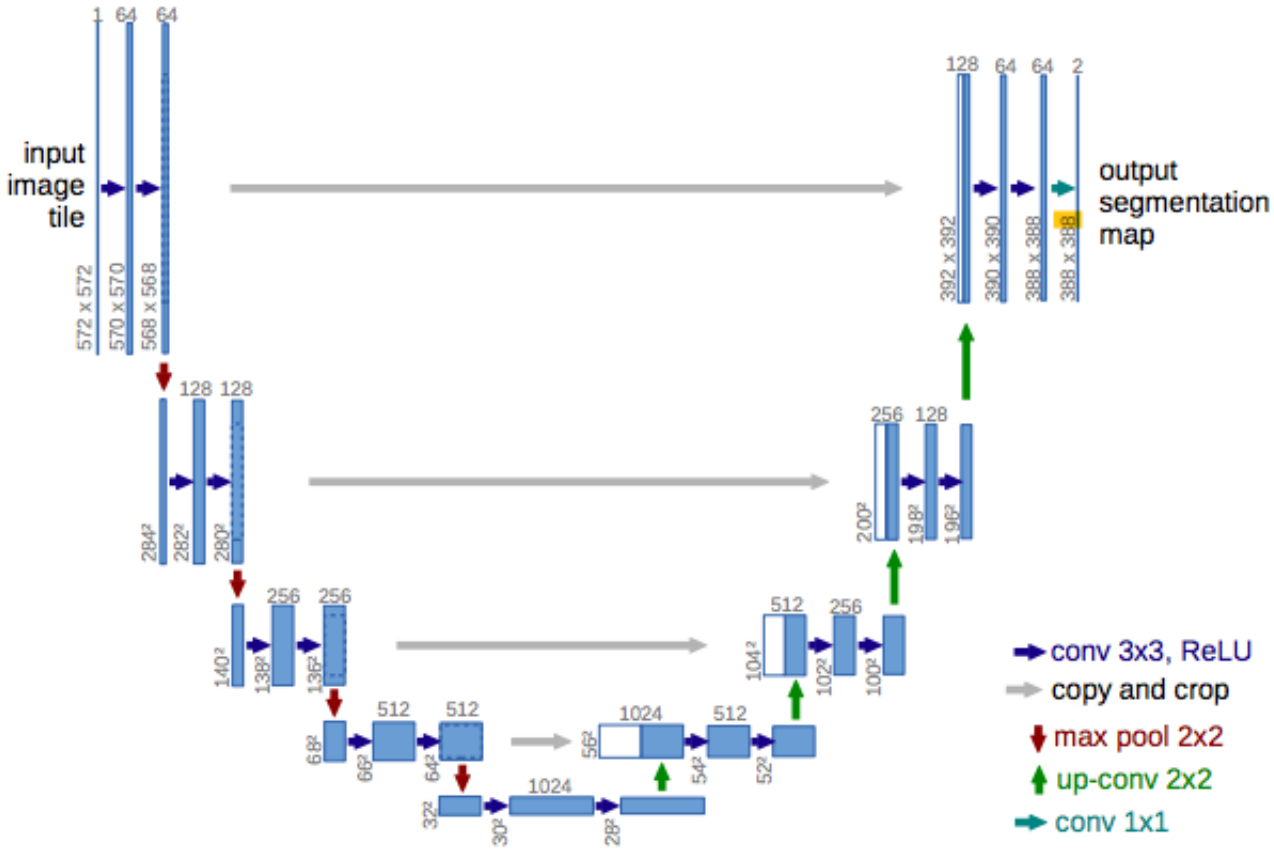

Figure I.2: U-Net Architecture [2].

Structurally, U-Net is a symmetric "U"-shaped network with three fundamental parts: the contracting path (encoder), bottleneck, and expanding path (decoder). The encoder captures contextual information by successively downsampling the spatial resolution with convolution and max-pooling. The decoder restores the segmentation map by upsampling and blending the high-resolution features from the encoder while preserving spatial details.

The network employs a Softmax activation function followed by a pixel-wise loss function—usually cross-entropy or Dice loss—to classify each pixel into one of the target classes [2]. The greatest strength of U-Net is the ability to train it on relatively small annotated training

sets, making U-Net a better fit for the medical field with scarce annotated data. It has a very modular architecture, easily adjustable and extendable for various tasks. For instance, U-Net has been used in successful applications of brain tumor segmentation,vessel detection in retina, and organ delineation [2].

I.5.3 Key Techniques for Feature Extraction and Selection

Feature extraction is a fundamental process in data analysis that enhances model accuracy and efficiency by isolating the most relevant features while eliminating redundancy. Several techniques are widely employed for this purpose:

- Principal Component Analysis (PCA): A widely used dimensionality reduction technique that transforms correlated variables into a smaller set of uncorrelated principal components. PCA helps retain essential data structures while minimizing complexity, ultimately improving computational efficiency and model performance.
- Gabor Filters:These filters are extensively utilized in image processing, particularly for texture analysis. They effectively capture spatial frequency details, making them valuable for feature selection and object recognition in medical imaging and other computer vision applications. By detecting edges and texture patterns, Gabor filters enhance classification and segmentation accuracy. [45]

Both PCA and Gabor filters significantly contribute to refining feature extraction processes, leading to improved interpretability and performance in AI-driven models across various fields [45].

I.6 Convolutional Neural Networks (CNNs) for Medical Imaging

Convolutional neural networks (CNNs), a highly sought-after tool of deep learning, are specifically designed architectures for image processing and analysis. In medical imaging, CNN models have played a tremendous role in the diagnosis and treatment of disease by efficiently extracting features from medical images. Transformative learning, which makes use of pre-trained CNN models, helps to bridge the challenges of limited datasets and computational resources[46].

CNNs hierarchical extraction of features of convolution and clustering layers of images. CNNs

are also position, size and orientation change insensitive in images. This is because they use convolutional layers where segments of an image are mapped with local filters and the use of clustering layers where the dimension is reduced at the expense of preserving relevant features. CNNs can process large images with high computational efficiency because of the weight-sharing mechanism of convolutional filters. This significantly reduces the model parameters relative to fully connected architectures [46].

CNNs have demonstrated state-of-the-art performance in most computer vision tasks and competitions and outperformed conventional approaches in object detection and image classification tasks[46]. CNNs have numerous advantages for image processing, including:

- Increased accuracy: CNN models have made unprecedented strides in image analysis and classification, therefore enhancing the accuracy of medical diagnosis.
- Save time and resources: Transformative learning saves time and resources required, through the use of pre-trained models to speed up processes and gain improved efficiency.
- Highly efficient image processing: CNNs are able to process big images at high speed effectively, decreasing the need for enormous resources to process big medical images.
- Adaptability to diverse tasks with outstanding performance in various applications: CNNs offer superior adaptability to various tasks in multiple domains.
- Present challenges: Despite the benefits present challenges such as demanding large and diverse datasets and interpreting deep learning models are a challenge [46].

I.6.1 Fundamentals of CNNs

Convolutional Neural Networks (CNNs) image processing automatically determines spatial hierarchies of features, which allows them to learn patterns like edges, textures, and shapes—important features in medical image analysis. The building blocks of CNNs are:

- **Convolutional Layer**

The convolutional layer identifies distinctive features from input images through filters that recognize edges, textures, and patterns. The filters are learned and enable the network to fine-tune for particular tasks such as classification and object detection [3]. Figure I.3 represents how a convolutional layer works

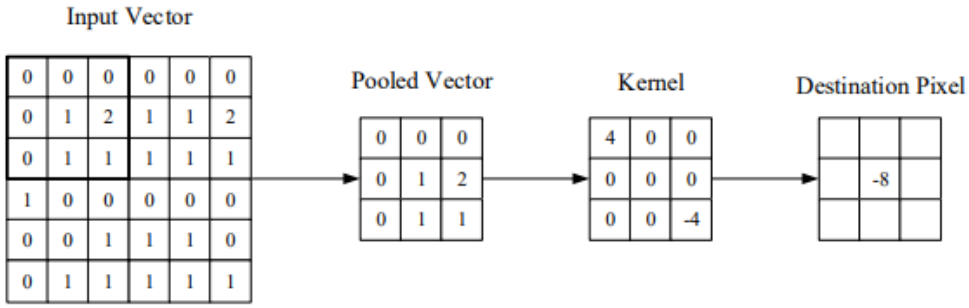

Figure I.3: Principle of Convolutional layer [3].

- **Pooling Layer**

The pooling layer lowers the spatial size of input images without losing important features, which reduces computational cost. Typical techniques are Max Pooling (taking the maximum pixel value) and Average Pooling (taking the average pixel value) [3].

- **Fully Connected Layer**

The fully connected layer receives processed features and outputs final predictions and is a critical component of classification and regression [3].

- **Activation functions**

Activation functions add nonlinearity to models, helping deep learning systems understand and learn complex patterns that simple linear models can't capture [46]. The most common functions are:

ReLU activation function is a simple yet widely used nonlinear activation function in convolutional neural networks. It improves efficiency by activating only a few neurons at a time. As shown in Equation I.1, ReLU returns the input if it is positive, and zero otherwise [46]:

$$f(x) = \max(0, x) \tag{I.1}$$

Sigmoid function is a widely used nonlinear function, especially in binary classification tasks. It transforms input values into a range between 0 and 1, making it suitable for probabilistic interpretations [46]. The sigmoid function is defined as follows:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{I.2}$$

Softmax function used for multiclass classification, converting outputs into probabilities that sum to 1 [46]. It is defined as:

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}, \quad j = 1, \ldots, K \tag{I.3}$$

I.6.2 Overview of Convolutional Neural Network Architectures

Here we provide an overview of widely used CNN architectures, including SegNet, VGGNet, AlexNet, and ResNet.

SegNet

Figure I.4 illustrates the architecture of SegNet, a deep learning model specifically designed for semantic segmentation tasks

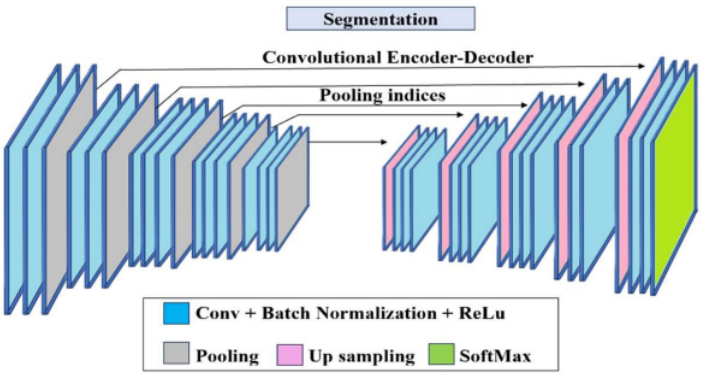

Figure I.4: SegNet Architecture [4].

VGGNet

Figure I.5 shows the architecture of VGGNet deep learning model.

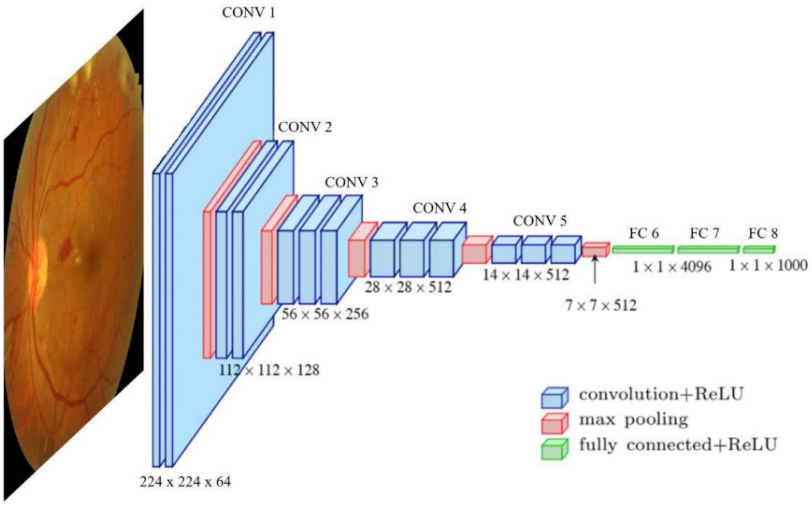

Figure I.5: VGGNet Architecture[5].

AlexNet

Figure I.6 presents the architecture of AlexNet, one of the earliest deep convolutional neural networks.

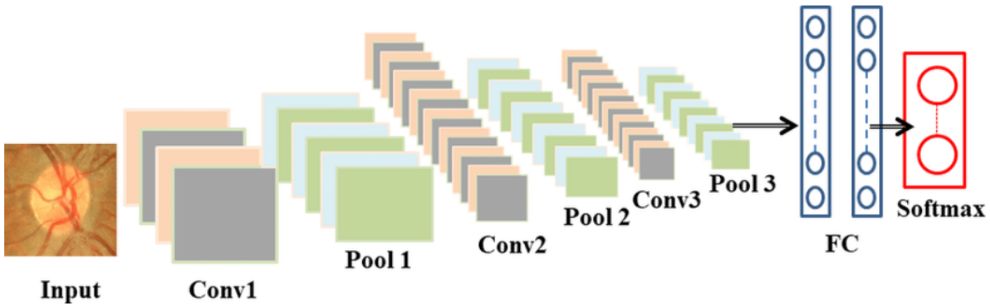

Figure I.6: AlexNet Architecture[6].

ResNet

A deep convolutional neural network architecture designed to address the challenges of training very deep networks. It introduces a residual learning framework, where layers learn to predict residuals [7].

ResNet achieved state-of-the-art performance on several benchmarks, including winning first place in the ILSVRC 2015 classification task [7].

As shown on the bottom of Figure I.7, the ResNet architecture introduces shortcut connections that allow the network to learn residual functions, facilitating the training of deeper models with improved performance

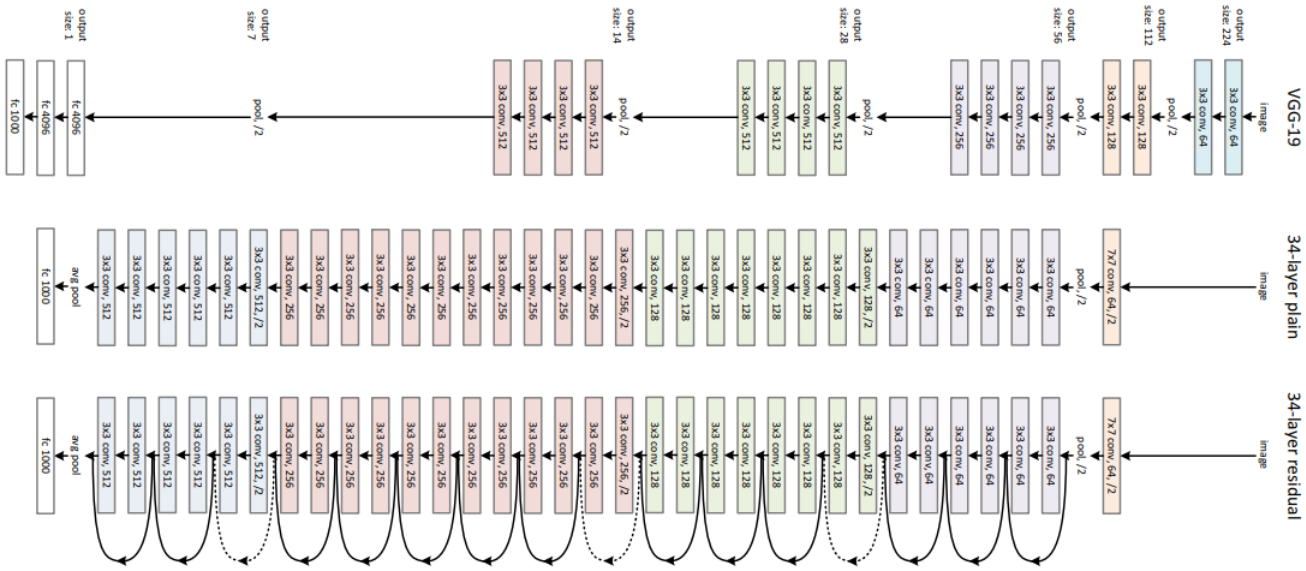

Figure I.7: The bottom network represents the ResNet model with 34 layers[7].

I.7 Transformers

The Transformer is a deep learning model solely founded on the attention mechanism, without recurrent or convolutional networks. It was presented by Vaswani et al. (2017) in their work "Attention Is All You Need", which outlines its architecture and its remarkable results on machine translation and other sequence modeling tasks [8].

In contrast to sequential models like RNNs and LSTMs, the Transformer handles input sequences in parallel, which not only greatly boosts efficiency but also the capacity for modeling long dependencies. Its defining mechanism, Self-Attention, enables the model to assign importance to each point in the sequence in relation to the others [8].

The Transformer model adheres to an Encoder-Decoder framework:

- Encoder: Uses the input sequence through a series of self-attention layers and nonlinear transformations to build up rich feature representations.
- Decoder: Employs these learned representations to produce an output sequence taking into account both the context of the input and the elements previously generated [8].

A key component of the Transformer is the Multi-Head Attention mechanism, enabling the model to pay attention to several types of relationships between elements in a sequence at the same time, enhancing its capability to model complex dependencies. Ever since its origin, the Transformer model has transformed several domains, more specifically Natural Language Processing (NLP) and Computer Vision. BERT, GPT, and T5 are some of the models derived

from this architecture that have set state-of-the-art results for several AI tasks[8].
The whole Transformer model architecture is illustrated in Figure I.8.

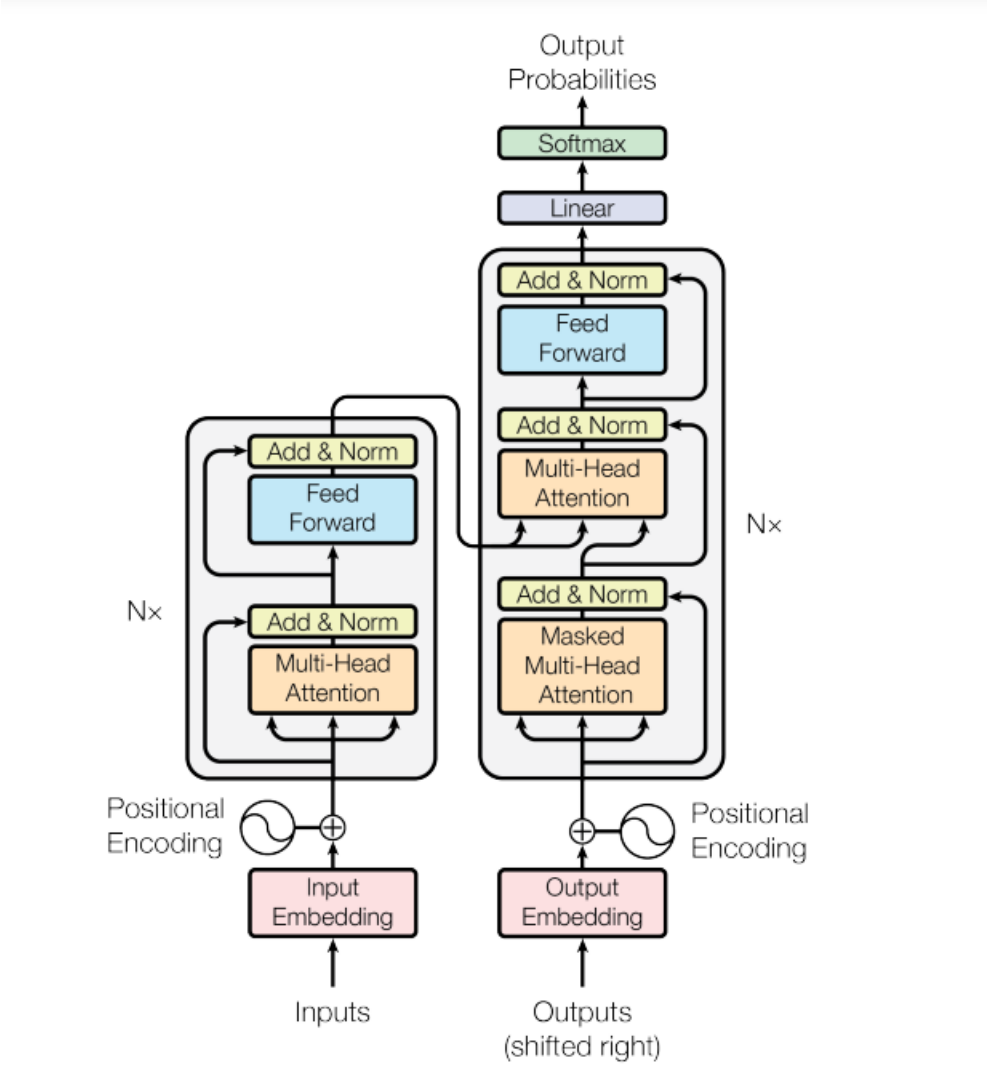

Figure I.8: Transformer Architecture [8].

I.7.1 Vision Transformer (ViT)

The Vision Transformer (ViT) introduces a novel architecture for image analysis by eliminating the use of conventional convolutional layers and instead leveraging a pure attention mechanism. Unlike CNNs that extract local features using convolutional filters, ViT splits an input image into fixed-size patches, embeds each patch as a token, and then applies standard Transformer encoders to model the global relationships between these tokens. ViT has shown competitive performance in image classification tasks, especially when trained on large-scale datasets. However, it may be less effective in scenarios requiring fine-grained object localization or when trained on small datasets [47]. Figure I.9 presents a schematic overview of the Vision

Transformer architecture.

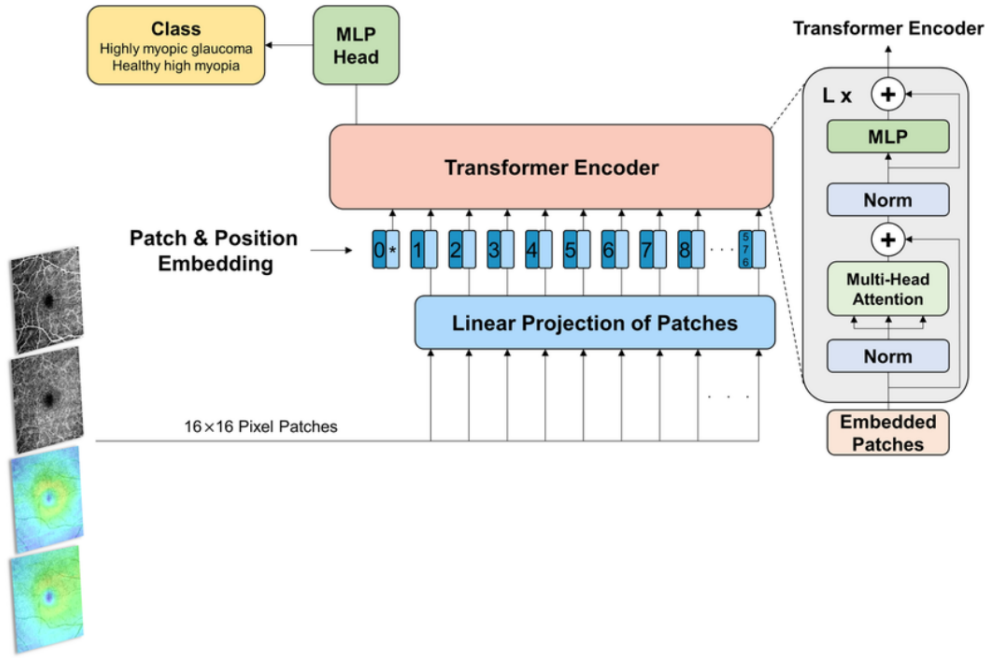

Figure I.9: Schematic of Vision Transformer (ViT) Architecture[9].

I.7.2 Swin Transformer

The Swin Transformer is a vision-specific Transformer model that is designed to be a general-purpose backbone for computer vision tasks. It introduces a hierarchical structure and a shifted window approach for efficient computation of local attention. This not only truncates the computational complexity by limiting the self-attention operation within lapping local windows but also provides the possibility of cross-window connection among different regions of the image . With this architecture, the Swin Transformer achieves multi-scale transformations and reaches state-of-the-art performance on image classification, object detection, and semantic segmentation and outperforms current models on all available benchmarks [48].

I.7.3 SegFormer

SegFormer is a new semantic image segmentation model that synergistically combines Transformers and efficient, lightweighted MLP layers. One of the key features of SegFormer is its novel hierarchical Transformer encoder, which generates multi-scale features from images independently of positional encodings. The architecture avoids position code interpolation, a common issue that has a negative effect on performance if testing resolution is not equal to training resolution. Additionally, SegFormer avoids the requirement of sophisticated decoders.

Instead, it suggests a lightweight MLP decoder that effectively fuses information from different stages of the encoder through the utilization of both local and global attention mechanisms in order to produce accurate and robust representations [10].

Figure I.10 illustrates the SegFormer architecture, which is composed of two main components: a hierarchical Transformer encoder and a lightweight MLP decoder. The encoder extracts multi-scale features from the input image without relying on positional encodings, enabling it to generalize better across different image resolutions. These features are then passed to the decoder, which consists of simple MLP layers that fuse information from various encoder stages. This design allows SegFormer to capture both local and global contexts efficiently, resulting in accurate and scalable semantic segmentation performance with minimal computational complexity.

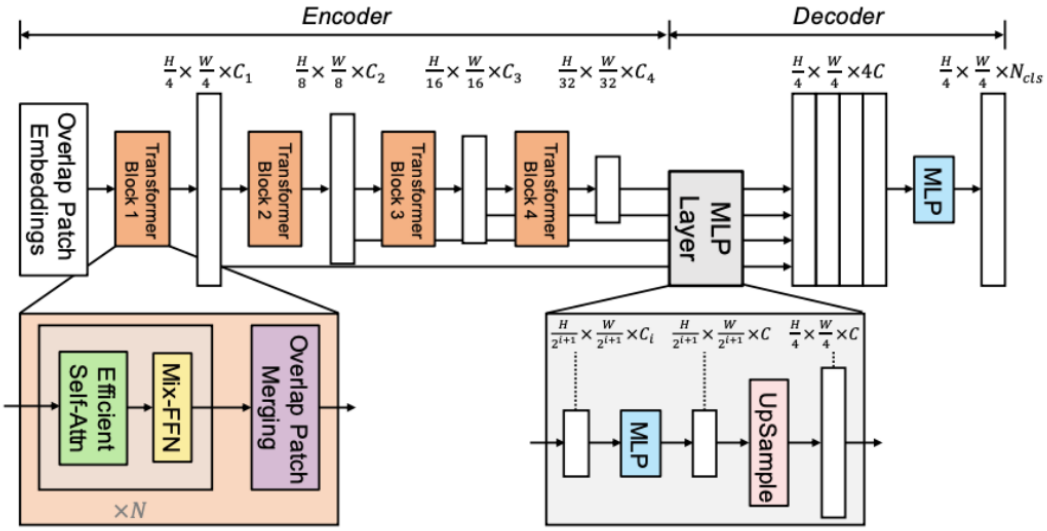

Figure I.10: SegFormer architecture: a hierarchical Transformer encoder with a lightweight MLP decoder[10].

I.7.4 SAINT

SAINT (Self-Attention and Intersample Attention Transformer) is a modern artificial intelligence model specifically designed to process tabular data, which is one of the most commonly used types of data in vital sectors such as medicine, finance, and logistics. SAINT is distinguished by its integration of two mechanisms: Self-Attention for analyzing relationships between features within each sample, and Intersample Attention for extracting relationships between different rows in the table, which gives it the ability to understand the complex structure of tabular data. The model handles both continuous, categorical, and ordinal features

by projecting them into a shared vector space, and then passing them through a Transformer encoder. SAINT is trained in two phases: first, the self-supervised pretraining phase to enhance the initial representations, and then the supervised finetuning phase to perform predictive tasks. Experiments have shown that SAINT outperforms many traditional models, such as XGBoost, CatBoost, and LightGBM, in both supervised and semi-supervised tasks[11].

Figure I.11 presents the architucture of SAINT model

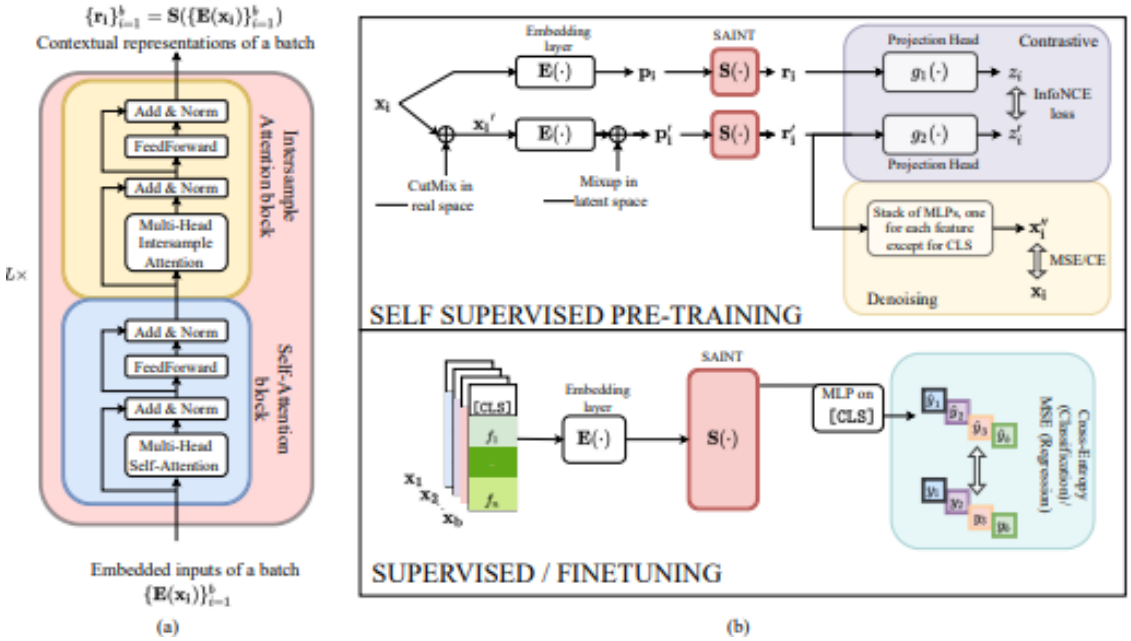

Figure I.11: SAINT model architucture [11] .

I.7.5 Hybrid CNN-Transformer Architectures for Medical Image Segmentation

Transformer-based models have taken center stage in image segmentation in recent times. Compared to traditional CNN-based models that are primarily concerned with capturing local spatial features via convolutional operations, transformers are more capable of representing long-range interactions within a picture. This allows them to recognize the overall context and interdependencies between distant areas better, thereby improving segmentation performance. The models have been implemented across numerous application domains, including medical imaging, where correct identification of anatomical structures is a top priority [49].

Swin UNETR

Swin UNETR integrates Swin Transformers with an encoder-decoder U-Net architecture for state-of-the-art 3D medical image segmentation, namely for brain tumor diagnosis using MRI

analysis. With hierarchical self-attention mechanisms, it effectively acquires both local and global contextual representations, overcoming convolutional model limitations in representing long-range dependencies. The model topped performance in the BraTS 2021 segmentation task, which shows its effectiveness in medical imaging tasks [50].

TransUNet

TransUNet uses the strength of CNNs and Transformers to perform medical image segmentation. It uses a transformer encoder for long-range dependency capture and a CNN decoder for spatial detail improvement with both global context awareness and local accuracy. It has a hybrid framework for generating improved performance for applications such as multi-organ and cardiac segmentation and outperforms baseline models [51].

U-Netmer

U-Netmer marries transformers and U-Net everywhere locally to enhance the medical image segmentation. U-Netmer preserves local pixel relations but utilizes self-attention for distant dependencies. By addressing token-flattening and scale-sensitivity issues, U-Netmer boosts the segmentation quality of various organs and imaging modalities [52].

I.7.6 Transformer-Based Segmentation Models

Image deraining is suggested to restore images polluted by rain streaks, where CNNs struggle to deal with spatial details and non-uniform patterns. A hybrid CNN-transformer model enhances recovery through the integration of local feature extraction, global context modeling, and frequency-domain contrastive learning, achieving improved performance over state-of-the-art methods [53].

CoTr (Convolutional Transformer)

CoTr combines CNN feature extraction with a deformable transformer to achieve efficient long-range dependency modeling with low computational cost and higher 3D medical image segmentation accuracy [54].

HiFormer (Hierarchical Multi-Scale Representations)

HiFormer combines CNNs and Swin Transformers to learn local and global features using a Double-Level Fusion mechanism to achieve better segmentation accuracy in medical

imaging[55].

UNetFormer

UNetFormer marries a U-Net-like structure with a transformer-decoder, drawing on the ResNet18 for encoding effectiveness and a global-local attention mechanism to enhance segmentation accuracy. UNetFormer performs better than others in urban scene segmentation with great computational efficiency[56].

I.8 Conclusion

Artificial intelligence has become an important tool in medical image analysis, contributing significantly to disease diagnosis and image analysis for decision-making. In this chapter, we have provided a complete review of AI models, explaining the most pertinent concepts related to machine learning, deep learning, and transformer-based models, highlighting their applications in medical image processing. The discussion also emphasized the critical role of computer vision and convolutional neural networks (CNNs) in enhancing diagnostic accuracy and clinical efficiency, In the second chapter, new advances in automated glaucoma diagnosis will be discussed, with a focus on methods for early detection of glaucoma.
In the next chapter, we will review how these AI models can be used for automated glaucoma diagnosis, with a focus on the latest developments, existing challenges, and the most prominent scientific studies in this field.

CHAPTER II

ARTIFICIAL INTELLIGENCE PARADIGMS FOR AUTOMATED GLAUCOMA DIAGNOSIS: CHALLENGES AND LITERATURE REVIEW

II.1 Introduction

Glaucoma is a progressive and irreversible condition that damages the optic nerve, leading to partial or complete vision loss. Its diagnosis presents considerable challenges, as it often depends on subjective assessments by specialists. Retinal imaging provides crucial information about ocular health. However, advancements in imaging technology have made it possible to develop systems capable of analyzing these images for more accurate diagnosis.

Artificial intelligence (AI), particularly deep learning models, has emerged as a powerful tool in medical imaging, offering high-accuracy automated detection and classification of glaucoma. Techniques such as Transformers, Convolutional Neural Networks (CNNs), and other AI-driven methods have demonstrated exceptional performance in analyzing retinal scans, aiding ophthalmologists in making more objective and consistent diagnoses. However, despite these advancements, challenges such as data variability, model interpretability, and generalization remain significant obstacles.

This chapter delves into the various AI paradigms applied to glaucoma diagnosis, examining current methodologies, key challenges, and the latest advancements in the field. It provides a comprehensive review of AI-based approaches, highlighting their potential and limitations.

II.2 Clinical overview of Glaucoma

II.2.1 Definition

Glaucoma is a chronic and progressive eye disease that damages the optic nerve, the critical structure responsible for transmitting visual information from the eye to the brain. The optic nerve can be conceptualized as a bundle of over one million retinal ganglion cell axons. In glaucoma, these axons undergo progressive degeneration , ultimately leading to irreversible vision loss [57].

One of the most common risk factors associated with glaucoma is elevated intraocular pressure (IOP) When this pressure becomes too high, it starts to compress the optic nerve. as shown in Figure II.1

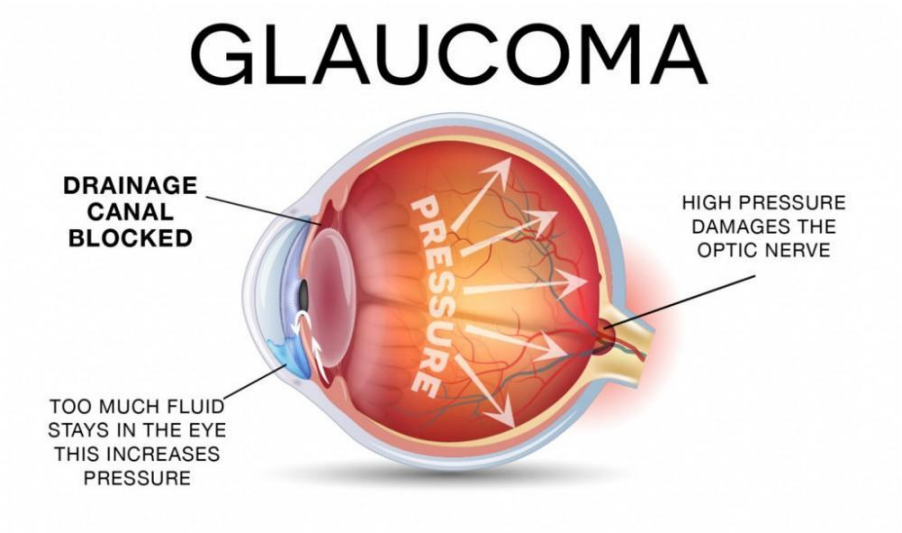

Figure II.1: the progression of optic nerve damage [12].

This pressure affects a part of the eye called the optic disc, which is the visible portion of the optic nerve at the back of the eye. Within this disc is a central depression called the optic cup. In healthy eyes, this cup is small relative to the disc, but in glaucoma, as nerve fibers are lost, the cup becomes larger. This leads to an increased cup-to-disc ratio (CDR)—a key indicator used by eye doctors to detect glaucoma [58] This structural change reflects underlying neural damage.

As more nerve fibers are damaged, the Retinal Nerve Fiber Layer (RNFL)—a layer of tissue containing these fibers—begins to thin.[59] At the same time, the ganglion cells (neurons in the retina) also die off, leading to further deterioration of vision increases with age[60].

However, In some individuals, this neurodegeneration occurs despite intraocular pressure remaining within the statistically normal range—a condition known as normal-tension glaucoma (NTG) or low-tension glaucoma.thus, While vision loss from glaucoma is irreversible, the primary objective of clinical management is to halt or slow disease progression. Early detection and prompt intervention are therefore essential to preserving remaining visual function[60]. If left untreated, this vision loss becomes irreversible and can eventually lead to irreversible vision loss.

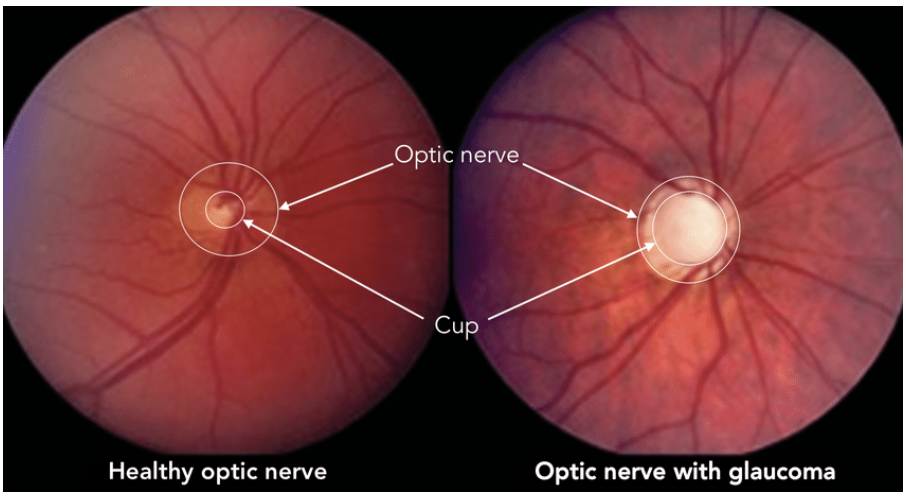

Figure II.2: The enlargement of the optic cup between healthy optic nerve and damaged one "cupping" [13].

II.2.2 Optical Coherence Tomography in Glaucoma diagnosing

Glaucoma diagnosis and monitoring rely on a combination of structural and functional assessments. Historically, the most commonly employed tools include intraocular pressure (IOP) measurement using tonometry and fundus photography, both of which remain widely used in clinical practice [57]. Fundus imaging allows for the evaluation of morphological features such as the cup-to-disc ratio (CDR) and the neuroretinal rim (NRR)—parameters that are essential for assessing optic nerve head changes associated with glaucoma.[58] However, these tools provide limited insight into the internal retinal architecture and may not detect early-stage structural damage

The introduction of Optical Coherence Tomography (OCT) has significantly enhanced the structural evaluation of glaucomatous damage. OCT is a non-invasive imaging modality that generates high-resolution, cross-sectional scans of the retina and optic nerve head using low-coherence interferometry. In glaucoma management, it enables quantitative analysis of several structural biomarkers including retinal nerve fiber layer (RNFL) thickness, ganglion cell complex (GCC) integrity, neuroretinal rim area, and cup-to-disc ratio (CDR)[58].

Thus, Ophthalmologists utilize OCT as a fundamental tool in the clinical evaluation of ocular diseases including Glaucoma as it enables quantitative assessment of key structural biomarkers associated with it. OCT has demonstrated high reproducibility across multiple studies , making it a reliable method for longitudinal disease monitoring. Furthermore, OCT is less demanding for patients, offers rapid acquisition times, and is considered no more complex to interpret than visual field assessments or fundus photography [61]. Advanced OCT systems

are now equipped with automated segmentation algorithms that calculate parameters such as CDR and RNFL thickness with high precision [62]. These measurements form the basis of many clinical decisions and support the development of artificial intelligence (AI)-based diagnostic tools.

Despite the widespread adoption of OCT for glaucoma diagnosis, clinical observations suggest that its performance may vary depending on patient demographics. At the Cooperative Ophthalmology Hospital Foundation of Cuba -El Oued in Algeria where this study is conducted, practitioners use the Mocean 4000, an OCT device manufactured in China illustrated in Figure II.3. This system provides a range of analyses including macular imaging, anterior-segment OCT, angiography, and Heidelberg Retinal Tomography (HRT)—the latter being critical for quantifying glaucoma biomarkers illustrating its capability in providing detailed 3D topographical maps essential for early glaucoma detection and monitoring. However, clinicians have reported inconsistencies in diagnostic outcomes, particularly in patients of non-Eastern Asian descent, highlighting a gap between automated measurements and expert evaluation.

Figure II.3: The OCT device at the CUBA-ELOUED cooperative ophthalmology hospital.

II.3 Artificial intelligence in glaucoma diagnosis

Currently, glaucoma screening primarily relies on fundus photography and intraocular pressure (IOP) measurement using Tonometry, which provide useful but limited indicators of disease

presence. These conventional methods often lack the ability to detect early-stage glaucoma, leading to delayed diagnosis and irreversible vision loss as a result. However, fundus imaging alone does not provide a comprehensive assessment of the optic nerve and retinal layers, which are crucial for accurate glaucoma diagnosis. Thus, Optical Coherence Tomography (OCT) has become a preferred imaging modality, offering high-resolution, cross-sectional scans of the retinal structure[63]. Given the increasing availability of OCT devices in clinical settings, integrating AI-driven analysis into OCT-based diagnostics has become a focal point in ophthalmic research.

II.4 Challenges in AI based glaucoma diagnosis

While AI-enhanced OCT analysis holds great potential for early detection and objective assessment of glaucoma, its performance and clinical applicability remain highly dependent on the composition of the datasets used to develop these algorithms. A particularly pressing issue is the impact of ethnic variability on OCT-derived biomarkers.

At the ophthalmology hospital, clinicians raised serious concerns about the diagnostic reliability of the Mocean 4000 OCT device that developed and validated in China. Although this device offers advanced capabilities, its automated results often conflicted with expert evaluations, especially in local patients. In many cases, the device classified patients as glaucomatous or non-glaucomatous in ways that contradicted comprehensive clinical examinations. As a result, ophthalmologists frequently reverted to traditional diagnostic protocols, such as 24-hour intraocular pressure monitoring, fundus imaging, and visual field testing, to validate or correct the device's conclusions.

These inconsistencies are not isolated to one institution—they reflect a broader challenge in AI-based ophthalmology. Ethnic differences in key structural features such as retinal nerve fiber layer (RNFL) thickness, optic disc shape, and cup-to-disc ratio (CDR) have been well-documented [64]. Despite this, most commercial OCT systems and the AI models used to interpret their outputs are built using normative databases dominated by Caucasian populations [65]. When such models are applied to individuals from African, Asian, or Hispanic backgrounds, the mismatch in anatomical baselines often leads to misclassification, inaccurate risk scoring, or underestimation of disease progression [65].

To ensure diagnostic equity, there is an urgent need for the development of population-specific normative databases that accurately reflect the anatomical diversity of the global population. Moreover, future AI algorithms should be trained on large, ethnically diverse datasets

that account for genetic, anatomical, and demographic variability. Doing so would not only enhance diagnostic accuracy but also build clinical trust in AI-driven glaucoma tools across all regions and populations [58].

II.5 Problem Statement in Early Detection of Glaucoma Using Artificial Intelligence

Glaucoma is a chronic, neurodegenerative disorder and a leading global cause of irreversible yet preventable blindness, affecting approximately 2% of the population. Its etiology is multifactorial, often involving elevated intraocular pressure (IOP) or impaired ocular blood flow. The condition frequently progresses undetected due to the absence of early clinical symptoms, resulting in delayed diagnosis and treatment [66], which ultimately leads to optic nerve damage and constriction of the visual field. The optic nerve and the retinal nerve fiber layer design critical components in transmitting visual signals from the eye to the brain, undergo progressive degeneration [67].

1. Glaucoma typically manifests without overt warning signs, and its progression is insidiously gradual, with many individuals unaware of visual deterioration until the disease reaches an advanced stage. This asymptomatic nature has earned it the moniker "the silent thief of sight" [66]. Given the irreversible nature of glaucomatous vision loss, early detection is essential. Timely diagnosis enables the implementation of therapeutic strategies that can significantly decelerate or even prevent further visual impairment.

2. Fundus photography and Optical Coherence Tomography (OCT) represent the most cost effective and widely utilized imaging modalities for glaucoma screening. Each provides critical structural biomarkers indicative of glaucomatous damage, specifically, the vertical cup-to-disc ratio (vCDR) in fundus images and the thickness of the retinal nerve fiber layer (RNFL) in OCT volumes. In clinical settings, it is generally advised to employ both modalities concurrently to enhance diagnostic accuracy and reliability. Despite the development of numerous automated glaucoma detection algorithms based on either fundus images or OCT data, integrative approaches that exploit the complementary strengths of both modalities remain relatively scarce. Notably, 2D fundus photography and 3D OCT continue to be the principal tools for structural assessment of the optic nerve in routine ophthalmologic evaluations [66].

3. Recent studies indicate that relying exclusively on either fundus imaging or OCT volumetric data may result in the oversight of approximately 46.3% of glaucoma cases. Nevertheless, the majority of computer-aided diagnostic (CAD) systems for glaucoma have been developed using a single imaging modality. Although both fundus photography and OCT are integral to routine clinical screening, few computational models have been designed to leverage their complementary diagnostic information. This limited integration can be attributed to two primary challenges: (a) the absence of publicly available multimodal datasets suitable for training and evaluating such models, and (b) the substantial disparity in data characteristics and dimensional representations between the two modalities, which introduces significant technical complexity [66].

4. It is important to acknowledge a limitation of the GAMMA dataset, namely the lack of ethnic diversity, as all imaging data are derived from individuals of Chinese descent. While variations in OCT measurements across populations may be minimal, fundus photographs can exhibit substantial inter-ethnic differences, primarily due to variations in fundus pigmentation, which may affect image characteristics and diagnostic performance [66].

5. Moreover, it is noteworthy that the methodologies employed in the GAMMA challenge predominantly rely on black-box neural network architectures, with limited emphasis on model interpretability. Although deep learning has demonstrated high efficacy in automated glaucoma detection, the lack of transparency poses a significant barrier to clinical adoption. Explainability is a critical requirement for the integration of computer-aided diagnostic systems into routine ophthalmologic practice, yet it remains an underexplored aspect in this domain [66].

6. In clinical settings, glaucoma diagnosis commonly involves the manual estimation of the cup-to-disc ratio (CDR) from fundus images, wherein accurate localization of the optic disc (OD) constitutes a critical preprocessing step. To enhance diagnostic efficiency and consistency, particularly in large-scale screening scenarios, and to mitigate the variability and potential inaccuracies inherent to subjective assessment, the development of automated algorithms for OD extraction from retinal fundus images is highly desirable [68]. Nevertheless, conventional hand-crafted feature-based approaches are often highly sensitive to confounding factors such as variations in illumination and the presence of coexisting ocular pathologies, which can substantially alter the visual characteristics of

the OD region. Consequently, considerable research has been directed toward the robust segmentation of the optic disc to address these challenges.

7. A critical but often overlooked challenge in clinical glaucoma diagnosis is the visual and mental fatigue experienced by ophthalmologists during manual examination of complete OCT reports. These reports often span multiple sections and require careful interpretation to accurately assess structural parameters of glaucoma, such as RNFL thickness maps, cup-to-disc ratios, and deviation analyses. Prolonged exposure to many of these reports, especially in high-volume examination environments, can lead to reduced diagnostic accuracy due to physician fatigue.

II.6 Literature review of AI-based glaucoma detection

To address the challenges associated with the early and accurate diagnosis of glaucoma, researchers have started developing different model architectures based on both traditional machine learning techniques and deep learning methods. These models aim to improve diagnostic accuracy and reduce the reliance on manual human examination. The following table shows a variety of methods that have been used in research for the automated detection of glaucoma using artificial intelligence.

Table II.1: State-of-the-art approaches for Glaucoma Detection Using Artificial Intelligence

Type	Method	Dataset	Results	Ref
Segmentation + Classification	Improved U-Net + CNN	Glaucoma Dataset(Kaggle)	96.90%	[69]
Segmentation + Classification	U-net + EfficientNet b3	ACRIMA	99%	[57]
multi class Classification	Transferable Ranking Convolutional Neural Network (TRk-CNN)	Glaucoma image dataset(Korea university medical center)	92.96%	[67]
Classification	SVM for hybrid features	DRISHTI-GS, RIM-ONE	97.22% (split method) 93.2% (5-fold cross-validation)	[70]
Segmentation + Classification	U-Net + SVM.	DRISHTI-GS	93.33%	[71]
Classification	CNN models: Base CNN, VGG16, ResNet50, DenseNet121	Retinal Image Dataset	79%	[72]
Classification	A CNN-based model.	Ulsan University Hospital NTG Dataset	Sensitivity: 92%, Specificity: 86.9%	[73]
Classification	A custom 18-layer CNN trained to classify digital fundus images as normal or glaucomatous.	Private Fundus Image Dataset (1426 images)	91.67%	[74]

Table II.1: State-of-the-art approaches for Glaucoma Detection Using Artificial Intelligence

In order to improve accuracy and overcome the limitations of previous works, we decided in this study to propose a multi-stage approach that includes the analysis of blood vessels in addition to the optic disc and optic nerve head. We have also adopted Transformer models instead of traditional Convolutional Neural Networks (CNNs), both in the segmentation and classification stages. Moreover, we rely on real and ethnically diverse data from the local environment, which enhances the model's generalization ability and increases its clinical credibility.

II.7 Conclusion

Artificial intelligence has significantly transformed glaucoma diagnosis, particularly through deep learning-based OCT analysis. Advances in CNNs, ViTs, hybrid models, and segmentation algorithms have enhanced diagnostic accuracy, enabling early disease detection and longitudinal monitoring. However, challenges such as dataset bias, explainability, and model generalization must be addressed for AI-powered glaucoma diagnosis to reach its full clinical potential. Future research should focus on developing robust, interpretable, and ethically sound AI models to bridge the gap between AI research and real-world ophthalmic practice.

In the next chapter, we present our proposed deep learning-based methodology for early detection of glaucoma using OCT imaging. This includes a detailed explanation of the dataset used, the preprocessing steps and the model architecture.

CHAPTER III

THE PROPOSED APPROACH FOR EARLY GLAUCOMA DETECTION INTEGRATING BOTH OCT VOLUMES AND FUNDUS IMAGES

III.1 Introduction

In this chapter, we present the proposed methodology to tackle the problem addressed in the previous chapter, which is the early detection of glaucoma using Optical Coherence Tomography (OCT) images. This methodology was designed based on well-studied steps that include data collection, processing, model construction, and performance evaluation. Since OCT images provide accurate structural information about the eye, integrating them with deep learning techniques opens new horizons for automated and accurate diagnosis extracted from images.

III.2 Proposed methodology for detecting glaucoma

During the development of the baseline model, several techniques emerged as particularly effective in enhancing task performance. In routine clinical practice, glaucoma assessment often involves the manual estimation of the cup-to-disc ratio (CDR) from fundus images, a process that relies heavily on the accurate identification of the optic disc. This step is critical, as glaucomatous damage typically manifests in the optic disc region through structural changes such as increased CDR and optic disc hemorrhages. To improve diagnostic efficiency and consistency, particularly in large-scale screening contexts, and to reduce errors stemming from subjective visual assessment, the implementation of automated and standardized preprocessing techniques is strongly warranted.

To guide the model's attention toward clinically relevant structures, we extracted the optic disc region from the fundus images, thereby allowing the network to concentrate on the optic disc and cup areas. This region of interest was localized using a pre-trained optic disc segmentation network. Additionally, given the established diagnostic value of the ratio between blood vessel area in the inferior-superior axis and that in the nasal-temporal axis [66], we employed the SegFormer architecture to perform precise segmentation of blood vessels. Subsequently, the encoded features from both the fundus and OCT imaging branches were fused via feature concatenation and passed to the SAINT classifier for final glaucoma prediction.

During the training phase, the model was optimized using the cross-entropy loss function in a supervised learning framework. Training was conducted on the designated training set, with performance evaluated on a separate validation set and subsequently reported on the final test set. Fundus images were resized to a resolution of 512 x 512 pixels, while OCT images were processed at a resolution of 512 x 512 pixels. Model optimization was performed using the Adam optimizer. Ultimately, the outputs of three independently trained networks were aggre-

gated to generate the final diagnostic prediction.

To enhance the interpretability of our model, we adopted a multi-task learning framework that simultaneously performs glaucoma grading, blood vessel segmentation, and optic disc–cup segmentation. This approach enables the model to generate clinically relevant segmentation outputs, which serve as visual evidence of the network's focus on diagnostically meaningful regions. In the context of deep learning, explainability is often associated with the model's ability to highlight critical regions of interest within an image. For glaucoma diagnosis, these regions include not only the optic disc and cup but also the surrounding blood vessels. Importantly, it is the quantitative relationships, such as the cup-to-disc ratio, the diameter of retinal blood vessels and both OCT and fundus texture characterization, that provide key discriminative features for diagnosing glaucoma. Figure III.1 depicts the architecture of the proposed system.

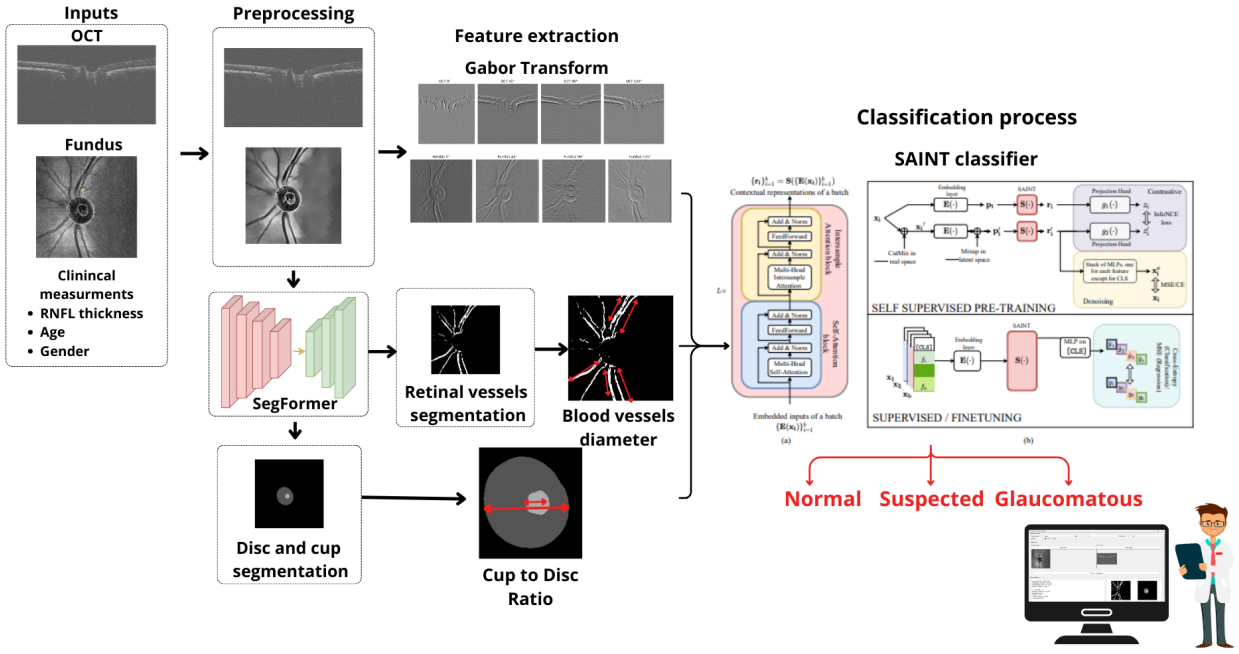

Figure III.1: The overall architecture of the proposed multi-modal glaucoma diagnosis system.

III.3 Data Collection and Preprocessing

The dataset used in this work consists of 320 OCT report images (commonly referred to locally as 'HRT'), each containing multiple views and clinical annotations per patient. These reports were collected during an internship at the Cooperative Ophthalmology Hospital Foundation of Cuba -El Oued in Algeria , under the supervision of Dr Ali Saadoun. These images were obtained using an OCT device and contain accurate structural information that contributes to

the early detection of glaucoma.From each OCT scan, two key images were manually extracted:

- Fundus Image: A 2D top-down view that emphasizes the optic disc, cup, and retinal blood vessels.
- Optic Disc Tomogram: A crosssection of the internal structure of the optic disc.

These regions were cropped manually from the original images using Bulk Image Crop online tool, FigureIII.2 and FigureIII.3 illustrate the extracted fundus and tomogram images

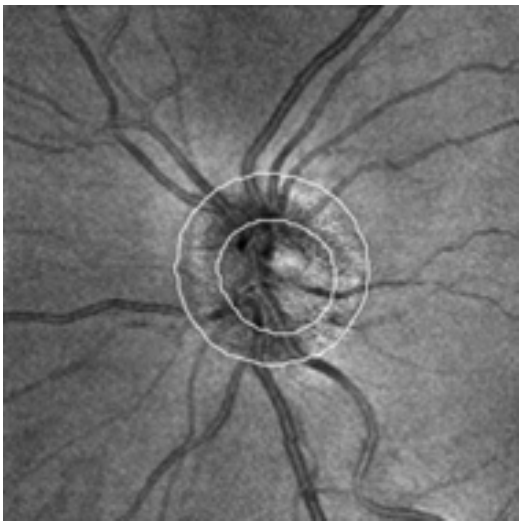

Figure III.2: Cropped fundus image extracted from the OCT report.

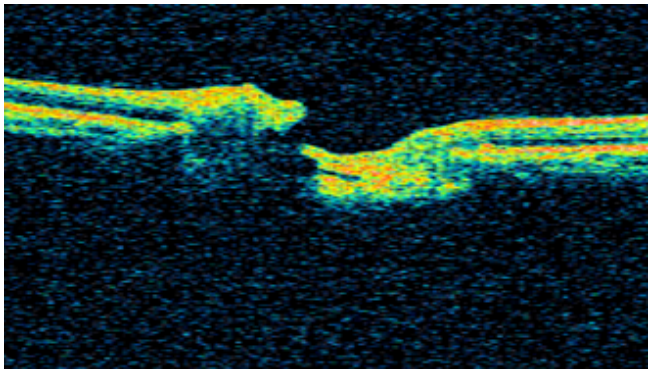

Figure III.3: Cropped optic disc tomogram (B-scan) extracted from the OCT report.

Preprocessing

To improve the quality of the images and ensure their compatibility with the requirements of deep learning models, a series of pre-processing steps aimed at standardising the characteristics of the images were applied. These steps included the following points :

1. Data Annotation and Organization: Before applying preprocessing techniques, all data was organised and annotated in an Excel file (.csv). This file includes patient age,gender, RNFL thickness value, and both oct and fundus image paths.

2. Data Augmentation: Data augmentation techniques were applied to improve image diversity and enhance the generalisability of the model. These techniques were implemented using the `torchvision.transforms` library and summarized in table III.1:

Augmentation Technique	**Parameters**
Horizontal Flip	$p = 1.0$
Vertical Flip	$p = 1.0$
Rotation	degrees = $(-15°, 15°)$
Color Jitter	brightness = 0.1, contrast = 0.1
Random Affine	degrees = 0, translate = $(0.1, 0.1)$
Random Resized Crop	size = 512, scale = $(0.9, 1.0)$

Table III.1: Summary of data augmentation techniques and their parameters

These techniques were applied only to the images, with the clinical data related to each case fully preserved without any modification.

3. Image Enhancement: In the context of improving image quality and preparing them for deep processing, a series of filtering techniques were applied aimed at enhancing contrast, reducing noise, and improving detail clarity. These techniques were applied only to fundus images because they will undergo a subsequent segmentation phase in the processing pipeline, while for OCT images, we will limit ourselves to extracting texture features without applying visual enhancements. However, before applying any enhancement we first converted the images into grayscale to better highlight the medical structures and eliminate the influence of color variations. The quality improvement techniques used include three main steps, which we summarize as follows..

   - **CLAHE** (Contrast-Limited Adaptive Histogram Equalisation): a technique to enhance local contrast by enhancing contrast within small regions of the image known as tiles, To avoid artificial edges between these tiles, the algorithm blends their borders smoothly using bilinear interpolation. This technique effectively improves local

contrast and enhances the visibility of important structures [75]. We implemented CLAHE techninque using a clipLimit factor of 2 and a tile Grid Size of 8×8.

- **Non-local means**: This method differs from traditional filtering techniques that rely on the direct neighborhood of the pixel, as the NLM algorithm searches for similar patches across the entire image, not just in the local neighborhood, making it very effective in preserving fine details.

  The principle of the algorithm is based on the idea that the color of a pixel can be better estimated by the average colors of other pixels that are visually similar to it, even if these pixels are spatially distant. The similarity between patches is calculated using Euclidean distance, and a decreasing weight function is used to determine the contribution of each similar pixel to the final estimate [76].

  The mathematical formula for the NLM algorithm is given as follows:

$$NL[u](p) = \frac{1}{C(p)} \int_{\Omega} f\left(d(B(p), B(q))\right) u(q)\, dq \tag{III.1}$$

  Where:

  - $C(p)$ is the normalizing factor
  - $d(B(p), B(q))$ is an Euclidean distance between image patches centered respectively at p and q
  - f is a decreasing function

- **Unsharp Masking** This method relies on calculating the difference between the original image and a blurred version of it using any image filter, then amplifying this difference and adding it to the original image to obtain a sharper image [77].The general formula for this process is:

$$I_{\text{sharp}} = I_{\text{original}} + \alpha \cdot (I_{\text{original}} - I_{\text{blurred}}) \tag{III.2}$$

  In our case, we used the Gaussian filter with a blur radius equivalent to =1, and the amplification factor was set to 1.3 This means that the original image had the difference between it and its blurred version added after amplifying it by 1.3, using the add Weighted function from cv2 library, which allows merging the two images with different weights. This technique effectively contributes to enhancing edges and fine details, which are important elements in tasks such as segmentation and optical diagnosis.

4. Resizing and Normalization: All images have been resized to a uniform resolution of 512×512 pixels to ensure standardization of input dimensions. Then, a normalization process was performed to adjust the pixel values to match the requirements of the pre-trained deep models, improving training stability and model performance.

To illustrate the impact of the applied preprocessing techniques on the fundus images, Figure III.4 presents a visual comparison between the original image and the enhanced versions results.

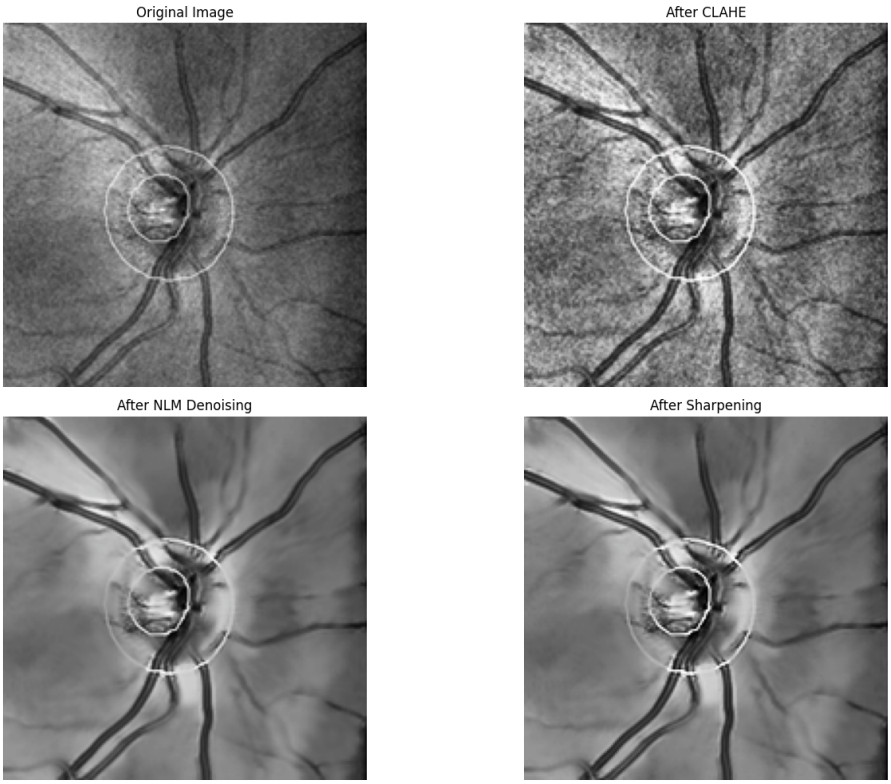

Figure III.4: Progressive enhancement of a fundus image using a filtering pipeline.

III.4 Segmentation of Retinal Vessels based SegFormer

The process of segmenting the retinal blood vessels is a crucial stage in extracting the eye's vascular network, as it plays a pivotal role in the early detection of the initial indicators of glaucoma. During the practical internship conducted at the Cooperative Ophthalmology Hospital Foundation of Cuba El Oued in the city of EL Oued, under the supervision of Dr. Ali Saadoun and in collaboration with a team of ophthalmologists, important clinical observations were recorded that formed the starting point for our project idea. It was observed that many patients whose OCT results showed good indicators of eye health had wide retinal blood vessels. However, subsequent medical follow-up revealed that a large number of them were actually

carriers of glaucoma. It was also found that some cases classified in the high-risk category exhibited the same vascular characteristic. These field results highlight the importance of the structural characteristics of blood vessels as an additional factor that can be exploited to improve the accuracy of glaucoma diagnosis in its early stages. Based on these clinical data, it was necessary to develop a precise methodology for segmenting the optical blood vessels, with the aim of extracting the structural features associated with the risk of developing Glaucoma. To achieve this, the CHASEDB1 database, known for its high quality and data diversity, was relied upon, providing a suitable environment for training the models and testing their effectiveness.

III.4.1 Utilized Datasets for model training

The public CHASEDB1 dataset is a dataset for retinal vessel segmentation which contains 28 color retina images with the size of 999×960 pixels which are collected from both left and right eyes of 14 school children. Each image is annotated by two independent human experts [78]. We expose in the figure III.5 samples from the CHASEDB1 dataset.

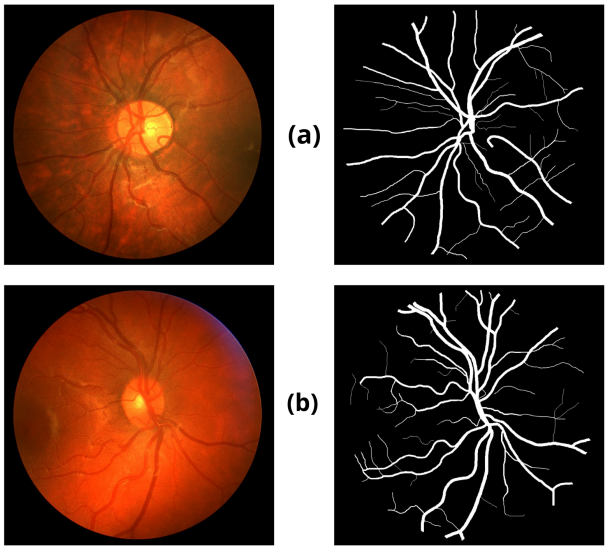

Figure III.5: Samples from the CHASEDB1 dataset: Fundus images and corresponding vessel masks annotated by experts.

III.4.2 Retinal Vessels segmentation pipeline

Preprocessing

Before being fed to the model, the data underwent several preprocessing steps aimed at improving its quality and increasing training efficiency

1. Data Augmentation :Data augmentation techniques were applied to obtain qualified data to feed into the SegFormer model and improve the model's generalization capability. The techniques are summerized in table III.2

Technique	**Parameters**
Horizontal Flip	$p = 1.0$
Vertical Flip	$p = 1.0$
`Rotate`	limit = $\pm 30°$, $p = 1.0$
Random Brightness Contrast	brightness_limit = 0.2, contrast_limit = 0.2, $p = 1.0$
Gaussian Blur	blur_limit = 3, $p = 1.0$

Table III.2: Summary of Applied Augmentation Techniques and Their Parameters

2. Image Enhancement: Achieving high clarity and visibility of anatomical structures is crucial in medical image processing high clarity and visibility of anatomical structures. Therefore, each fundus image was first converted to grayscale (single-channel) to emphasize the structural content and reduce unnecessary color information. However, since the SegFormer model expects three-channel RGB inputs, the single grayscale channel was duplicated across three channels, to create an equivalent image in terms of dimensions without losing any information. After the conversion, the images were enhanced through:

   - Contrast Limited Adaptive Histogram Equalization (CLAHE): to enhance the contrast of fundus images by redistributing the intensity values in localized regions, making subtle anatomical details more distinguishable. We implemented CLAHE techninque using a clip Limit factor of 2 and a tile Grid Size of 8×8.
   - Gaussian filter: to remove datails and noise while preserving important edges and structures. we set a filter kernel of size 7 and Standard deviation in the horizontal direction equal to zero.

3. Resizing and normalization : All images were resized to a standardized resolution of 512×512 pixels as the pre trained model requires to ensure uniform input dimensions across the dataset. As for normalization, it was applied automatically during the preprocessing stage using the feature extractor object of the SegFormer model, which adjusts pixel values to be within the appropriate range for the pre-trained model.

Below in figure III.6 a sample of CHASEDB1 dataset after preprocessing steps

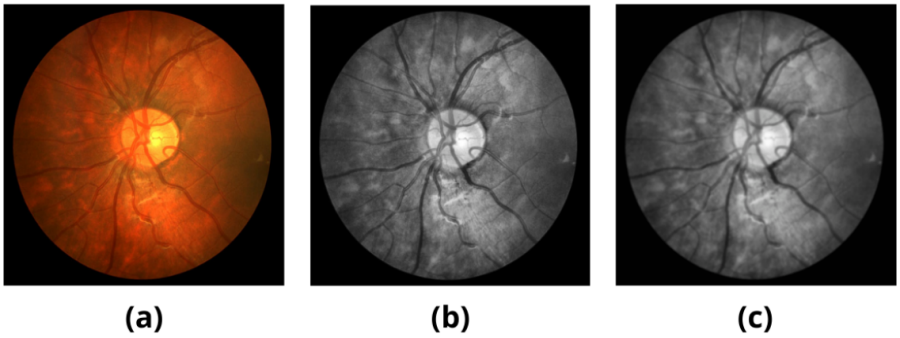

Figure III.6: (a) Original image from the CHASEDB1 dataset, (b) image after applying CLAHE for local contrast enhancement, (c) image after applying both CLAHE and Gaussian filtering for improved contrast and noise reduction.

Training phase

At beginning, Dataset was split into 85% for training set and 15% for validation set. A transfer learning approach were adopted based on SegFormer B2 model, originally fine-tuned on the ADE20K dataset. To adapt the model to our specific task of retinal vessel segmentation, we performed fine-tuning by modifying the final classification head to match the number of required classes. Training was performed using Adam optimizer with a learning rate set to 2×10^{-5} , a batch size of 16, and two data loading workers to enhance processing efficiency. To improve the robustness of training ,Early Stopping was applied based on validation loss monitoring with patience of 10 epochs. Adittionally, a Learning Rate scheduler integrated to dynamically reduce learning rate during training once the validation performance plateaued, allowing the model to converge more effectively. The Model Checkpointing feature was enabled to automatically save the best model performance at every decrease in validation loss. training continued for a maximum of 300 epochs, with performance validation after each complete epoch.

III.5 Segmentation of Optic Disc and optic Cup based Segformer

This stage of segmentation is considered one of the most important fundamental steps in the methodology, as the accuracy of calculating the cup-to-disk ratio is directly related to the precision of extracting the cup and disk masks from the segmentation process. Therefore, the success of this stage has a significant impact on the system's effectiveness.

III.5.1 Utilized Datasets for model training

The dataset used to train the optic disc and cup segmentation model was composed of 25 manually selected fundus images, extracted from the original dataset. These images were annotated under the supervision of Dr. Ali Saadoun using Fiji (ImageJ) software. Each image was labeled with two separate binary masks: one for the optic disc and another for the optic cup. To facilitate model training, both masks were subsequently combined into a single multi-class mask. This specialized characterization in the field provided highly reliable comments tailored to fit our medical context.

III.5.2 Optic Disc and Optic Cup segmentation pipeline

Preprocessing

The preprocessing steps discussed in the retinal vessels segmentation section were adopted to ensure standardisation of image quality and improve compatibility with the deep learning models used in the later stages. FigureIII.7 shows a sample of the utilized dataset before and after preprocessing

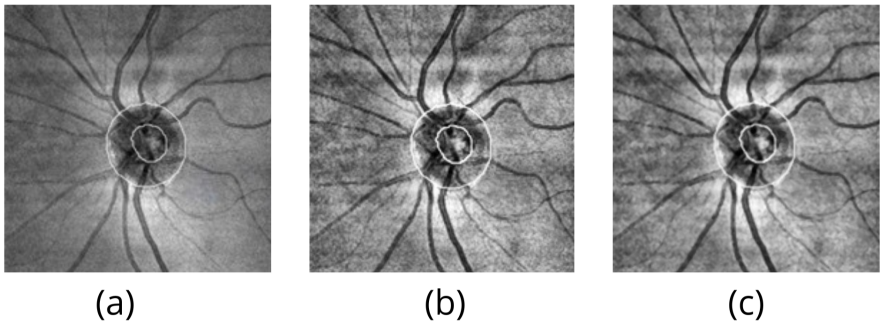

Figure III.7: (a) Original image from the custom dataset, (b) image after applying CLAHE for local contrast enhancement, (c) image after applying both CLAHE and Gaussian filtering for improved contrast and noise reduction.

Training phase

To perform the segmentation of the optic disc and cup, we used transfer learning with the pretrained SegFormer B2 model on the REFUGE dataset—a benchmark dataset specifically designed for glaucoma-related optic disc segmentation tasks.This provided a strong foundation of knowledge in the same field. The final classification head of the model was adjusted to fit our binary classification task (disk and cup), and then training was resumed on our labeled local dataset. Training was conducted using:

- Adam optimizer with a learning rate of 2×10^{-5}
- Batch size: 16
- Data loading workers: 2

To enhance convergence and model robustness, we applied:

- Early Stopping: Patience of 10 epochs based on validation loss
- Learning Rate Scheduler: Reduced the LR when performance plateaued
- Model Checkpointing: Automatically saved the best model based on validation loss

III.6 Extraction of Multi-Source Features for Glaucoma Diagnosis

In order to build a robust and accurate early detection system for glaucoma, we adopted a feature extraction approach from multiple sources. This strategy includes integrating features

derived from the various anatomical and structural components of the eye, using both OCT volumes and fundus images. The extracted features capture changes related to shape, geometry, and texture that are typically associated with damage caused by glaucoma. These features are essential inputs for the next stage of classification. We present the various types of extracted features.

III.6.1 Features Extracted from Segmentation results

A set of morphological and geometric features that represent the changes in the eye structure associated with glaucoma was extracted based on the obtained segmentation results.

1. Features obtained from retinal vessels regions: As we mentioned earlier about the importance of retinal vessels structure in determining the early symptoms of Glaucoma, we relied on the skeletonization technique applied to the retinal vessels region extracted from the fundus image, which enabled us to represent the vessels in a linear manner . We calculated a set of properties related only to the main blood vessels, ignoring small vessels with insufficient length or area. The main vessels were identified based on a minimum length and area threshold to ensure the exclusion of noise and non-significant structures. Specifically, the following criteria were calculated:

    - The overall average diameter of the major vessels: This indicator reflects the overall thickness of the vascular network, which may indicate the presence of vascular changes associated with the early stages of optic nerve diseases.
    - Maximum recorded diameter: Highlights the presence of dilated blood vessels, which may be an early sign of eye pressure disorders.
    - Number of major vessels: It represents the numerical density of important vessels, as a decrease in this number may indicate a deterioration in retinal blood flow.

    These combined characteristics provide an accurate quantitative view of the state of the retinal blood vessels and contribute to improving the accuracy of diagnostic models based on fundus characteristics.

2. Features obtained from Optic Disc and Cup regions: After performing the segmentation process on both the optic disc and the optic cup, and obtaining precise masks thanks to the SegFormer model, we extracted the key feature used in evaluating structural changes associated with glaucoma, which is the cup-to-disc ratio (CDR).

- Extracting the cup area: It was calculated by the number of pixels classified as "cup" in the resulting mask.
- Extracting the disc area: It represents the total area of the cup with the rim, which is the area surrounding the cup within the disc boundaries.
- CDR calculation :

$$\text{CDR} = \frac{\text{Cup Area}}{\text{Disc Area}} \tag{III.3}$$

- Converting to millimeters: we converted the areas to square (mm^2)using the conversion factor 1pixel(X) = 0.2645833333 mm

The converted CDR value is used as an important input unit in the classification stage, as it helps distinguish between healthy cases and those potentially affected by glaucoma, and it is considered one of the fundamental pillars of our smart early diagnosis system.

III.6.2 Features Extracted within Fundus Images and OCT scan

In order to further improve the performance of the model in identifying early glaucoma, we extracted an additional set of features from fundus and optic disc tomogram images.

Features Extracted from Fundus Images

Because retinal blood vessels exhibit clear directional and linear alignment, a Gabor filter has been used to emphasize and analyze the pattern of vessels in fundus images. With this approach, fine details of the vascular structure can be accurately identified to achieve notable features such as vessel number, thickness, and distribution. Such parameters are particularly useful in glaucoma detection during its early stages.

The Gabor filter is a very effective tool employed to examine textures and it combines a Gaussian envelope with a sinusoidal wave in such a way that it best suits the purpose of detecting edges and patterns of specific orientation and frequency. Its ability to capture both spatial as well as frequency information has made it one of the most commonly used filters in many image processing operations[79]. Mathematically, Gabor filter is given by:

$$h(x,y) = \frac{1}{\sigma_x \sigma_y 2\pi} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \exp\left(j2\pi u_0 (x\cos\theta + y\sin\theta)\right) \tag{III.4}$$

Where:

- u_0 is the spatial frequency, which defines the number of cycles of the wave per unit distance.

- θ is the orientation angle of the sinusoidal wave, which specifies the direction of filter application.
- σ_x **and** σ_y are the standard deviations of the Gaussian envelope in the x and y directions, respectively. These specify the spatial scale and spread of the filter.
- The $exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right)$ is a space-anchored, two-dimensional Gaussian function.
- The $exp\left(j2\pi u_0(x\cos\theta + y\sin\theta)\right)$ is a complex sinusoidal plane wave that modulates the Gaussian and gives orientation- and frequency-tuned filtering capabilities.

This formula provides for the capability of the filter to identify retinal vessel-specific features at numerous different orientations in order to more effectively analyze changes in vessels related to glaucoma.

In our approach, we constructed a Gabor filter bank using four orientations: 0°, 45°, 90°, and 135°, to ensure comprehensive directional coverage. The filter bank was generated using a kernel size of 21, a standard deviation (sigma) of 3.0, a wavelength (lambda) of 10.0, and an aspect ratio (gamma) of 0.5. These parameters were chosen to effectively capture vessel patterns with varying orientations and scales across the retinal fundus images. In each of the four orientations , mean, standard deviation, and maximum response values resulting from each filter were calculated .These values represent quantitative characteristics of the vascular tissue in each direction and are used as distinctive features that are later fed into the final classifier to improve the prediction accuracy of early signs of glaucoma.

Features Extracted from OCT Images

To further enrich the structural information obtained from OCT images, Gabor filters were applied to highlight fine-grained textural differences between retinal layers. These filters are capable of capturing orientation- and frequency-based texture features, which can reflect subtle pathological changes linked to early glaucoma.

Unlike direct measurements such as retinal layer thickness, Gabor-based features provide insights into spatial texture variations within the retinal cross-sections. These textural patterns offer complementary diagnostic information by revealing structural irregularities that may not be evident through conventional analysis.

In our implementation, we used a Gabor filter bank with the following parameters: kernel size

= 21, $\sigma = 3.0$, $\lambda = 10.0$, $\gamma = 0.5$, and orientations = $[0°, 45°, 90°, 135°]$. These settings were chosen to capture diverse directional features across the retinal layers, contributing to improved discrimination between normal, threatened, and glaucomatous cases.

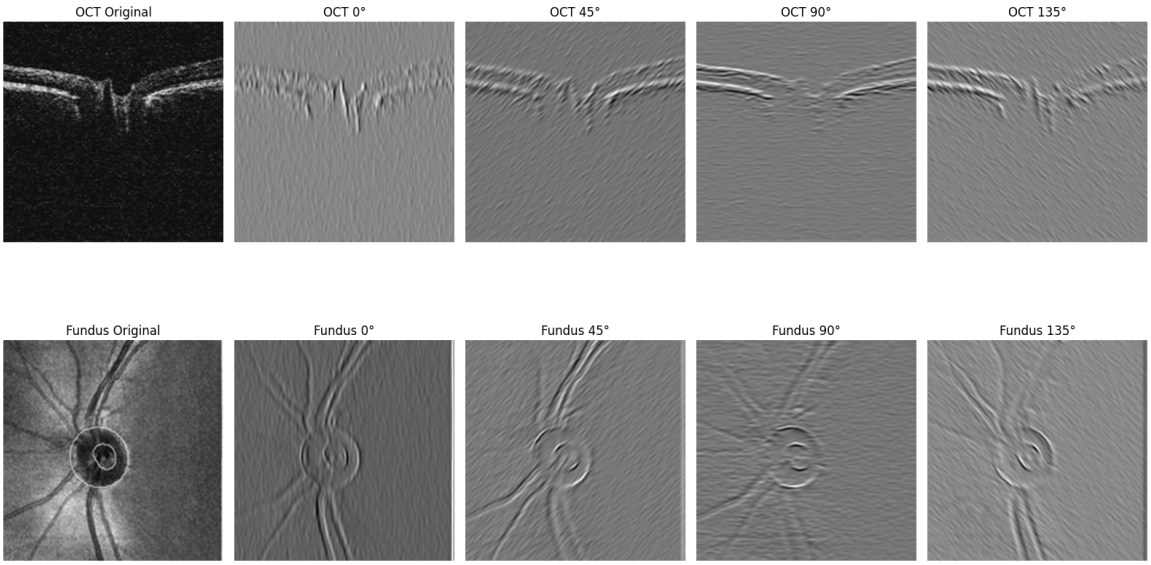

Figure III.8: Gabor filter application on both OCT and fundus image at four orientations (0°, 45°, 90°, and 135°).

Below is a table summarizing all the extracted and used features in training the final classification model.

Feature Type	**Image Source**	**Extracted Features**	**Diagnostic Purpose**
Optic Disc & Cup Morphology	Optic Disc & Cup segmentation masks	- Cup Area - Disc Area - Cup-to-Disc Ratio (CDR)	Detect structural enlargement of cup area
Retinal Vessels Characteristics	retinal vessels segmentation masks	- Average vessels diameter - Maximum diameter - Number of major vessels	Reveal vascular wideness linked to glaucomatous damage
Fundus Texture Features	Fundus photography in HRT	- Gabor filter responses (multi-scale, multi-orientation) - Texture statistics	Capture texture irregularities in the whole fundus image
Optic disc Tomogram Features	OCT scan in HRT	- Gabor filter responses (multi-scale, multi-orientation) - Texture statistics	Assess microscopic damage in nerve fiber layers and structural cupping
Demographic Data	Clinical records (patient data)	- RNFL thickness - Age - Gender	RNFL thickness, Age and gender influences glaucoma risk and progression. Used to personalize diagnosis.

Table III.3: Summary of Extracted Features for Glaucoma Diagnosis

III.7 Features Fusion and Classification

In this section, we adopted the SAINT (Self-Attention and Intersample Attention Transformer) architecture as the backbone of our final classification model, SAINT is a modern model used for tabular data processing. To train the classifier we first merged all the extracted features and saved them in a CSV file to ensure proper alignment with the corresponding fundus and OCT images. To perform the final classification step, we implemented a custom deep learning

classifier based on the SAINT architecture. The data was randomly split into 75% training and 10% validation and 15% testing sets using stratified sampling to maintain class balance. A preprocessing pipeline was applied using StandardScaler for numerical features and OneHotEncoder for categorical variables (gender). Feature engineering steps included interactions such as age × number of vessels and CDR × vessel diameter. After transformation, the resulting input vector was passed to the SAINT classifier. The hyperparameters used for training the SAINT model are summarized in the table below:

Hyperparameter	**Value**
`dropout`	0.3
`epochs`	150
`batch_size`	32
`learning_rate (lr)`	0.001
`weight_decay`	0.0001
`scheduler`	Cosine Annealing
`patience (early stopping)`	10
`class_weights`	Inverse of class frequency

Table III.4: Hyperparameters used for training the SAINT classifier

III.8 Conclusion

In this chapter, we presented a comprehensive and straightforward strategy that uses different paths to extract features to help improve glaucoma detection. This strategy combines several techniques, such as the SegFormer model, Gabor filters, and data from OCT images. By using these methods together, we aim to enhance the system's ability to locate the parts of the eye related to glaucoma. This approach shows how deep learning and image analysis can work together to support the early diagnosis of glaucoma.

In the next chapter, we will present and evaluate the experimental results of the proposed methodology, focusing on the performance of the segmentation and classification stages.

CHAPTER IV

EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

IV.1 Introduction

This chapter is dedicated to reviewing the results of the implementation of the methodology described in the previous chapter. First, we summarise the software tools and algorithms used during the implementation. Next, we present the results obtained after the implementation of the methodology. We then discuss the empirical results and assess their conformity with the research objectives and scientific quality standards.

IV.2 Technical Framework and Tools

IV.2.1 Python language

Python is a high-level, object-oriented programming language with dynamic behaviour in terms of binding variables and types. It is designed to be easy to read and quick to develop, making it suitable for rapid application development as well as a scripting language for linking software components. Python provides high-level built-in data structures and supports modules and packages, encouraging code reuse and programme organisation. The Python interpreter is widely supported across all major platforms and is free and open source. Python is recognised for its ability to accelerate the editing, testing and debugging cycle by eliminating the need for a pre-compilation phase, increasing programmer productivity. Furthermore, the language has advanced debugging tools, and its debugger is written in Python itself, reflecting its introspective power [80].

IV.2.2 Kaggle

Kaggle is an online platform that hosts the largest global community of data scientists, with more than 536,000 active members across 194 countries, and around 150,000 engagements per month. The platform offers powerful tools and diverse resources to support developments in the field of data science. Similar to modern education platforms such as DataScientest, Kaggle offers a customised online Jupyter Notebooks environment with no prior setup required, free access to graphical processing units (GPUs), along with a huge library of community-shared data and code, which includes over 50,000 public datasets and 400,000 publicly available notebooks [81]. In general, Kaggle allows users to:

- Search and publish datasets.
- Explore and build models in an online environment.

- Collaborate with other professionals and amateurs to share experiences.
- Participate in competitions that push them to innovate and evolve [81].

Kaggle is trusted by many leading data science companies. The platform allows professionals and developers to participate in machine learning competitions, write and share code, and host datasets[81].

IV.2.3 Visual Studio Code

Visual Studio is an integrated development environment (IDE) that supports the entire software development lifecycle, offering advanced features such as code editing, debugging, building, and application deployment. It integrates compilers, code completion, graphical design tools, and other utilities to streamline development. In contrast, Visual Studio Code is a lightweight yet versatile source code editor compatible with Windows, macOS, and Linux. It natively supports languages like JavaScript, TypeScript, and Node.js, and can be extended to support many others, including C++, C#, Java, Python, PHP, Go, and .NET through a vast extension ecosystem[82].

IV.3 Hardware Configuration

IV.3.1 Local Devices

Table IV.1 shows the specifications of the local devices used during the training.

Component	Machine 1	Machine 2
Manufacturer	Lenovo	HP
Model	82C4	HP EliteBook 840 G3
Processor	Intel(R) Core(TM) i3-1005G1 CPU @ 1.20GHz 1.19 GHz	Intel(R)Core(TM)i5-6300U @2.40GHz – 2.50GHz
Hard Disk	256 GB SSD	512 GB SSD
RAM	4GB	8 GB
Operating System	Windows 11 Pro	Windows 10

Table IV.1: Specifications of the local hardware used.

IV.3.2 Online Environment

In addition to using local hardware, the training process was also conducted on the Kaggle platform, which offers access to a P100 GPU. This environment provided substantial computational power, significantly accelerating the training of deep learning models and allowing for more efficient experimentation.

IV.4 Model Performance Evaluation

After a SAINT transformer model is trained, the next crucial step is to evaluate its performance. It is typically performed on a test dataset not seen by the model during training. Choosing the right metrics for evaluation is crucial, as it determines our view of the model's strengths and weaknesses, especially in sensitive domains like medical diagnosis.

IV.4.1 Confusion Matrix

Confusion matrix is a simple table that is used to estimate classification models by comparing labels that are predicted to the actual labels, it is also an important tool in statistics, data mining, machine learning models, and artificial intelligence applications as a whole [83].

In Figure IV.4.1, an example of a confusion matrix of a binary classification model is presented:

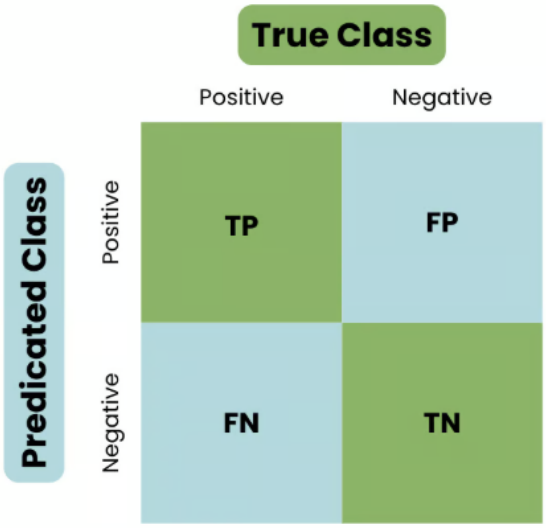

Figure IV.1: The structure of a confusion matrix.

The four main elements of the confusion matrix are [83]:

- True Positives (TP): Cases that were correctly classified as positive.
- False Positives (FP): cases that were incorrectly classified as positive.
- True Negatives (TN): Cases that were correctly classified as negative.
- False Negatives (FN): cases that were incorrectly classified as negative.

IV.4.2 Accuracy

Accuracy is one of the most critical measures employed for the evaluation of the performance of classification models. It indicates the proportion of precise predictions made by the model either positive or negative among all instances. Accuracy is calculated using the formula below [57] :

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{IV.1}$$

While accuracy provides an overall indication of model performance, it can be misleading in the case of imbalanced datasets, such as in glaucoma detection, where positive samples are rare. In such cases, other metrics such as recall, precision, and F1-score offer more insightful evaluations of model performance.

IV.4.3 Precision and Recall

Precision and recall are fundamental metrics that are utilized to evaluate the quality of classification models, especially when dealing with imbalanced datasets.

- **Precision**: also known as the positive predictive value, measures the proportion of relevant instances among the retrieved instances. In other words, it indicates how many of the predicted positive cases are actually correct. It is calculated using the formula [57]:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{IV.2}$$

- **Recall**: also called sensitivity or true positive rate, measures the proportion of actual positive instances that were correctly identified by the model. It is calculated as [57]:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{IV.3}$$

IV.4.4 F1-Score

The F1-score is a metric that gives a balanced mesurement between precision and recall. It is calculated by the formula[57]:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{IV.4}$$

This metric is particularly useful in clinical diagnosis issues such as glaucoma diagnosis, where it is crucial to minimize both false positives and false negatives due to the catastrophic consequences of wrong predictions.

IV.5 Segmentation Results within Fundus Images

IV.5.1 Retinal Vessels Segmentation Results

The performance of the SegFormer model for retinal blood vessel segmentation was evaluated on the validation set. After data augmentation, the number of images increased to 168. By applying the Early Stopping technique based validation loss, the best performance of the model was recorded in epoch number 144.At this stage,the intersection over union (IoU) reached 79.5%, loss of 0.087 and accuracy of 87%. These results reflect the model's ability to learn

effectively without overfitting, the results of validation IoU versus epochs for the SegFormer b2 segmentation model trained on CHASEDB1 dataset can be shown in the figure IV.2

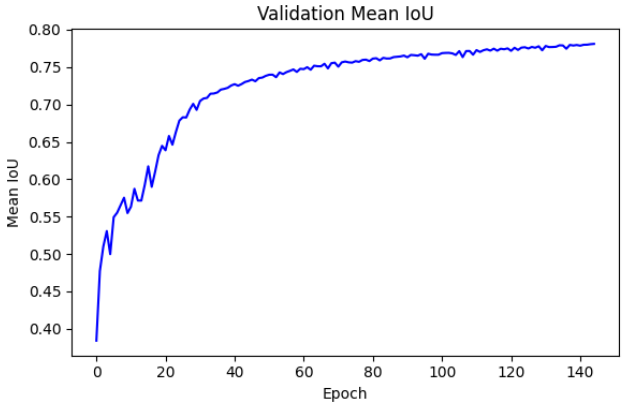

Figure IV.2: Mean IoU on the validation set during training based SegFormer b2 for retinal vessels segmentation.

Figure IV.3 shows the evolution of the loss value across epochs, where a clear stability in the values can be seen during the last epochs, reflecting the model's performance stability and no further improvement on the validation set. This behavior is a direct indication of the effectiveness of applying the Early Stopping technique, which allowed for the automatic cessation of training upon reaching the best possible performance, while avoiding overfitting.

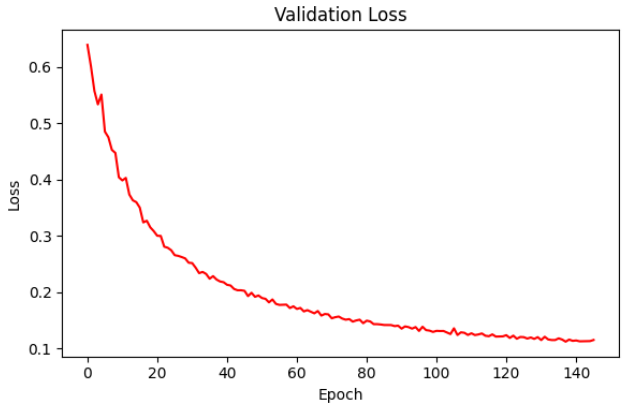

Figure IV.3: Evolution of the average loss during training epochs based SegFormer b2 for retinal vessel segmentation.

To affirm the performance of the Segformer b2 segmentation model, we present in figure IV.4 a visual comparison between the ground truth masks of the CHASEDB1 dataset and the predicted masks generated by the model, in order to visually assess the accuracy of the match between the expected and actual results.

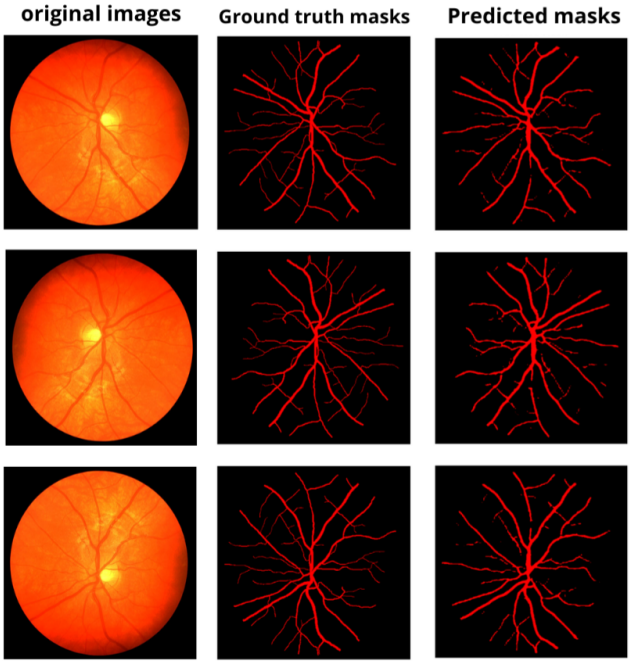

Figure IV.4: Samples of CHASEDB1 input images, Ground truth masks and predicted masks.

After evaluating the model on the training data, we also tested it on our local dataset, the resulting predicted masks can be presented in the figure IV.5.

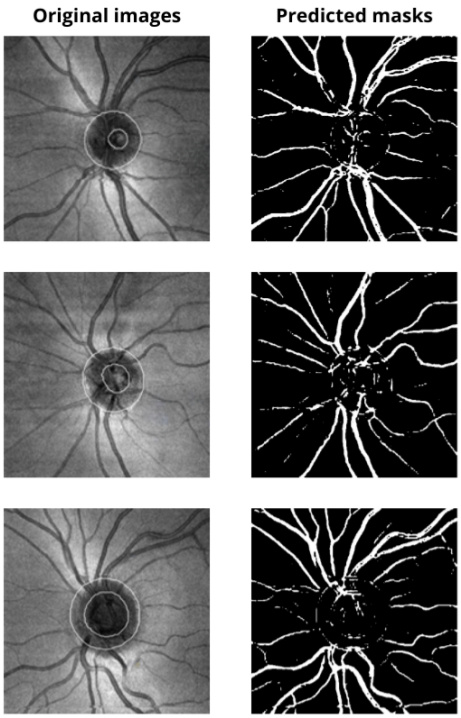

Figure IV.5: Blood vessels segmentation results on the local dataset

IV.5.2 Optic Disc and Optic Cup Segmentation Results

The performance of the segmentation model for the optic disc and optic cup was quantitatively evaluated using validation data. After applying data augmentation, the number of images increased to 160 image. The model demonstrated a satisfactory capability in delineating these important anatomical structures, which are essential for glaucoma diagnosis.

The obtained results are as follows: an Intersection over Union (IoU) score of 82.6% a validation loss of 0.044 and accuracy of 89.4%. These metrics indicate that the model achieved a high level of precision and consistency in segmenting the optic disc and cup regions in fundus images.

Since the segmentation model was trained directly on the manually annotated images by Dr. Ali Saadoun, the results obtained reflect the model's true performance in the same clinical context in which the data was collected. In figure IV.6, we present a visual comparison between the original ground truth masks for the disc and cup, and the Predicted Masks produced by the SegFormer b2 trained model.

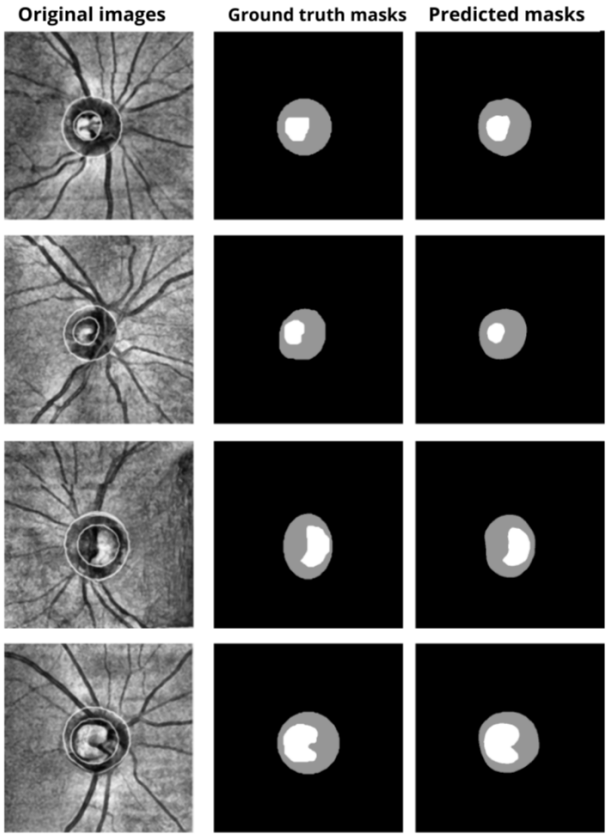

Figure IV.6: Samples of the original OCT-based retinal images, manually annotated masks(ground truth) and predicted masks.

The experimental performance of the segmentation model is summarized for each architec-

ture evaluated in this study. These metrics reflect the model's ability to accurately and reliably identify key anatomical structures. The values of these metrics for retinal vessels, optic disc and optic cup are summarized in the table IV.2.

Task	**loss**	**IoU**	**Accuracy**
Retinal vessels	0.087	0.795	0.872
Optic disc and cup	0.044	0.826	0.894

Table IV.2: Segmentation performance metrics for each task

IV.6 Classification results based SAINT

The classification performance of the SAINT based classification model was evaluated using training, validation and testing datasets. After applying data augmentation, the number of training samples increased to 1,283, significantly enhancing the model's ability to generalize and reducing the risk of overfitting. The model achieved a validation accuracy of 95.41% and a testing accuracy of 94.27%, reflecting a strong generalization ability and effective learning behavior. Figure IV.7 presents the confusion matrix generated from the test data. The matrix reveals that the model correctly classified the majority of instances across all three categories: Normal, Suspected, and Glaucomatous cases. Most misclassifications occurred between the Normal and Suspected classes, which is expected due to their subtle clinical differences.

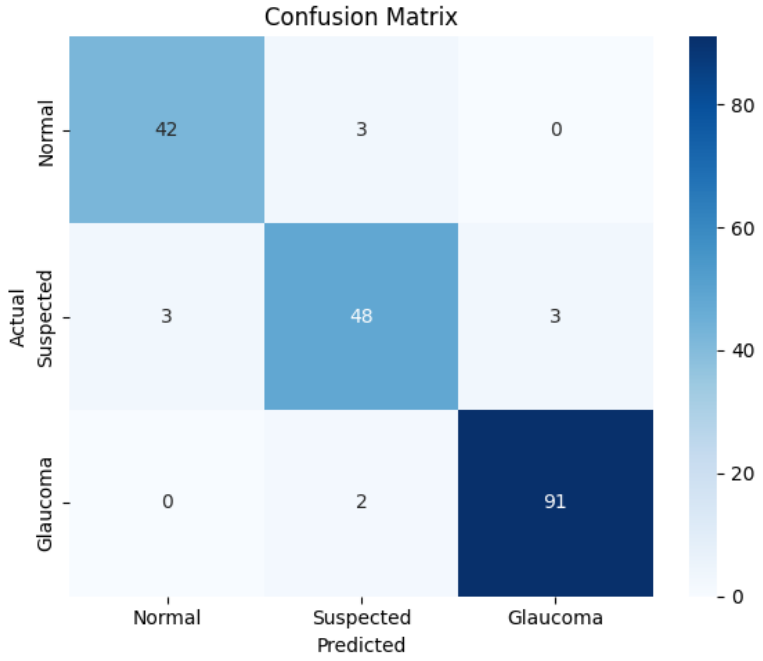

Figure IV.7: Confusion matrix showing classification performance of SAINT model on test set.

The detailed classification report is summarized in table IV.3. The model achieved high precision, recall, and F1-scores across all classes. Notably, the Glaucoma class achieved the highest F1-score of 0.98, indicating the model's robustness in detecting this critical condition.

Class	**Precision**	**Recall**	**F1-score**
Normal	0.93	0.93	0.93
Suspected	0.91	0.89	0.90
Glaucoma	0.97	0.98	0.97

Table IV.3: Classification report of the SAINT model on the test set.

The evolution of training and validation accuracies and losses over epochs is illustrated in figure IV.8

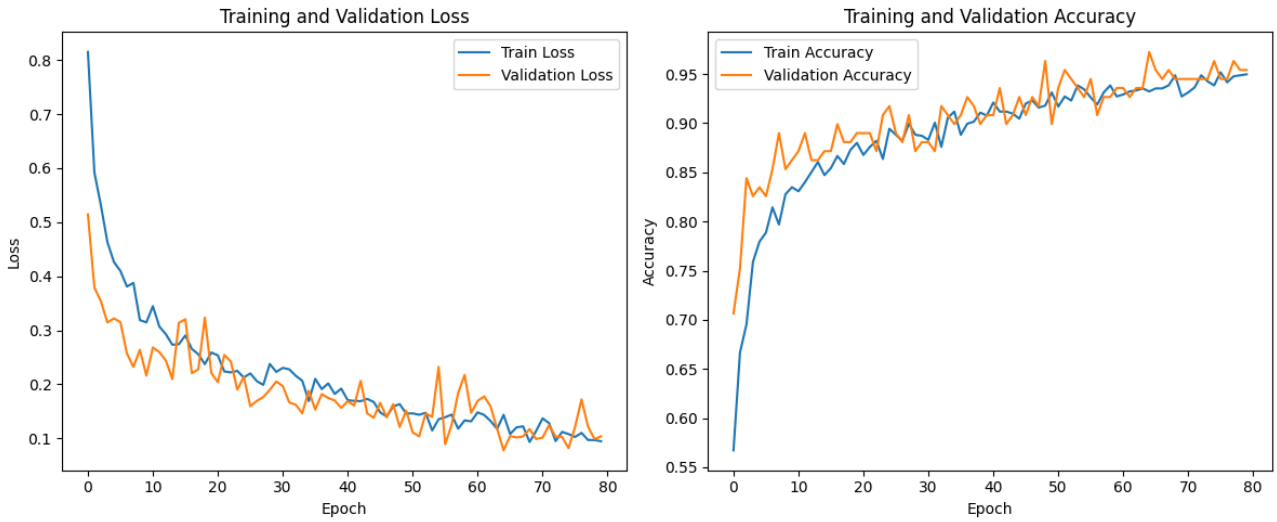

Figure IV.8: Training and validation accuracy and loss curves of the SAINT classifier over epochs.

IV.7 Comparison with Previous Works

Table IV.4 presents a comparison between our proposed method and several state-of-the-art approaches that employ both segmentation and classification for glaucoma detection. These hybrid methods typically start with a segmentation phase to extract anatomical structures such as the optic disc and cup, followed by the computation of diagnostic features which are then used as input to classification models.

Method	**Dataset**	**Accuracy**
U-Net for optic disc/cup segmentation, followed by SVM classification[71]	DRISHTI-GS	93.33%
WSMTL: segmentation and classification [84]	ORIGA650	92.4%
Transferable Ranking Convolutional Neural Network (TRk-CNN) multi classififcation [67]	Glaucoma image dataset(Korea university medical center)	92.96%
Proposed approach	Private OCT reports collected at Cooperative Ophthalmology Hospital El Oued, Algeria	94.27%

Table IV.4: Comparison of the proposed approach against state-of-the-art approaches for glaucoma detection

As Table IV.4 shows, the proposed method outperforms previous methods that combine segmentation and classification in terms of accuracy. This is due to the use of the SegFormer model, which provides accurate segmentation, and the SAINT model, which contributes to a more accurate classification thanks to its ability to understand the spatial relationships between the extracted clinical features.

IV.8 Case Study: Correction of Device Misdiagnoses by the Proposed Model

In this section, we present a focused evaluation of the proposed model on a set of 16 new OCT report images not seen during training or validation. These images were manually annotated and verified by ophthalmologist Dr. Ali Saadoun.

The goal of this analysis is to identify cases where the standard diagnostic interpretation provided by the OCT device was inaccurate, and to compare it with the prediction of our

model. For each case, the doctor's expert opinion was used as the ground truth.

Table IV.5 presents selected examples where the device either failed to detect glaucoma or gave a misleading diagnosis, while our model successfully classified the image correctly, in alignment with the doctor's final assessment.

Case ID	Ground Truth (Doctor)	OCT Device Diagnosis	Model Prediction
C1	Suspected	Normal	Suspected 97.93%
C2	Suspected	Normal	Suspected 99.06%
C3	Suspected	Normal	Suspected 99.84%
C4	Glaucoma	Normal	Glaucoma 84.26%
C5	Glaucoma	Suspected	Glaucoma 100%
C6	Normal	Normal	Normal 99.86%
C7	Suspected	Normal	Suspected 99.70%
C8	Normal	Normal	Normal 70.08%
C9	Normal	Normal	Suspected 86.61%
C10	Glaucoma	Glaucoma	Glaucoma 100%
C11	Glaucoma	Normal	Glaucoma 99.95%
C12	Glaucoma	Glaucoma	Glaucoma 100%
C13	Suspected	Normal	Suspected 82.75%
C14	Normal	Normal	Normal 98.02%
C15	Normal	Normal	Suspected 98.96%
C16	Susppected	Normal	Suspected 99%

Table IV.5: Comparison of the OCT device diagnosis and the model's prediction against the expert-labeled ground truth

IV.9 System Interface Overview

Figure IV.9 shows the main interface of the developed system. The displayed interface represents the intelligent glaucoma diagnosis system. It has been organized clearly and simply

to facilitate data entry and diagnosis operation, consisting of a prominent title at the top of the interface (Glaucoma Diagnosis System), followed by a section dedicated to entering patient information, then a section dedicated to uploading the fundus and OCT scan images, and finally a diagnosis button.

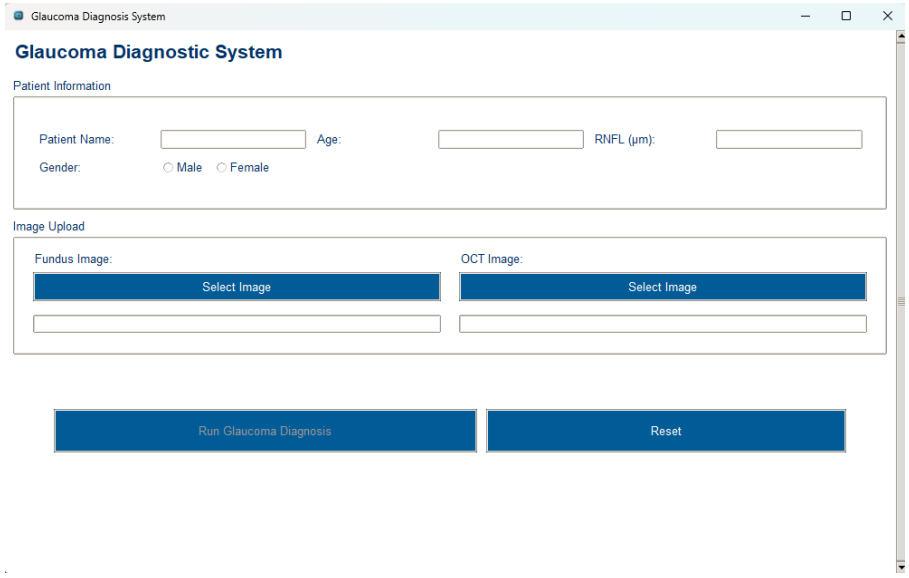

Figure IV.9: Glaucoma Early Detection System Main Interface.

After pressing the diagnosis button, a separate window appears displaying the patient's information and the diagnosis results as shown in the figure **??** while figure IV.11 presents Visualization of the retinal vessels and optic disc and cup results provided by the system.

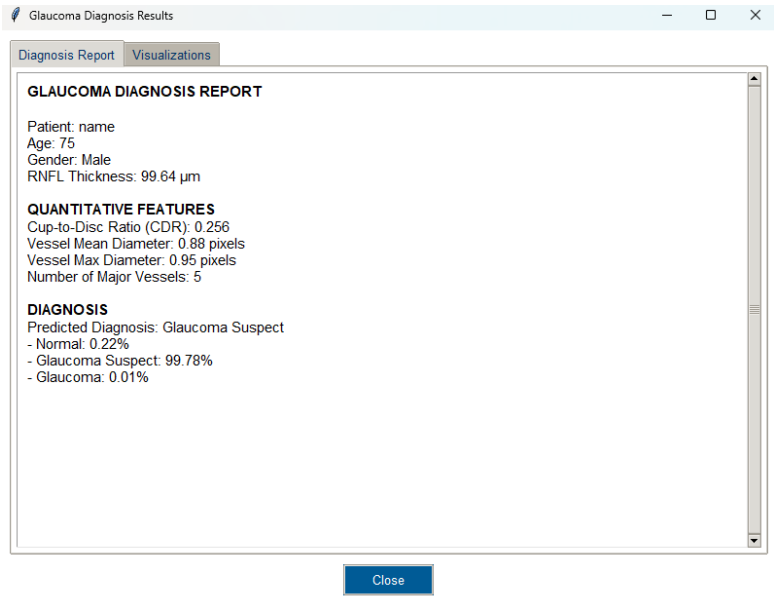

Figure IV.10: Diagnosis report provided by the system.

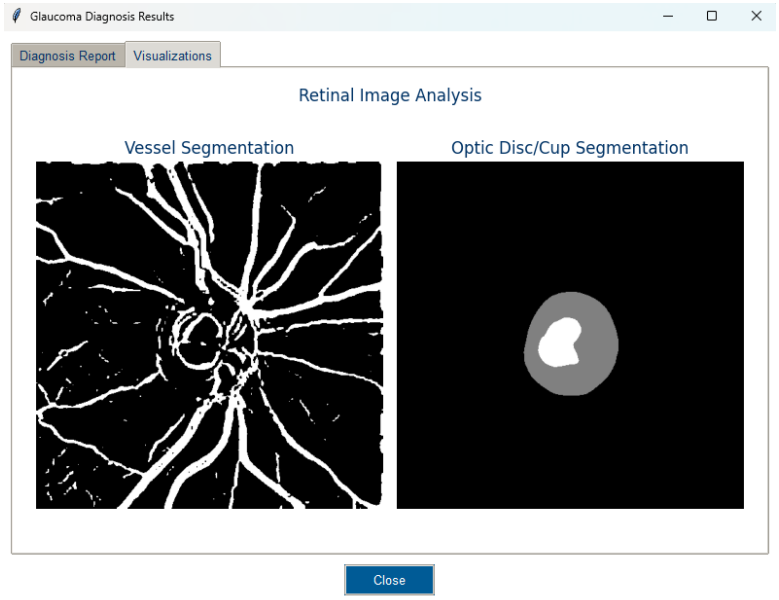

Figure IV.11: Visualization of the segmentation results provided by the system.

IV.10 Discussion

This study aims at early detection of glaucoma using deep learning techniques applied to OCT images. A series of experiments were carried out to evaluate the effectiveness of the proposed model, which included the stages of segmenting retinal vessels and optic disc using SegFormer, and then classifying the images using the SAINT model.

The experimental results indicate that the proposed model performs well in classifying glaucoma cases with high accuracy. The confusion matrix showed that the model is able to effectively discriminate between healthy and diseased cases, with calculated values for Precision, Recall, and F1-score reaching encouraging levels compared to previous studies.

It was also found that using the SegFormer model in the segmentation phase helped to improve the quality of visual feature extraction, which enhanced the performance of the classification phase. This is attributed to the model's accuracy in characterising optical vessels and important structures in OCT images.

Overall, this study emphasises the importance of integrating segmentation models and clinical feature extraction with modern classification techniques to enhance the accuracy of early glaucoma detection. This work provides a basis for the development of intelligent aids for ophthalmologists in the early diagnosis of this silent disease.

IV.11 Conclusion

In this chapter, we have presented our proposed methodology for early detection of glaucoma based on deep learning techniques, through three main stages: Segmentation of retinal vessels using the SegFormer model, extraction of clinically important features such as cup-to-disc ratio (CDR) and retinal nerve fibre layer thickness (RNFL), and final classification of the eye condition using the Transformers-based SAINT model.

The experimental performance of the models was presented and tracked through graphs of accuracy, loss and intercept rate (IoU) indices, showing the effectiveness of the proposed methodology in optimising diagnostic results. The work was enhanced by the development of an interactive graphical interface that facilitates patient data entry, automatic uploading and analysis of medical images, as well as the presentation of the final results in a simplified manner to help physicians make the right decision.

GENERAL CONCLUSION

Glaucoma is one of the world's most serious diseases, due to its progressive and silent nature that leads to permanent vision loss if not detected in its early stages. Recent studies suggest that more than 111.8 million people may develop glaucoma by 2040, highlighting the urgent need to develop smart and effective tools for early detection [16]. The World Health Organisation and other research confirms that late diagnosis is one of the biggest challenges in ophthalmology, especially in resource-limited countries [14].

Our work is part of a research and development effort that aims to utilise Artificial Intelligence (AI) technologies to provide an effective solution for early stage glaucoma diagnosis from OCT images.

Initially, a comprehensive review of previous works was presented, focusing on smart models applied in the medical field and the difficulties in glaucoma diagnosis. Then, we proposed a methodology based on three main stages: First, segmentation of retinal vessels using the SegFormer model.second, extraction of critical clinical features such as cup-to-disc ratio (CDR), mean vessel diameter,and cup depth profiles . Finally, classifying cases as normal, early-stage, or progressive glaucoma.

The results obtained showed good performance in both the segmentation and classification phases, allowing to support clinicians in decision making. Building on these findings, the study paves the way for promising future research directions. In particular, future work will focus on integrating intraocular pressure (IOP) measurements and visual field test results to develop a fully automated glaucoma detection model that adheres closely to established clinical diagnostic criteria.

Overall, this work represents a significant contribution to the early detection of glaucoma and underscores the vast potential of modern AI techniques in advancing prevention strategies and enhancing the accuracy of diagnosing this serious and often asymptomatic disease.

BIBLIOGRAPHY

[1] W. S. Noble, "What is a support vector machine?" *Nature biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.

[3] K. O'shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458*, 2015.

[4] L. L. Scientific, "Enhancing glaucoma diagnosis: Deep learning models for automated identification and explainability using fundus images," *Journal of Theoretical and Applied Information Technology*, vol. 102, no. 13, pp. 5346–5363, 2024.

[5] B. Goutam, M. F. Hashmi, Z. W. Geem, and N. D. Bokde, "A comprehensive review of deep learning strategies in retinal disease diagnosis using fundus images," *IEEE Access*, vol. 10, pp. 57 796–57 823, 2022.

[6] A. U. Rehman, I. A. Taj, M. Sajid, and K. S. Karimov, "An ensemble framework based on deep cnns architecture for glaucoma classification using fundus photography." *Mathematical Biosciences and Engineering: MBE*, vol. 18, no. 5, pp. 5321–5346, 2021.

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

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, . Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[9] [Online]. Available: https://www.geeksforgeeks.org/region-and-edge-based-segmentaion/

[10] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "Segformer: Simple and efficient design for semantic segmentation with transformers," *Advances in neural information processing systems*, vol. 34, pp. 12 077–12 090, 2021.

[11] G. Somepalli, M. Goldblum, A. Schwarzschild, C. B. Bruss, and T. Goldstein, "Saint: Improved neural networks for tabular data via row attention and contrastive pre-training," *arXiv preprint arXiv:2106.01342*, 2021.

[12] [Online]. Available: https://www.astorinoeyecare.com/blog/glaucoma-who-is-a-suspect

[13] A. Sánchez-Morales, J. Morales-Sánchez, O. Kovalyk, R. Verdú-Monedero, and J. L. Sancho-Gómez, "Improving glaucoma diagnosis assembling deep networks and voting schemes," *Diagnostics*, vol. 12, p. 1382, 06 2022.

[14] A. K. Schuster, C. Erb, E. M. Hoffmann, T. Dietlein, and N. Pfeiffer, "The diagnosis and treatment of glaucoma," *Deutsches Ärzteblatt International*, vol. 117, no. 13, p. 225, 2020.

[15] K. Allison, D. Patel, and O. Alabi, "Epidemiology of glaucoma: the past, present, and predictions for the future," *Cureus*, vol. 12, no. 11, 2020.

[16] Y.-C. Tham, X. Li, T. Y. Wong, H. A. Quigley, T. Aung, and C.-Y. Cheng, "Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis," *Ophthalmology*, vol. 121, no. 11, pp. 2081–2090, 2014.

[17] S. Shan, J. Wu, J. Cao, Y. Feng, J. Zhou, Z. Luo, P. Song, I. Rudan, G. H. E. R. Group *et al.*, "Global incidence and risk factors for glaucoma: A systematic review and meta-analysis of prospective studies," *Journal of global health*, vol. 14, p. 04252, 2024.

[18] R. N. Weinreb, T. Aung, and F. A. Medeiros, "The pathophysiology and treatment of glaucoma: a review," *Jama*, vol. 311, no. 18, pp. 1901–1911, 2014.

[19] K.-H. Yu, A. L. Beam, and I. S. Kohane, "Artificial intelligence in healthcare," *Nature biomedical engineering*, vol. 2, no. 10, pp. 719–731, 2018.

[20] P. Rajpurkar, J. Irvin, R. L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. P. Langlotz *et al.*, "Deep learning for chest radiograph diagnosis: A retrospective comparison of the chexnext algorithm to practicing radiologists," *PLoS medicine*, vol. 15, no. 11, p. e1002686, 2018.

[21] K. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *Computerized medical imaging and graphics*, vol. 31, no. 4-5, pp. 198–211, 2007.

[22] J. He, S. L. Baxter, J. Xu, J. Xu, X. Zhou, and K. Zhang, "The practical implementation of artificial intelligence technologies in medicine," *Nature medicine*, vol. 25, no. 1, pp. 30–36, 2019.

[23] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[24] A. M. Rahmani, E. Yousefpoor, M. S. Yousefpoor, Z. Mehmood, A. Haider, M. Hosseinzadeh, and R. Ali Naqvi, "Machine learning (ml) in medicine: review, applications, and challenges," *Mathematics*, vol. 9, no. 22, p. 2970, 2021.

[25] E. Y. Boateng, D. A. Abaye *et al.*, "A review of the logistic regression model with emphasis on medical research," *Journal of data analysis and information processing*, vol. 7, no. 04, p. 190, 2019.

[26] [Online]. Available: https://www.ibm.com/think/topics/decision-trees

[27] Z. Zhang, "Introduction to machine learning: k-nearest neighbors," *Annals of translational medicine*, vol. 4, no. 11, p. 218, 2016.

[28] M. Sbihi and N. Couellan, "Robust svm optimization in banach spaces," *arXiv preprint arXiv:2202.08567*, 2022.

[29] [Online]. Available: https://aws.amazon.com/fr/what-is/deep-learning/

[30] [Online]. Available: https://www.ibm.com/think/topics/neural-networks

[31] J. Gao, Y. Yang, P. Lin, and D. S. Park, "Computer vision in healthcare applications," *Journal of healthcare engineering*, vol. 2018, p. 5157020, 2018.

[32] R. Szeliski, *Computer vision: algorithms and applications*. Springer Nature, 2022.

[33] [Online]. Available: https://binariks.com/blog/computer-vision-in-healthcare/

[34] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, vol. 19, no. 1, pp. 221–248, 2017.

[35] StudySmarter. (s.d.) Analyse de l'image. [Online]. Available: https://www.studysmarter.fr/resumes/etudes-de-communication/communication-et-technologie/analyse-de-limage/

[36] A. Belsare and M. Mushrif, "Histopathological image analysis using image processing techniques: An overview," *Signal & Image Processing*, vol. 3, no. 4, p. 23, 2012.

[37] M. Turab, "A comprehensive survey on image signal processing approaches for low-illumination image enhancement," *arXiv preprint arXiv:2502.05995*, 2025.

[38] B. Fernando, E. Fromont, and T. Tuytelaars, "Mining mid-level features for image classification," *International Journal of Computer Vision*, vol. 108, pp. 186–203, 2014.

[39] A. Cavallaro and T. Ebrahimi, "Interaction between high-level and low-level image analysis for semantic video object extraction," *EURASIP Journal on Advances in Signal Processing*, vol. 2004, pp. 1–12, 2004.

[40] L. Cadena, A. Zotin, F. Cadena, A. Korneeva, A. Legalov, and B. Morales, "Noise reduction techniques for processing of medical images," in *Proceedings of the World Congress on Engineering*, vol. 1, 2017, pp. 5–9.

[41] S.-C. Pei and C.-N. Lin, "Image normalization for pattern recognition," *Image and Vision computing*, vol. 13, no. 10, pp. 711–723, 1995.

[42] E. Goceri, "Medical image data augmentation: techniques, comparisons and interpretations," *Artificial Intelligence Review*, vol. 56, no. 11, pp. 12 561–12 605, 2023.

[43] D. D. Patil and S. G. Deore, "Medical image segmentation: a review," *International Journal of Computer Science and Mobile Computing*, vol. 2, no. 1, pp. 22–27, 2013.

[44] [Online]. Available: https://www.geeksforgeeks.org/thresholding-based-image-segmentation/

[45] [Online]. Available: https://domino.ai/data-science-dictionary/feature-extraction

[46] A. W. Salehi, S. Khan, G. Gupta, B. I. Alabduallah, A. Almjally, H. Alsolai, T. Siddiqui, and A. Mellit, "A study of cnn and transfer learning in medical imaging: Advantages, challenges, future scope," *Sustainability*, vol. 15, no. 7, p. 5930, 2023.

[47] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[48] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 10 012–10 022.

[49] X. Li, H. Ding, H. Yuan, W. Zhang, J. Pang, G. Cheng, K. Chen, Z. Liu, and C. C. Loy, "Transformer-based visual segmentation: A survey," *IEEE transactions on pattern analysis and machine intelligence*, 2024.

[50] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images," in *International MICCAI brainlesion workshop*. Springer, 2021, pp. 272–284.

[51] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.

[52] S. He, R. Bao, P. E. Grant, and Y. Ou, "U-netmer: U-net meets transformer for medical image segmentation," *arXiv preprint arXiv:2304.01401*, 2023.

[53] C. Wang and W. Li, "A hybrid cnn-transformer architecture with frequency domain contrastive learning for image deraining," *arXiv preprint arXiv:2308.03340*, 2023.

[54] Y. Xie, J. Zhang, C. Shen, and Y. Xia, "Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24*. Springer, 2021, pp. 171–180.

[55] M. Heidari, A. Kazerouni, M. Soltany, R. Azad, E. K. Aghdam, J. Cohen-Adad, and D. Merhof, "Hiformer: Hierarchical multi-scale representations using transformers for medical image segmentation," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2023, pp. 6202–6212.

[56] L. Wang, R. Li, C. Zhang, S. Fang, C. Duan, X. Meng, and P. M. Atkinson, "Unetformer: A unet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 190, pp. 196–214, 2022.

[57] M. T. Islam, S. T. Mashfu, A. Faisal, S. C. Siam, I. T. Naheen, and R. Khan, "Deep learning-based glaucoma detection with cropped optic cup and disc and blood vessel segmentation," *IEEE Access*, vol. 10, pp. 2828–2841, 2022.

[58] M. Naveed, A. Ramzan, and M. U. Akram, "Clinical and technical perspective of glaucoma detection using oct and fundus images: a review," in *2017 1st international conference on next generation computing applications (NextComp)*. IEEE, 2017, pp. 157–162.

[59] P. Shanmugam, J. Raja, and R. Pitchai, "An automatic recognition of glaucoma in fundus images using deep learning and random forest classifier," *Applied Soft Computing*, vol. 109, p. 107512, 2021.

[60] M. K. Sahu, "A review on glaucoma: causes, symptoms, pathogenesis & treatment," *Journal of Clinical Research and Ophthalmology*, vol. 11, no. 1, pp. 001–004, 2024.

[61] D. C. Hood, S. La Bruna, E. Tsamis, K. A. Thakoor, A. Rai, A. Leshno, C. G. de Moraes, G. A. Cioffi, and J. M. Liebmann, "Detecting glaucoma with only oct: Implications for the clinic, research, screening, and ai development," *Progress in Retinal and Eye Research*, vol. 90, p. 101052, 2022.

[62] Á. S. Hervella, J. Rouco, J. Novo, and M. Ortega, "End-to-end multi-task learning for simultaneous optic disc and cup segmentation and glaucoma classification in eye fundus images," *Applied Soft Computing*, vol. 116, p. 108347, 2022.

[63] Y. Jin, L. Liang, J. Li, K. Xu, W. Zhou, and Y. Li, "Artificial intelligence and glaucoma: a lucid and comprehensive review," *Frontiers in Medicine*, vol. 11, p. 1423813, 2024.

[64] C. A. Girkin, G. McGwin Jr, M. J. Sinai, G. C. Sekhar, M. Fingeret, G. Wollstein, R. Varma, D. Greenfield, J. Liebmann, M. Araie *et al.*, "Variation in optic nerve and macular structure with age and race with spectral-domain optical coherence tomography," *Ophthalmology*, vol. 118, no. 12, pp. 2403–2408, 2011.

[65] C. Li, J. Chua, F. Schwarzhans, R. Husain, M. J. Girard, S. Majithia, Y.-C. Tham, C.-Y. Cheng, T. Aung, G. Fischer *et al.*, "Assessing the external validity of machine learning-based detection of glaucoma," *Scientific reports*, vol. 13, no. 1, p. 558, 2023.

[66] J. Wu, H. Fang, F. Li, H. Fu, F. Lin, J. Li, Y. Huang, Q. Yu, S. Song, X. Xu *et al.*, "Gamma challenge: glaucoma grading from multi-modality images," *Medical Image Analysis*, vol. 90, p. 102938, 2023.

[67] T. J. Jun, Y. Eom, D. Kim, C. Kim, J.-H. Park, H. M. Nguyen, Y.-H. Kim, and D. Kim, "Trk-cnn: transferable ranking-cnn for image classification of glaucoma, glaucoma suspect, and normal eyes," *Expert Systems with Applications*, vol. 182, p. 115211, 2021.

[68] L. Wang, J. Gu, Y. Chen, Y. Liang, W. Zhang, J. Pu, and H. Chen, "Automated segmentation of the optic disc from fundus images using an asymmetric deep learning network," *Pattern recognition*, vol. 112, p. 107810, 2021.

[69] R. Kashyap, R. Nair, S. M. P. Gangadharan, M. Botto-Tobar, S. Farooq, and A. Rizwan, "Glaucoma detection and classification using improved u-net deep learning model," in *Healthcare*, vol. 10, no. 12. MDPI, 2022, p. 2497.

[70] N. Thakur and M. Juneja, "Classification of glaucoma using hybrid features with machine learning approaches," *Biomedical Signal Processing and Control*, vol. 62, p. 102137, 2020.

[71] S. Dasgupta, R. Mukherjee, K. Dutta, and A. Sen, "Deep learning based framework for automatic diagnosis of glaucoma based on analysis of focal notching in the optic nerve head," *arXiv preprint arXiv:2112.05748*, 2021.

[72] V. P. C. Magboo and M. S. A. Magboo, "Batch size selection in convolutional neural networks for glaucoma classification," *arXiv preprint arXiv:2112.05748*, 2021.

[73] H. Yang, Y. Ahn, S. Askaruly, J. S. You, S. W. Kim, and W. Jung, "Deep learning-based glaucoma screening using regional rnfl thickness in fundus photography," *Diagnostics*, vol. 12, no. 11, p. 2894, 2022.

[74] U. Raghavendra, H. Fujita, S. V. Bhandary, A. Gudigar, J. H. Tan, and U. R. Acharya, "Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images," *Information Sciences*, vol. 441, pp. 41–49, 2018.

[75] P. Musa, F. A. Rafi, and M. Lamsani, "A review: Contrast-limited adaptive histogram equalization (clahe) methods to help the application of face recognition," in *2018 Third International Conference on Informatics and Computing (ICIC)*, 2018, pp. 1–6.

[76] A. Buades, B. Coll, and J.-M. Morel, "Non-local means denoising," *Image Processing On Line*, vol. 1, pp. 208–212, 2011.

[77] G. Deng, "A generalized unsharp masking algorithm," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1249–1261, 2011.

[78] [Online]. Available: https://paperswithcode.com/dataset/chase-db1

[79] L. Armi and S. Fekri-Ershad, "Texture image analysis and texture classification methods-a review," *arXiv preprint arXiv:1904.06554*, 2019.

[80] [Online]. Available: https://www.python.org/doc/essays/blurb/

[81] [Online]. Available: https://datascientest.com/en/kaggle-everything-you-need-to-know-about-this-platform

[82] T. Kahlert and K. Giza, "Visual studio code tips & tricks vol. 1," *Microsoft Deutschland GmbH*, 2016.

[83] M. Heydarian, T. E. Doyle, and R. Samavi, "Mlcm: Multi-label confusion matrix," *Ieee Access*, vol. 10, pp. 19 083–19 095, 2022.

[84] R. Zhao, W. Liao, B. Zou, Z. Chen, and S. Li, "Weakly-supervised simultaneous evidence identification and segmentation for automated glaucoma diagnosis," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 809–816.