Glaucoma Detection Using Macular OCT Images Based on Deep Convolutional Neural Networks

By

Hana Mekonen Tamiru

In Partial Fulfillment of the Requirements for the Degree of Master of Science in Biomedical Engineering (Bioinstrumentation and Imaging)

Center of Biomedical Engineering

Addis Ababa Institute of Technology

Addis Ababa University

Advisor: Dr. Dawit Assefa Haile

Clinical Advisor: Dr. Tesfaye Tadesse

Addis Ababa, Ethiopia
March , 2023

DECLARATION

I hereby declare that this MSc thesis is submitted in partial fulfillment of the Master's Degree is my original work and that all contributions from any other sources are properly and duly cited. I further declare that the material has not been submitted either in whole or in part, for a degree at this or any other university.

Name: Hana Mekonen Tamiru

Signature: ________________

Date: ____________________

This MSc thesis has been submitted for examination with my approval as an advisor.

Advisor: Dawit Assefa Haile (PhD) Signature _______________Date ______________

Clinical Advisor: Dr. Tesfaye Tadesse Signature _______________Date ______________

Certificate of Examination
Addis Ababa University
School of Graduate Studies

This is to certify that the thesis prepared by Hana Mekonen Tamiru entitled: "Glaucoma Detection Using Macula SD-OCT B-scans Based on Deep Convolutional Neural Networks" submitted in partial fulfilment of the requirements for the Degree of Master of Science in Biomedical Engineering (Bioinstrumentation and Imaging) complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the examining committee

Name of External Examiner	Signature	Date
Name of Internal Examiner	Signature	Date
Name of Advisor	Signature	Date

Chief of Department or Graduate Program Coordinator

ABSTRACT

Glaucoma is a major public health problem as it is the second leading cause of blindness after cataract. Since vision loss due to glaucoma can't be recovered, an early, reliable diagnosis is desirable. Although complete eye examination is recommended for assessment of both structural and functional states of the disease, glaucomatous structural changes precede functional changes. For instance, many studies reported that 25-30% of ganglion cell loss precedes the manifestation of visual field defect and loss of retinal nerve fiber layers (RNFL) occurs approximately six years before any detectable visual field defect. Therefore, the early diagnosis of glaucoma relies on the detection of these structural changes. Recently, classifying glaucomatous images taken from different modalities based on Deep Learning (DL) is increasingly being studied. Most of the researchers, however, relied on images generated from a fundus camera and others on OCT scans taken from the optic nerve head (ONH). Various others relied on specific information derived from the OCT machine itself including thickness and deviation maps of macular and ONH scans, and en-face images. However, the glaucomatous eye can be more effectively detected by analyzing the degeneration of the ganglion cell complex (GCCs) by using original OCT complete scans as input. The current thesis study used deep segmentation models to extract the GCC region which is composed of the retinal nerve fiber layer and ganglion cells with the inner plexiform layer. The study also used Convolutional Neural Network (CNN) based classifiers for detecting glaucomatous pathologies by paying attention to the GCC region of the macula Spectral Domain Optical Coherence Tomograpgy (SD-OCT) scans. The data set utilized for training and validation of the models composed of 1,262 locally acquired macula SD-OCT B-scans (431 non-glaucomatous and 830 glaucomatous) from four different regions of the macula: superior outside, inferior outside, inferior inside and central macula regions. Transfer learning was employed for segmentation as well as classifying the dataset. Deep segmentation models, SegNet, PSPNet, and $\mathrm{RAG-Net}_{\mathrm{v2}}$, were employed for segmentation and CNN models namely VGG16, VGG19, and ResNet50 were used for classification purpose. SegNet showed the best performance for retinal layer segmentation with 97.89% accuracy, 87% recall, 87.5% f1-score, 88% precision, 89% mean dice coefficient, and 81% mean_IOU. In terms of classification of glaucomatous and normal images, the best accuracy of 94.3% was obtained using VGG16 computed on the superior outside macula region, with 93.3% precision, 91.7% recall, 91.8% f1-score and 91.7% AUC. The study has demonstrated that using GCC aware deep learning model based on macula B-scans show great promises in accurate screening of glaucoma and suggested that incorporating DL technology into macula SD-OCT for glaucoma assessment has the potential to fill some gaps in current practices and clinical workflow.

Keywords: Glaucoma, Optical Coherence Tomography, Deep Learning, Retinal Nerve Fiber Layers, Ganglion Cell Complex

ACKNOWLEDGMENTS

First and foremost, praise and thanks to God, the Almighty, for his help throughout my research work.

Next, I would like to express my deep and sincere gratitude to my research advisor, Dr. Dawit Assefa Haile, and clinical Advisor Dr. Tesfaye Tadesse for their time, constructive and noticeable comments, assistance, and guidance throughout this study. My special thanks also go to all Ophthalmology Department Staff at Myungsung Christian Medical Center (Korean Hospital) in Addis Ababa. Special thanks to Mr. Biruk Tekle for his unwavering help during study data collection.

I am extremely grateful to my family for their love, prayers, care, understanding and sacrifices throughout the research work. This accomplishment would not have been possible without them. Finally, my appreciation goes to all the people who supported me during the process of studying.

Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

ADAM	Adaptive Moment Estimation
AI	Artificial Intelligence
ANN	Artificial Neural Network
CAD	Computer-aided Diagnosis
CNN	Convolutional Neural Network
DL	Deep Learning
FC	Fully Connected Layer
FN	False Negative
FP	False Positive
GCC	Ganglion Cell Complex
GCIPL	Ganglion Cell Inner Plexiform Layer
IOP	Intraocular Pressure
IOU	Intersection over Union
ML	Machine Learning
PAS	Peripheral Anterior Synechiae
POAG	Primary Open-angle Glaucoma
ReLU	Rectified Linear Units
RNFL	Retinal Nerve Fiber Layers
SD-OCT	Spectral Domain - Optical Coherence Tomography
SGDM	Stochastic Gradient Descent
TN	True Negative
TP	True Positive
UWF	Ultrawide Field
WHO	World Health Organization

CHAPTER ONE

INTRODUCTION

1.1 Background

Glaucoma is a global public health problem that is the leading cause of irreversible blindness worldwide [1][2]. It represents a group of eye diseases which result in damage to the optic nerve (or retina) and cause vision loss. The damage is often caused by an abnormal increase of pressure inside the eye which usually happens because of the build-up of a fluid known as aqueous humor in the eye [3]. When this fluid gets blocked or obstructed, the intraocular pressure increases. The increased intraocular pressure can damage the optic nerve which in time can weaken the eyesight and eventually lead to complete loss of vision [1].

Glaucoma has two major types: primary and secondary glaucoma. According to the underlying anatomy and pathophysiology, both of these glaucoma types have two major subtypes: open-angle glaucoma and angle-closure glaucoma [3]. Open-angle glaucoma has several variants and is a long duration (chronic) condition whereas the latter may be either a sudden (acute) condition or chronic. Primary open-angle glaucoma (POAG) takes 74.0% of all cases worldwide according to a previous study [4]. It is estimated that 57.5 million people worldwide are affected by POAG [5]. Most researches indicate that the prevalence of POAG is analyzed according to various risk factors. The risk factors include people over 60 years of age, family members of those already diagnosed with glaucoma, steroid users, diabetics, and also those with high myopia, hypertension, central cornea thickness of <5 mm, and eye injury.

Based on a recent review estimation carried out in 2020, approximately 76 million people suffer from glaucoma with that number estimated to reach 111.8 million by 2040 [6]. Africa is found to have the second highest prevalence of glaucoma (13%) among all continents next to Asia which accounted for approximately 60% of the world’s total glaucoma cases [6][7]. In Ethiopia, glaucoma accounted for 18.8% of blindness next to cataracts, a relatively larger proportion of blindness than seen in other regions of the world, based on a study reported in 2021 [8].

To diagnose glaucoma, a complete eye examination is recommended. The examination mainly includes measurement of Intraocular pressure (IOP), assessment of visual fields, and use of imaging techniques to examine the inner details of the eye [9]. These tests enable ophthalmologists to detect any changes and abnormalities in both function and structure of the eye caused by glaucoma. Although both structural and functional tests are essential for detecting and monitoring glaucomatous damage, many studies reported that glaucomatous structural changes precede functional changes [10]. Currently, the visual field test is the most common test for the evaluation of glaucoma functional defects. However, visual field loss is not always apparent until the advanced stage of the disease. On the other hand, structural changes caused by glaucoma are mainly characterized by optic nerve head remodeling, loss of optic nerve axons, and loss of retinal ganglion cells. Studies showed that at least 25-30% of ganglion cells are lost before visual field defect manifests and retinal nerve fiber layers (RNFL) loss occurs approximately six years prior to any detectable visual field defect [11][12][10]. Therefore, the early diagnosis of glaucoma relies on the detection of structural changes.

Fundoscopy and Optical Coherence Tomography (OCT) are the most common imaging modalities used to analyze the internal structures of the eye [13]. Fundus imaging is the easiest test in clinics and can identify glaucomatous structural change [14]. However, it offers very limited information about the characterization of glaucoma and does not provide quantitative assessments of the retinal layers. Whereas, OCT enables high resolution cross-sectional images of the retina and optic nerve in vivo along with objective, reproducible and quantitative assessments of various structures [15]. Recently, macular OCT imaging of the ganglion cells and their surrounding layers has shown promise in early diagnosis and management of glaucoma where other changes may not be apparent. Since defects in retinal nerve fiber layers and ganglion cells are early signs of glaucoma before other changes are visible, a cross-sectional image of macular OCT image becomes preferable for early glaucoma diagnosis [16].

Indeed, the use of computers has become increasingly relevant to medical decision-making. Computer-aided diagnosis (CAD) of diseases introduces a great impact on medical diagnosis. Structural imaging modalities such as fundus image and OCT have been investigated for automatically estimating key parameters for glaucoma assessment including cup-to-disc diameter ratio, the thickness of the ganglion cell layer, and the thickness of the RNFL. An emerging area of diagnosis of diseases involves the use of automated interpretation of clinical data and digital images with the help of artificial intelligence (AI) [17][18].

Based on Oxford Dictionaries' definition, AI implies the theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages. AI methods have recently demonstrated the significant advances in medicine and the potential to revolutionize the screening, diagnosis, and classification of glaucoma. As it holds promise for fundamentally changing research aimed at understanding the development, progression, and treatment of glaucoma, applying AI based glaucoma detection is crucial [19][20]. Deep learning (DL) is a type of AI that enables automatic learning through the absorption of data such as images, video, or text. It is a subset of machine learning, which is essentially a neural network with three or more layers. DL is at the forefront of researches using fundus photography and OCT for glaucoma detection. This study also used the DL algorithm for early diagnosis and detection of glaucoma based on macular information taken from 2D B-scans of Spectral Domain OCT (SD-OCT) [21].

1.2 Problem Statement

Glaucoma is a major public health problem as it is the second leading cause of blindness after cataract [2][5]. Since vision loss due to glaucoma can't be recovered, an early reliable diagnosis and treatment are obviously desirable. Although early glaucoma recognition slows down and prevents vision loss, clinical practice has shown that identification of glaucoma at its earlier stage is still challenging. Since it has different types and stages, the signs and symptoms of glaucoma vary depending on the type and stage of patient conditions [9]. Plus, because glaucoma is a slowly progressing disease and its time course is often variable, diagnosis of glaucoma is still challenging.

Currently, retinal images from OCT and Fundus photography enabled earlier detection of glaucoma progression than ever before and are used widely by ophthalmologists [22]. However, early glaucoma diagnosis is still an issue. One of the biggest challenges when trying to diagnose glaucoma in its earliest stages is the significant variation in the appearance of the optic disc and peripapillary regions found in normal eyes. Some abnormal disc and retinal nerve fiber layers can be very difficult or impossible to distinguish from glaucomatous discs and layers. These characteristics resist the ophthalmologist's ability to accurately distinguish between glaucomatous and normal eyes. As it gives the quantitative analysis of the images, OCT has become more preferable tool for diagnosis of glaucoma. However, most clinically applicable OCT machines use basic image processing tools to analyze and segment OCT images. This is time-intensive and frequently error-prone task. For example, in glaucoma, it has been estimated that 36% of OCT scans could be subject to artifacts that affect estimations of the RNFL thickness [23]. The greater chances of misdiagnosis by the currently used methods for diagnosis open up ways for more advanced techniques like the use of DL.

Many researchers proposed their work on DL applied on images captured through a given imaging modality for glaucoma diagnosis. Most of the researchers, however, relied on images generated from a fundus camera and others on OCT scans taken from the optic nerve head (ONH). Various others relied on specific information derived from the OCT machine itself including thickness and deviation maps of macular and ONH scans, and en-face images. These methods, however, are meant to be effective for specific OCT machine designs. This calls for the development of an automated system that could take original OCT complete scans as input and process the images for use in a more effective detection of Glaucoma. Furthermore, recent researches proved that glaucomatous damage to the macular area happens earlier [22][12][24] and detection of it is unquestionably vital to detect early manifestations of the disease. Therefore, the current study attempted to investigate the efficacy of a DL-based optimum glaucoma diagnosis approach by using a macular SD-OCT B scan.

1.3 Research Questions

The research on the detection of glaucoma using macula SD-OCT B-scans and a deep convolutional neural network tries to answer the following research questions:

1. How can we use macula OCT B-scans for glaucoma detection?
2. What macular region is more sensitive for glaucoma detection?
3. What is the best set of hyper-parameters that can be used to train deep segmentation and classification models for extracting relevant features of macular SD-OCT scans for early detection of Glaucoma?
4. Which deep segmentation model can give the best performance in the segmentation of retinal OCT images?
5. Which CNN classification model can give the best performance in the early detection of glaucoma using macular SD-OCT scans?

1.4 Objectives of the Research

General Objective

- The general objective of this research is to find the most suitable deep segmentation and classification models for the early detection of glaucoma using macular SD-OCT images.

Specific Objectives

The specific objectives of this study are:

- To investigate the application of using macula SD-OCT B-scans for early detection of glaucoma.
- To find the most sensitive macula region for glaucoma detection.
- To train and evaluate the performance of the deep learning system for segmentation and early detection of glaucoma using macular SD-OCT.
- To find out the best set of hyper-parameters for training each model.
- To compare and choose the best performed segmentation and classification models for early detection of glaucoma.

1.5 Scope of the Study

The number of hospitals having OCT is very limited and most of them use optic nerve head (ONH) scans for glaucoma diagnosis, specifically in the Ethiopian context where the current research focuses on. Hence, the study used OCT data collected from only one hospital. It might be interesting to collect more data from different sites (hospitals) and check the performance of the proposed glaucoma detection tool. This might improve the overall performance of the approach and make it more general. Moreover, the study only detects the glaucomatous pathology without further grading them. Although different modalities are available for imaging purposes, the scope of the current research work is focused on automatic early glaucoma detection based on OCT images only. The methodology developed has been implemented on Matlab 2020a and Python 3.8.8 platforms.

1.6 Organization of the Thesis

The rest of the thesis has been organized into five chapters. Chapter 2 of the research flagged out key theoretical frameworks including the anatomy of the eye and basic concepts behind glaucoma, glaucoma diagnosis modalities and machine learning. A thorough literature review is included in Chapter 3. Chapter 4 provides information about the study population and methods used for segmentation and classification. Results and discussions are presented in Chapter 5. The last chapter concludes the thesis by presenting the study's limitations and giving some useful recommendations and possible future directions.

CHAPTER TWO

THEORETICAL BACKGROUND

2.1 Structure of the Eye and the Retina: Overview

The Eye is a specialized sense that provides living organisms with vision, the ability to receive and process visual details, as well as enabling several photo response functions that are independent of vision [25]. The human eye is part of the sensory nervous system that reacts to visible light and allows us to use visual information for a variety of purposes, including seeing things, keeping balance and maintaining circadian rhythms. Our eye collects light from the surrounding environment, regulates its intensity through a diaphragm, and focuses it through an adjustable assembly of lenses to form an image. Then it converts this image into a set of electrical signals and transmits these signals to the brain through complex neural pathways that connect the eye via the optic nerve to the visual cortex and other areas of the brain [26]. A schematic diagram showing the structure of the human eye is presented in Fig. 1.

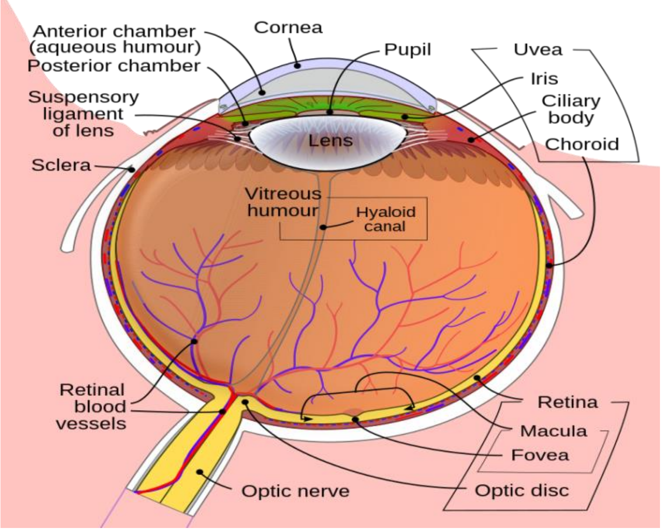


Figure 1: Schematic diagram of the human eye [27].

The eye consists of three coats, or layers, surrounding different anatomical structures. The outermost layer, called the fibrous tunic, consists of the cornea and sclera which give shape to the eye and support the deeper structures. The sclera is the white outer layer of the eyeball that helps to maintain the spherical shape of the eyeball whereas the cornea is the transparent front part of the eye that covers the iris, pupil, and anterior chamber. The middle layer, known as the vascular tunic or uvea, is divided into two parts: anterior (iris and ciliary body) and posterior (choroid).

Choroid supplies blood vessels to the eye and contains dark pigment granules that prevent the reflection of light in the eye. The ciliary body consists of the ciliary muscle that controls the shape of the lens, and the ciliary epithelium that produces aqueous humor. Iris is the color portion of the eye formed by circularly and radially arranged smooth muscle fibers. It regulates the amount of light entering the eye by constricting or dilating the pupil. The innermost layer of the eye is the retina. [25][26].

The retina is a filmy piece of tissue that lines the inside of the eyeball. Retina includes both the sensory neurons that respond to light and intricate neural circuits that perform the first stages of image processing; ultimately, an electrical message travels down the optic nerve into the brain for further processing and visual perception [28][29]. Ten layers (see also Fig. 2) of cells in the retina can be seen microscopically including the inner limiting membrane (ILM), nerve fiber layer, ganglion cell layer, inner plexiform layer, inner nuclear layer, outer plexiform layer, outer nuclear layer, external limiting membrane, inner segment/outer segment layer, and retinal pigment epithelium [30].

Although ten layers of cells in the retina can be seen microscopically in general, there are four main layers. Next to the choroid is the pigment epithelium. Above the epithelium is the layer of rods and cones, the light-sensitive cells. The changes induced in the rods and cones by light are transmitted to a layer of neurons (nerve cells) called the bipolar cells. These bipolar cells connect with the innermost layer of neurons, the ganglion cells, and axons, which constitute the optic nerve fibers. The spaces separating the later three layers are also anatomically distinct. The region containing synapses linking the photoreceptors with bipolar and horizontal cell dendrites is known as the outer plexiform layer; the area where the bipolar and amacrine cells connect to the ganglion cells is the inner plexiform layer [30][31].

Retina has to change thickness in various sections. The central retina close to the fovea is considerably thicker than the peripheral retina. This is due to the increased packing density of photoreceptors, particularly the cones, and their associated bipolar and ganglion cells in the central retina compared with the peripheral retina [32][33].

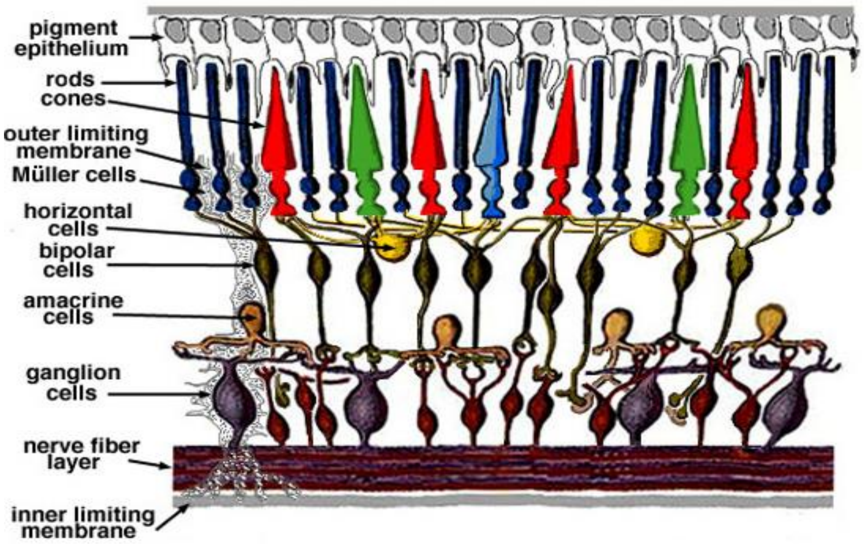


Figure 2: Schematic diagram of the human eye [34].

2.2 Glaucoma

The term glaucoma refers to a group of progressive optic neuropathies characterized by an excavated appearance of the optic disc, often described as cupped, together with loss of retinal ganglion cells and their axons and corresponding visual field loss [32][1]. The damage is often caused by an abnormal increase of pressure inside the eye which usually happens because of the build-up of a fluid known as aqueous humor in the eye. When this fluid gets blocked or obstructed, the intraocular pressure increases. The increased intraocular pressure can damage the optic nerve which in time can weaken the eyesight and eventually lead to complete loss of vision [3].

2.2.1 Classification of Glaucoma

Glaucoma is classified as primary or secondary and as open angle or angle closure. Primary glaucoma is glaucoma that develops due to an unknown cause, whereas secondary glaucoma is associated with known predisposing events including developmental abnormalities, systemic diseases, drug therapy, or trauma. We can take primary and secondary glaucoma as major types. Then, according to the underlying anatomy and pathophysiology, both of these have two major subtypes: open-angle glaucoma and angle-closure glaucoma. Open-angle glaucoma has several variants and is a long-duration (chronic) condition whereas angle closure glaucoma may be either a sudden (acute) condition or chronic [35].

In open-angle glaucoma (OAG), no obstruction of the trabecular meshwork is visible on gonioscopic examination of the anterior chamber angle. A previous study indicates that primary open-angle glaucoma takes 74.0% of all cases worldwide. It is estimated that 57.5 million people worldwide are affected by primary open-angle glaucoma (POAG) [36].

In angle closure, the peripheral iris partially or completely obstructs the trabecular meshwork. The obstruction may be caused either by appositional iridotrabecular contact or by adhesions, known as peripheral anterior synechiae (PAS), between the iris and the trabecular meshwork. When angle closure is present in association with glaucomatous optic neuropathy, it is known as angle-closure glaucoma [3][35].

2.2.2 Major Risk Factors for Glaucoma

Anyone can develop glaucoma, but several factors increase the risk. Major risk factors for glaucoma include [37][38][39][40]:

- Increased intraocular pressure (IOP): increased pressure inside the eye due to the obstructed flow of aqueous humor may lead to glaucoma.
- Increased age: risk for developing glaucoma increases a little with each year of age. People over age 60 are at increased risk for the disease. In one of the previous studies it was reported that, for African Americans, the increase in risk begins after age 40.
- Race: many studies implied that African-American or people of African descent are much more likely to get glaucoma than those who are Caucasian. POAG is three to four times more prevalent in African Americans than in Caucasian Americans, according to previous report. People of Asian descent have a higher risk of angle-closure glaucoma, and people of Japanese descent have a higher risk of developing low-tension glaucoma. Figure 3 presents estimated percentage prevalence of primary open-angle glaucoma with age for men and women combined by ethnicity.
- Family history of glaucoma: having a family history of glaucoma increases the risk of developing glaucoma.

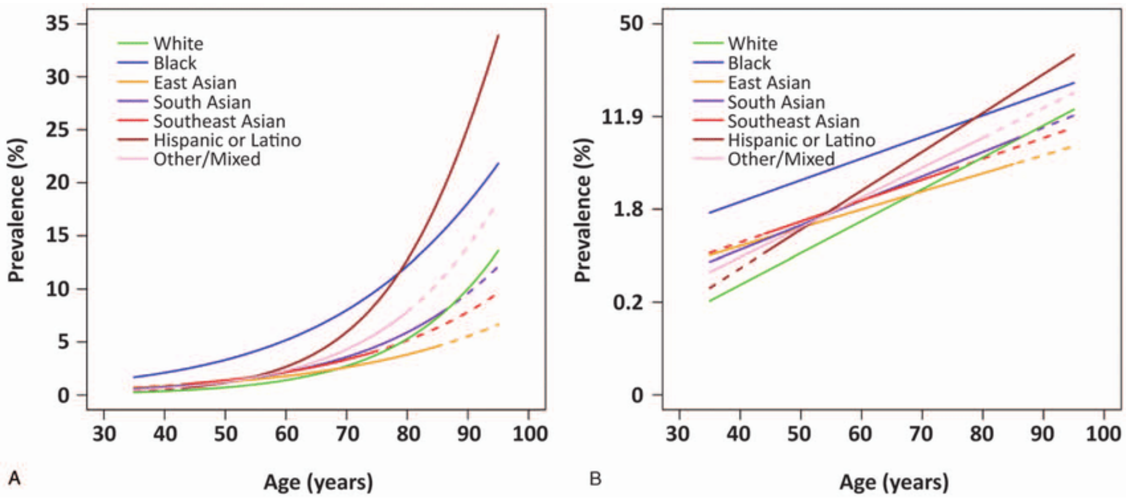


Figure 3: Estimated prevalence (%) of primary open-angle glaucoma with age for men and women combined by ethnicity; (A) prevalence on normal scale, (B) prevalence on log scale [40].

2.2.3 Glaucomatous Structural and Functional Changes

Histological studies suggest that a substantial number of retinal ganglion cells are lost before the development of visual field damage. A retinal ganglion cell (RGC) is a type of neuron located near the inner surface of the retina of the eye. These neurons receive signals from the photoreceptors, process them, and transmit them in axons through the optic nerve further to the brain. The axons run from the ganglion cell nuclei in the retina to the optic disc, and then together with the retinal vessels through the lamina cribrosa, a sieve-like structure composed of collagen [3] [41]. Behind the lamina cribrosa, the axons, surrounded by a myelin sheath, continue as the optic nerve.

Changes in lamina cribrosa morphology in glaucoma include posteriorization of the laminar insertion into the sclera, increased cupping or depth of the lamina cribrosa, and the development of focal lamina cribrosa defects. Elevated intraocular pressure, low perfusion pressure, and/or low cerebrospinal fluid pressure increase the gradient across the lamina cribrosa and cause papillary hypoperfusion, leading to structural changes and re-modeling of the lamina cribrosa and to impaired axonal transport in the optic nerve fibers [38][41]. In particular, the pores in the anterior region of the lamina cribrosa are elongated in open-angle glaucoma [39]. The increasing loss of retinal ganglion cells leads to progressive impairment of the visual field, generally beginning in the mid-periphery and then advancing until only a central or peripheral island of intact vision remains. Further functional disturbances include impaired contrast and color perception and difficulty in reading [9].

Glaucoma structural changes are generally characterized by morphological and structural changes in the ONH manifested by neuroretinal-rim thinning and increased cupping, associated with progressive loss of the retinal ganglion cells (RGC) and their axons in the RNFL. Its functional change is mainly demonstrated by visual field (VF) defects.

2.2.4 Diagnosis of Glaucoma

Appropriate management of glaucoma depends on the clinician's ability to diagnose the specific form of the disease in a given patient, determine the severity of the condition, predict the likelihood of progression, and detect progression when it occurs. The diagnosis begins with taking past ocular history and medical history as well as a review of available pertinent records.

Then the evaluation of patients with glaucoma and those at risk of developing glaucoma typically consists of determining IOP, structural properties (e.g. appearance of the ONH and thickness of the retinal nerve fiber layer and ganglion cell complex), and functional characteristics (e.g. visual field sensitivity) of the visual pathway. These tests enable ophthalmologists to detect any changes and abnormalities in both function and structure of the eye caused by glaucoma [1].

Currently, visual field test is the most common test for the evaluation of glaucoma functional defects. However, visual field loss is not always apparent until the advanced stage of the disease. Histological studies presented that a substantial number of retinal ganglion cells are lost before the development of visual field damage. Studies showed that at least 25-30% of ganglion cells are lost before visual field defect manifests and retinal nerve fiber layers (RNFL) loss occurs approximately six years prior to any detectable visual field defect. It has also been shown that in the earlier stages of glaucoma, optic disc changes are more prominent than visual field defects; but, in the later stages of the disease, visual field changes are more prominent compared with morphological changes. These imply that the early diagnosis of glaucoma relies on detecting of structural changes [3].

2.3 Retina Imaging Modalities for Glaucoma Diagnosis

Currently, different modalities are used for clinical retinal imaging, such as the fundus camera, scanning laser ophthalmoscope, and OCT. Fundus camera and OCT are the most common imaging modalities used for analyzing internal structures of the eye. Whereas, OCT enables high-resolution cross-sectional images of the retina and optic nerve in vivo along with objective, reproducible and quantitative assessments of various structures [13].

2.3.1 Fundus Camera

A fundus camera is a specialized low-power microscope with an attached camera designed for medical purposes. It provides an upright, magnified view of the fundus. The main structures that can be visualized in a fundus photo are the central and peripheral retina, optic disc, and macula. Two types of fundus cameras exist, the standard view and the ultrawide field (UWF). Standard fundus photography, a widely available modality, provides a 30 to 50-degree image that includes the macula and optic nerve. Whereas, UWF imaging can produce up to a 200-degree retinal view extending beyond the macula to the peripheral retina. This type of fundus photograph can image over 80% of the total surface retinal area [42].

Fundus imaging is the most widely applied technique for glaucoma detection [14]. It is the easiest test in the clinics and can identify glaucomatous structural change. However, it gives very limited information about characterization of glaucoma and does not provide quantitative assessments of the retinal layers.

2.3.2 Optical Coherence Tomography (OCT)

OCT is a non-invasive imaging technique that uses low-coherence light to capture micrometer-resolution, two- and three-dimensional images from within optical scattering media. The original time-domain OCT (TD-OCT) has been superseded by Fourier – Fourier-domain OCT (FD-OCT), which has improved image acquisition speed, resulting in better reproducibility [43]. One important type of FD-OCT is the so called spectral-domain OCT (SD-OCT), the type of OCT technique utilized in the current study.

OCT enables high-resolution cross-sectional images of the retina and optic nerve in vivo along with objective, reproducible and quantitative assessments of various structures. Structurally, glaucoma causes progressive loss of the RGC and their axons in the RNFL, and thinning of the neuroretinal rim in the optic nerve disc (OND). OCT provides quantitative measurements of the circular peripapillary retinal nerve fiber layer (cpRNFL) thickness, optic nerve head topography, and macular thickness, which can help discriminate glaucomatous eyes from healthy eyes [15][7].

Peri-papillary Retinal Nerve Fiber

OCT RNFL thickness measurements are usually acquired in the peripapillary area, that is, at a fixed radius around the ONH. Most commercially available OCT devices acquire RNFL thickness measurements in a peripapillary circle located at a certain distance from the ONH, usually

3.45 mm. RNFL thickness measurements are generally lower in glaucomatous eyes compared with those in nonglaucomatous eyes, although considerable interpersonal variability exists [44].

Measurement parameters presented in OCT reports include the global average peripapillary RNFL thickness, which corresponds to the average of all thickness measurements in the peripapillary circle, as well as average RNFL thickness in quadrants (superior, inferior, temporal, nasal) or in small clock- hour sectors. Among all parameters, the average peripapillary RNFL thickness and thicknesses in the inferior and superior quadrants have the best diagnostic accuracy [45].

Optic Nerve Head (ONH)

The optic nerve is composed of the axons of the RGCs, glial tissue, extracellular matrix, and blood vessels. The arcuate nerve fibers entering the superior and inferior poles of the optic disc are more susceptible to glaucomatous damage. An understanding of the normal and pathologic appearance of the optic nerve allows the clinician to detect and monitor glaucoma [45]. SD-OCT devices are also able to provide topographical measurements of the ONH, including measurements of the optic disc area, neuroretinal rim area, and cup– disc ratio [46].

Macula

Although analysis of cpRNFL and OND measurement using OCT is very useful for glaucoma assessment, macular measurement showed better glaucoma diagnostic performance and progression detection capability in some specific cases. For example, the cases of a high degree of myopia, and OND abnormalities (drusen, pit, papilledema, paripapillary changes, etc.,), suggest that macula measurements may have the superior capability to produce a better diagnostic strategy. In eyes with myopic discs or large areas of PPA (paripapillary area), which can present with artifacts on peripapillary RNFL assessment, macular evaluation may provide an enhanced ability to diagnose and monitor glaucomatous damage [45].

The macula includes the region surrounding the fovea with the highest density of RGCs. There is a growing body of evidence that early glaucomatous damage involves the macula. The anatomical basis of this damage can be studied using SD-OCT, by which the thickness of the RNFL and local retinal ganglion cell plus inner plexiform (RGC-IP) layer can be measured [24]. The macular RGC layer contains more than 50% of the RGCs of the entire retina. Because much of the total macular thickness consists of RNFL and ganglion cell bodies, this region is an attractive area for identifying structural glaucoma damage.

Moreover, SD-OCT also enables quantitative assessment of either the entire macular thickness or the thickness of specific layers that may be important in glaucoma. For thickness measurements of glaucoma evaluation, some OCT machines use the overall thickness of the ganglion cell complex (GCC), which is composed of the RNFL, the ganglion cell layer, and the inner plexiform layer and others uses only RNFL thickness. The thickness of the ganglion cell layer combined with the inner plexiform layer (GCIPL) is a commonly used parameter. Figure 4 presents an SD-OCT B-scan taken from the macula center at the fovea. The left panel shows the OCT en-face image while the right shows the B-scan taken around the central region.

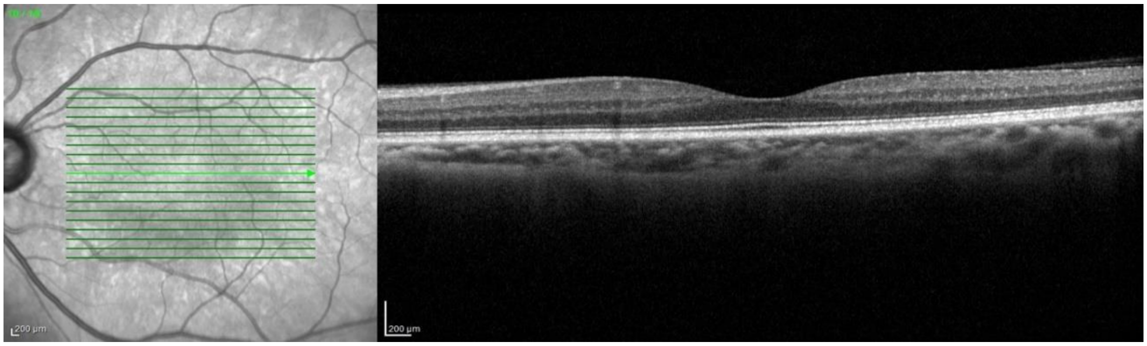


Figure 4: Macula SD-OCT B-scans taken from the macula center at the fovea.

2.4 Artificial Intelligence, Machine Learning, and Deep Learning

Artificial intelligence (AI) is a broad field of computer science that focuses on creating intelligent machines that can accomplish activities that would typically require human intelligence [47]. Although AI is an interdisciplinary science with multiple approaches, advancements in machine learning and deep learning are creating a paradigm shift in every sector of the tech industry. Machine learning is the part of AI that allows machines to learn from data it receives without explicitly being programmed. Contrary to classic, rule-based AI systems, machine learning algorithms develop their behavior by processing annotated examples, through the process called “training”. On the other hand, deep learning is a subdivision of machine learning that configures computers to perform tasks through experience (see also Fig. 5) [19].

Deep learning can solve more complex problems than other types of machine learning. Deep-learning algorithms solve problems using deep neural networks, a type of software architecture inspired by the human brain. It is an artificial neural network in which multiple layers of processing are used to extract progressively higher-level features from data. The neural networks are layers upon layers of variables that adjust themselves to the properties of the data they're trained on and become capable of doing tasks such as classifying images [48].

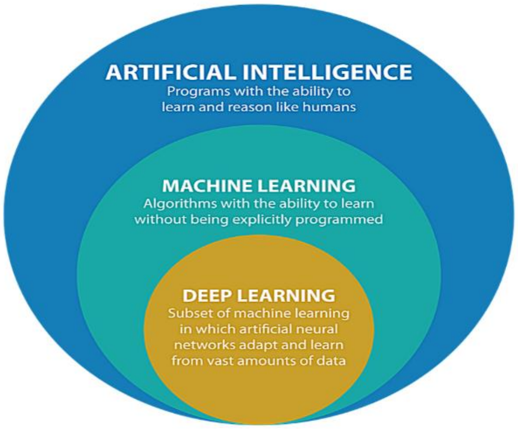


Figure 5: Difference between artificial intelligence, machine learning, and deep learning [49].

2.4.1 Convolutional Neural Network (CNN)

There are various types of neural nets, among those, convolutional neural networks (ConvNets or CNNs) are more often utilized neural nets used for classification and computer vision tasks especially for analyzing visual images. A CNN is a feed-forward neural network**,** which involves various layers and deeply connected architecture. CNN represents the input data in the form of multidimensional arrays. Then the various layers of the CNN extract every portion of the input image, which is known as receptive field, and finally learn to classify the images [50]. Based on the significant role of the receptive field, CNN assigns weights for each neuron. Hence, the outputs of a typical CNN represent the classes or the labels of the classes. They have three main types of layers, which are the Convolutional layer, Pooling layer, and Fully-connected (FC) layer as also been depicted in Fig. 6 [51].

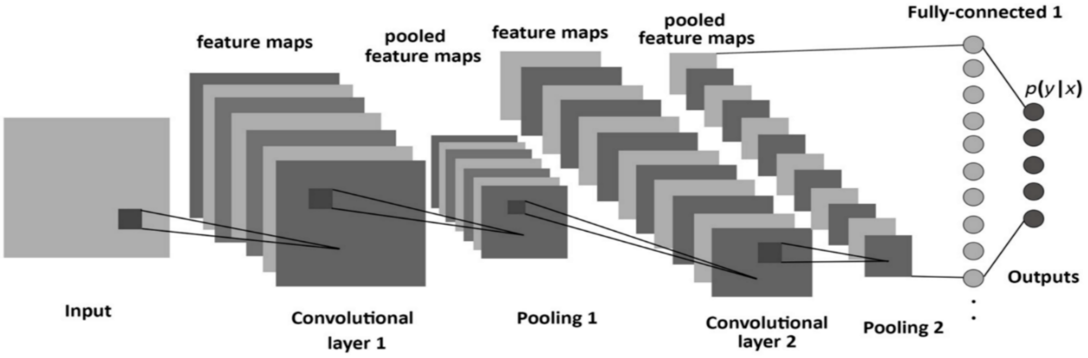


Figure 6: Diagram of Convolutional Neural Networks [52].

2.4.1.1 Convolutional Layer

The convolutional layer is the first layer and the core building block of a CNN. The convolutional layer applies a convolution operation to the input and passes the result to the next layer. This process converts all the pixels in its receptive field into a single value. It requires a few components, which are input data, a filter, and a feature map. The convolutional layer contains a set of filters whose parameters need to be learned. The filters (also known as kernels) use to detect what features, such as edges, are present throughout an image. The kernel filters are of the same dimension as the input images, but the filter size needs to be smaller compared to that of the input image with smaller constant parameters. Based on the type of tasks we need to do and the kind of features we are looking to learn, we can use different kinds of convolutions [51].

The 2D Convolution Layer

In this most common type of convolution, a filter or a kernel in a conv2D layer slides over the 2D input data, performing an element-wise multiplication. The filter mask slides across the width and height of the input image step by step and the dot products between the weights of the kernel filters with the value of the input image are computed at every spatial position. In this manner the kernel will perform the same operation for every location it slides over and transform a 2D matrix of features into a different 2D matrix of features. The convolution is given by Eqn. 1:

$$F * K\ (p) = \sum_{s+t=p} F(s)K(t) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

where $F(s)$ indicates the Input image and $K(t)$ the convolution filter. Figure 7 presents the convolution operator in action using a 3x3 size filter.

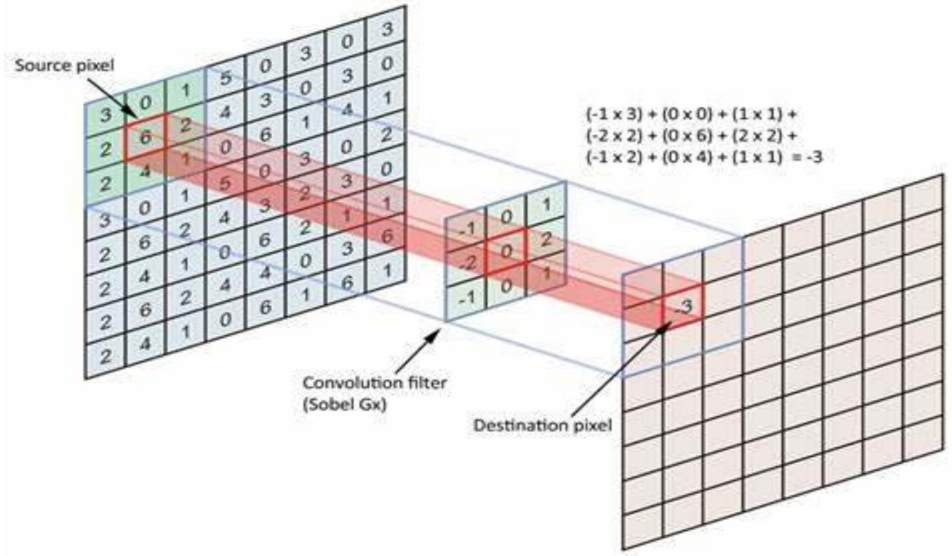


Figure 7: 2D convolution with filter size 3x3 [53].

The Dilated or Atrous Convolution

Dilated convolution was introduced in the paper "Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs" which appeared on the ICLR conference proceedings. The idea behind this convolution is to "inflate" the kernel which in turn skips some of the points [54]. Dilated operation expands window size without increasing the number of weights by inserting zero-values into convolution kernels. As a result, it operates with a less processing power. Indeed, as large-size kernels extract more useful information than the small-size kernel, the computation cost would be higher. Instead, dilated convolutions gather large-size area information with small-size kernels for higher dilation rates to keep dimensions the same as the input image, as also shown in Fig. 8**.** Since this convolution type support exponential expansion in the context of the receptive field, there is no loss of resolution [54].

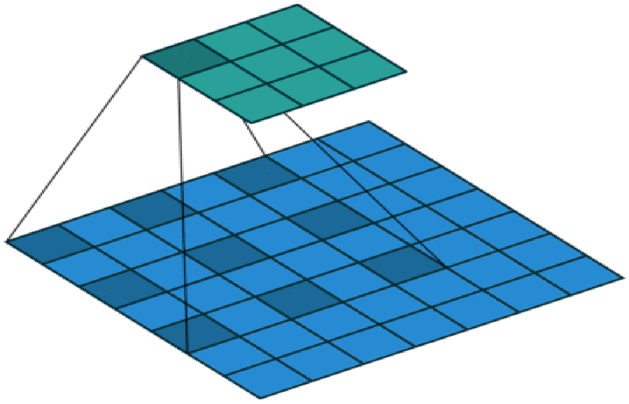


Figure 8: The Dilated or Atrous Convolution [55].

$$(F_l * K)(p) = \sum_{s+lt=p} F(s)K(t) \text{...(2)}$$

The l in the summation $s + lt = p$ tells us that some points are skipped during convolution. If l = 1, it is the same as a normal discrete convolution. But for l>1, the convolution becomes dilated convolution. The parameter l is known as the **dilation rate** which tells us how much we want to widen the kernel [54]. Figure 9 shows three different dilated convolutions: 1, 2 and 4 dilations.

2.4.1.2 Pooling Layer

The pooling layer, also called subsampling, is a layer of neural nodes in neural network that is used to reduce the dimensions of the feature maps. This is done by dividing the input feature set into many local neighbor areas. It pools one output value from each local neighbor area. Through dimensionality reduction, pooling layers reduces the number of parameters to learn and the amount of computation performed in the network.

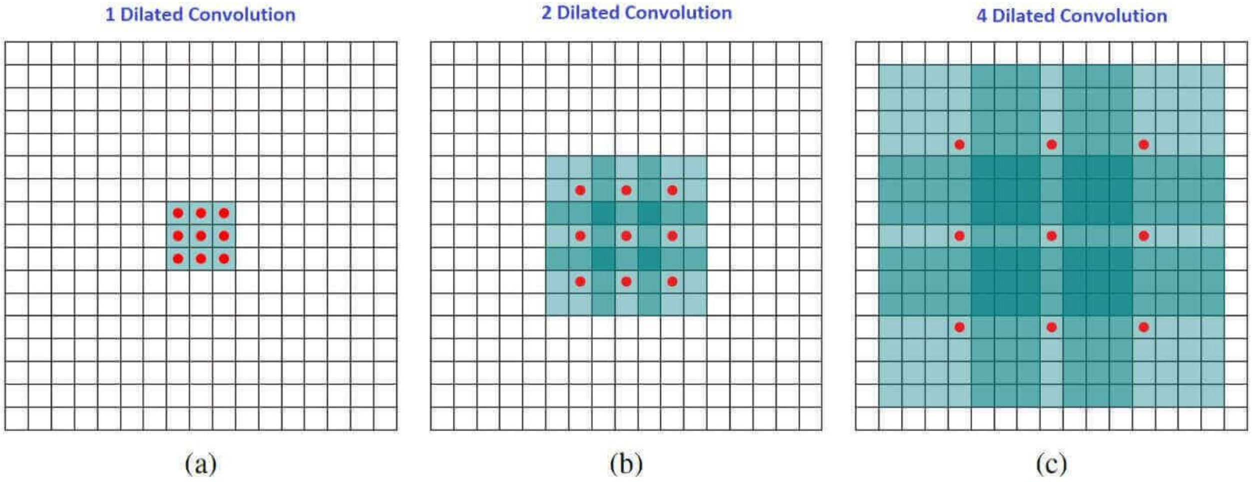


Figure 9: Three different dilated convolutions where l values 1, 2, and 4 respectively [54].

The pooling layer uses different filters to identify different parts of the images like edges, corners, body, feathers, etc. The most commonly used method in the CNN to pool an output value from a local neighbor area is to take the maximum and average input value of that area. Max pooling is a pooling operation that selects the maximum element from the region of the feature map covered by the filter whereas, average pooling computes the average of the elements present in the region of the feature map covered by the filter (see also Fig. 10) [56].

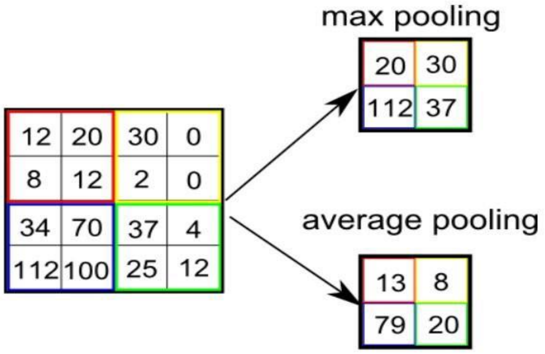


Figure 10: Types of pooling [57].

Then, the pooled feature map is going to be flattened. Flattening is the process of converting all the resultant 2-dimensional arrays from pooled feature maps into a single long continuous linear vector as also been depicted in Fig. 11. The flattened matrix from the pooling layer is fed as input to the fully connected layer to classify the image.

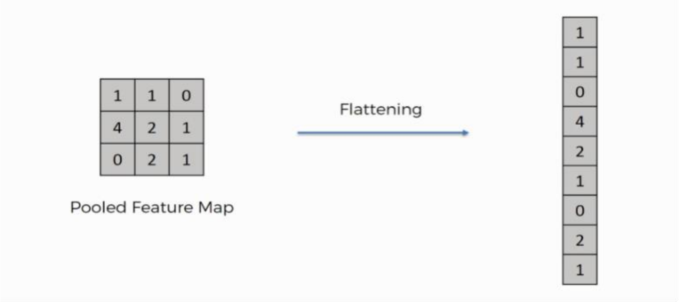


Figure 11: Example of flattening before feeding to the FC [58].

2.4.1.3 Fully Connected (FC) Layers

A fully connected layer refers to a neural network in which each neuron applies a linear transformation to the input vector through a weights matrix. Under the full connection process, the neuron in the fully-connected layer detects a certain feature, preserves its value, and communicates this value with each class. Then each class checks out the feature and decides whether it's relevant to them (see also Fig. 12).

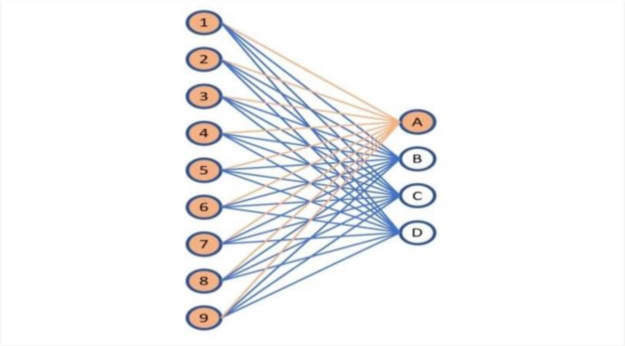


Figure 12: Illustration of a fully connected layer [59].

2.4.2 Parameters for CNNs

2.4.2.1 Activation Functions

An activation function is used to introduce non-linearity into the output of a neuron. Plus, it is used to learn and approximate any kind of continuous and complex relationship between variables of the network. This parameter decides whether a neuron should be activated or not by calculating the weighted sum and further adding bias to it. There are several commonly used activation functions such as the ReLU, Softmax, TanH, and Sigmoid functions. Although each of these functions has a specific usage, for a multi-class classification Softmax function is preferred [60][61]. On the other hand, ReLU is less computationally expensive and learns much faster than Sigmoid and TanH functions [62]. Therefore, in order to choose an activation function, we need to know the function characteristics we are trying to approximate.

2.4.2.2 Optimizers

Optimizers are algorithms or methods used to change the attributes of the neural network such as weights and learning rate in order to reduce the losses and to provide the most accurate results possible. There are different types of optimizers used in CNN implementation and some are described below.

Gradient Descent

Gradient descent is a first-order optimization algorithm that is dependent on the first-order derivative of a loss function [63]. When the cost function has been set to those neurons, gradient descent can be used to update the weight at each neuron. It calculates a minima way that the weights should be altered [64].

The mathematical form for gradient descent is given by:

$\theta = \theta - \alpha \cdot \nabla J(\theta)$……………………………….……………………………………...…….(3)

where θ is the weight, α is the learning rate, and $\nabla J(\theta)$ is the partial derivatives of the loss function with respect to the weight θ.

Gradient descent optimizer is easy to implement and understand. However, weights are changed after calculating the gradient on the whole dataset. So, if the dataset is too large, it may take a long time to converge to the minimum.

Stochastic Gradient Descent

Stochastic Gradient Descent is a variant of Gradient Descent. This optimizer attempts to update the model's parameters more frequently. Under the Stochastic Gradient Descent optimizer, weights are changed after loss computation on each training sample and it is able to converge in less time. But to get the same convergence as gradient descent, it needs to slowly reduce the value of the learning rate [65][66].

$\theta=\theta-\alpha\cdot\nabla J(\theta;x(i);y(i))$ ……………………………………………………………………….(4)

where {x(i), y(i)} are the training examples.

Adam (Adaptive Moment Estimation)

Adam uses the momentum of the first and second order. This optimizer can be used instead of the classical stochastic gradient descent procedure to update network weights iterative based on training data. Previous studies indicate that Adam outperformed all other methods in various training setups and experiments [67]. However, Adam introduces two new hyperparameters and complicates the hyperparameter tuning problem. The implementation of the Adam optimizer could be written as:

$$\widehat{\mathrm{m}_{\mathrm{t}}} = \frac{\mathrm{m}_{\mathrm{t}}}{1-\beta_1^{\mathrm{t}}} \tag{5}$$

$$\widehat{\mathrm{v}_{\mathrm{t}}} = \frac{\mathrm{v}_{\mathrm{t}}}{1-\beta_1^{\mathrm{t}}} \tag{6}$$

where m_{t} and v_{t} are values of the first moment (the *Mean)* and the second moment the uncentered variance) of the gradients, respectively.

The weights are updated based on:

$$\theta_{t+1}=\theta_t - \frac{\eta}{\sqrt{\widehat{v_t}}+\epsilon}\widehat{m_t} \tag{7}$$

The values for the parameters β1 = 0.9, β2 = 0.999, the step size parameter η = 0.001, while the parameter used to avoid division by zero $\epsilon = 10^{-8}$.

2.4.2.3 Batch Normalization

Normalization is the process of organizing a database used to bring the numerical data to a common scale without distorting its shape. The main aim of this process is to reduce redundancy and improve data integrity. Batch normalization is a process to make neural networks faster and more stable through the normalization of the layers' inputs by re-centering and re-scaling. It scales layers outputs to have a mean of 0 and a variance of 1 [68]. The outputs are scaled in such a way as to train the network faster and reduce problems due to poor parameter initialization [69][70].

Given the batch input from layer h, we calculate the mean of this hidden activation as:

$$\mu = \frac{1}{m}\sum h_i \tag{8}$$

where m is the number of neurons at layer h.

After calculating the mean, the next step is to calculate the standard deviation of the hidden activations as:

$$\sigma = \left[\frac{1}{m}\sum(h_i-\mu)^2\right]^{1/2} \tag{9}$$

Then, we subtract the mean from each input and divide the whole value with the sum of the standard deviation and the smoothing term (ε).

$$h_{i(norm)} = \frac{(h_i-\mu)}{\sigma+\epsilon} \tag{10}$$

where ϵ assures numerical stability within the operation by stopping a division by a zero value.

2.4.2.3 Hyperparameters

As its name indicates they are ‘top-level’ parameters that control the learning process and the model parameters that result from it. They are usually fixed before the actual training process begins. Learning rate, batch size, and the number of epochs are common training parameters that determine the learning algorithm’s behavior. Moreover, Hyper-parameters include architectural characteristics such as the number of layers and layer size. Additional parameters introduced by CNNs include the number of filters, filter size, padding, and stride and these are some common examples of hyperparameters [71][72].

2.4.3 Transfer Learning and Fine-tuning

Transfer learning is the reuse of a pre-trained model on a new but related problem. For example, knowledge gained while learning to recognize food could apply when trying to recognize a drink. Indeed, most real-world problems typically do not have millions of labeled data points to train deep-learning complex models. Currently, transfer learning becomes very popular in deep learning because it helps to train deep neural networks with comparatively little data [73].

The general idea is to use the knowledge a model has learned from a task with a lot of available labeled training data in a new task that doesn’t have much data. The first step to performing transfer learning is taking layers from a previously trained model. Then, to maintain the information they contain during feature training rounds we freeze its layers. After freezing, add some new trainable layers on top of the frozen layers. It helps the model to learn to turn the old features into predictions on a new dataset. Then train the new layers on a new dataset. The last step is fine-tuning**,** which consists of unfreezing the entire model and re-training it on the new data with a very low learning rate. This can potentially achieve meaningful improvements by incrementally adapting the pre-trained features to the new data [51].

There are a lot of pre-trained models. Keras, for example, provides numerous pre-trained models. Keras applications are deep learning models that are made available alongside pre-trained weights. A few keras provided pre-trained models are discussed below.

2.4.3.1 VGG16

VGG16 stands for Visual Geometry Group. VGG16 is a CNN model proposed by Karen Simonyan and Andrew Zisserman in 2014 in their work titled “Very deep convolutional networks for large-scale image recognition” [74]. Their model achieved 92.7% top-5 test accuracy on the ImageNet dataset which contains 14 million images belonging to 1000 classes.

The input to the network is an image of dimensions (224, 224, 3). Unlike other previous models, VGG16 uses a very small 3 x 3 receptive field (filters) throughout the entire network with a stride of 1 pixel. In some of the layers, it also uses 1*1 pixel which is used to manipulate the number of input channels. A stack of multiple (usually 1, 2, or 3) convolution layers of filter size 3 x 3 and stride one followed by the same padding of 1 pixel to prevent the spatial feature of the image. Then a max-pooling layer of size 2 x 2 is followed. Finally, three fully connected layers are presented, two with size 4,096 and the last one with size 1,000. The last one is the output layer with Softmax activation. The size of 1,000 refers to the total number of possible classes in ImageNet [75]. Figure 13 presents the architecture of the VGG16 model.

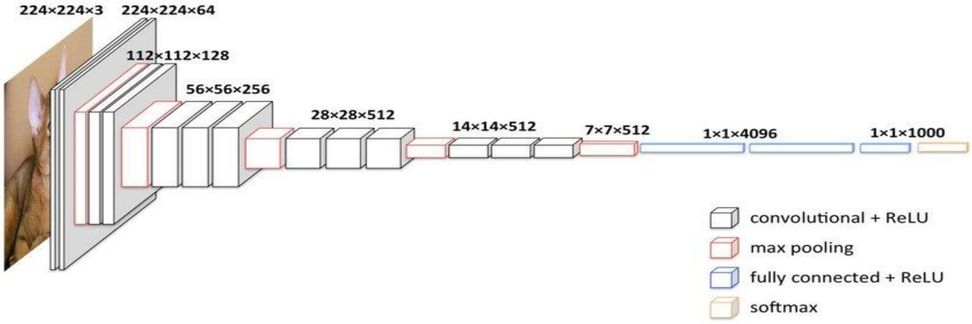


Figure 13: The architecture of the VGG-16 model [76].

2.4.3.2 VGG19

The concept of the VGG19 model is the same as the VGG16 except that it supports 19 layers (see also Fig. 14). The number 16 and 19 indicates the number of weight layers in the model. Therefore ‘19’ tells that VGG19 has three more convolutional layers than VGG16.

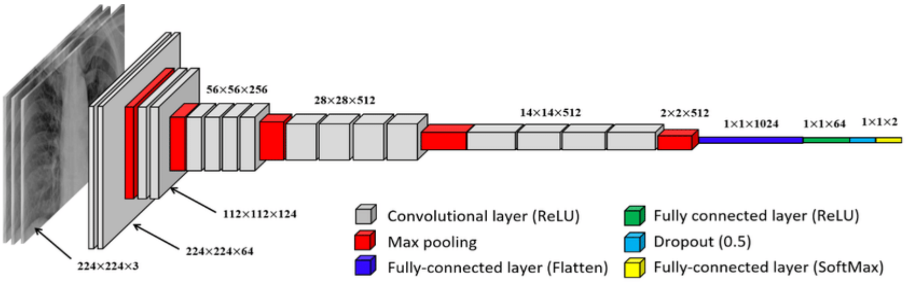


Figure 14: The architecture of the VGG-19 model [77].

2.4.3.3 ResNet50

ResNet-50, short for Residual Networks is a convolutional neural network that is 50 layers deep. It is proposed by Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun in their 2015 computer vision research paper titled “Deep Residual Learning for Image Recognition”**.** In ResNet-50, whose architecture has been depicted in Fig. 15, there are two kinds of blocks named *Identity Block* and *Convolutional Block*. The identity block is the standard block used in ResNets and corresponds to the case where the input activation has the same dimension as the output activation. Whereas, convolutional block uses when the input and output dimensions don’t match up. The ResNet-50 model consists of 5 stages each with a convolution and identity block. Each convolution block has 3 convolution layers and each identity block has 3 convolution layers. It has over 23 million trainable parameters [78].

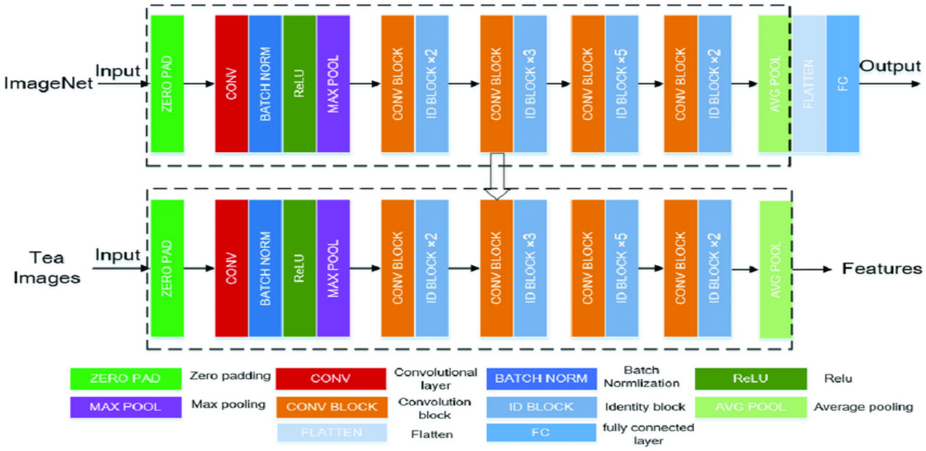


Figure 15: The architecture of ResNet50 model [79].

2.4.4 Deep Segmentation Model

Depending on the user’s interest, any image consists of both useful and useless information. Image segmentation separates an image into regions, each with its particular shape and border, extracting potentially meaningful areas for further processing, like classification and object detection. The goal of a segmentation task is to label each pixel of the input image with the respective class representing a specific object/body. Therefore, it provides pixel-by-pixel details of an object. Segmentation is different from classification tasks as it entails partitioning image pixels into different classes [80].

Although numerous methods are available in the literature that are successfully employed for image segmentation including edge-detection based, K-means clustering, histogram-based, thresholding and watersheds, in recent years, DL based methods showed remarkable performance for image segmentation tasks with higher accuracy [81]. There are various deep-learning models available for image segmentation tasks. Some established deep-learning models used for image segmentation are explained below.

2.4.4.1 Fully Convolutional Network (FCN)

In 2015, Jonathan Long, Evan Shelhamer, and Trevor Darrell introduced "fully convolutional" networks that take input of arbitrary size and produce correspondingly-sized output with efficient inference and learning. They defined the space of fully convolutional networks (FCN), explained their application to spatially dense prediction tasks and drew connections to prior models. Their study adapted contemporary classification networks (AlexNet, VGG Net, and GoogLeNet) into fully convolutional networks and transfer their learned representations by fine-tuning to the segmentation task [80].

A FCN is a special type of convolutional neural network that provides a segmented image of the original image where the required elements are highlighted as needed. It uses convolutional layers to process varying input sizes and can work faster. Under FCN networks, the features are merged from different stages in the encoder which varies in the coarseness of semantic information. Then, upsampling of learned low-resolution semantic feature maps is done using learned deconvolutions or partially avoiding the reduction of resolution altogether in the encoder using dilated convolutions at the cost of computation. At the end, the final output layer has a large receptive field and corresponds to the height and width of the image, while the number of channels corresponds to the number of classes (see also Fig. 16).

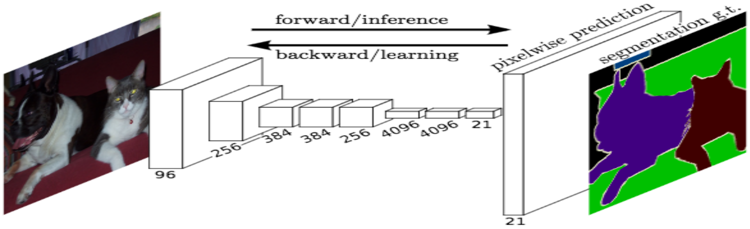


Figure 16: The FCN end-to-end dense prediction pipeline [82].

2.4.4.2 SegNet Neural Network

In 2015, Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla introduced a novel deep fully convolutional Encoder-Decoder Architecture for semantic pixel-wise segmentation termed SegNet [83]. The main idea behind the SegNet segmentation model is that its core trainable segmentation engine consists of an encoder network, and a corresponding decoder network followed by a pixel-wise classification layer. Unlike FCN, SegNet encoder doesn't use the fully connected layers and hence is lightweight network with lesser parameters. SegNet involves encoding the input image into low dimensions and then recovering it with orientation invariance capabilities in the decoder. In order to use and keep high-frequency details intact in the segmentation, SegNet uses unpooling to upsample feature maps in the decoder. This generates a segmented image at the decoder end (see also Fig. 17 for the architecture) [83].

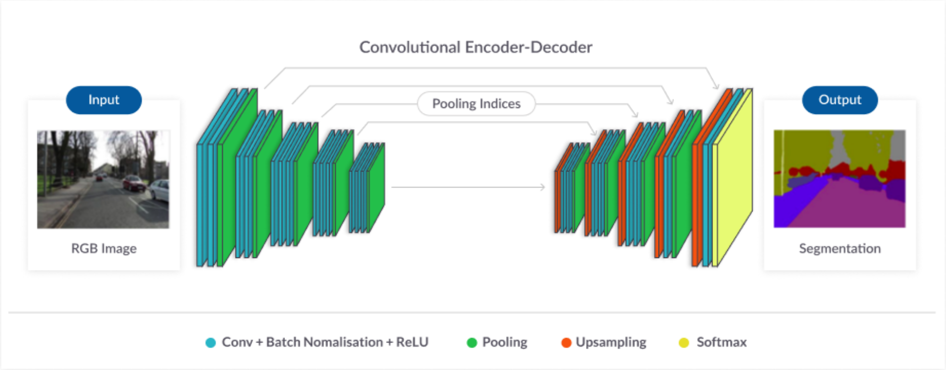
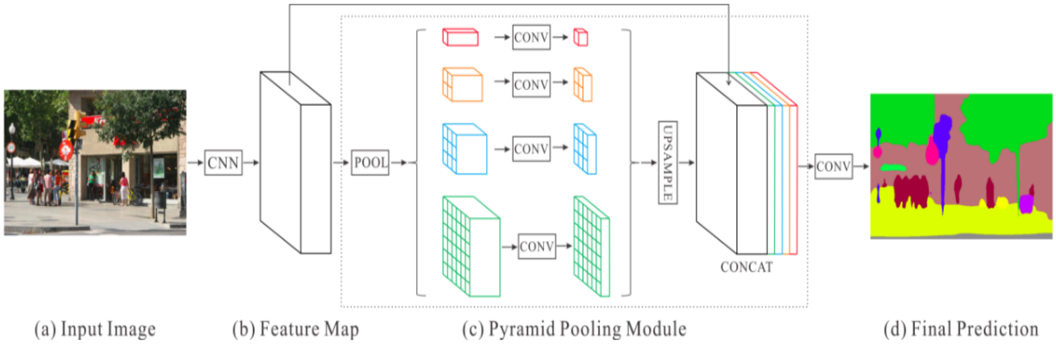


Figure 17: SegNet Neural Network [84].

2.4.4.3 Pyramid Scene Parsing Network (PSPnet)

In 2016, Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia proposed a pyramid scene parsing network (PSPNet) that provides a superior framework for pixel-level prediction tasks. This segmentation model modifies the base ResNet architecture by incorporating dilated convolutions and the features, after the initial pooling, are processed at the same resolution (1/4th of the original image input) throughout the encoder network until it reaches the spatial pooling module (see Fig. 18). PSPnet uses the introduction of auxiliary loss at intermediate layers of the ResNet to optimize overall learning [85].



(a) Input Image
(b) Feature Map
(c) Pyramid Pooling Module
(d) Final Prediction

Figure 18: PSPnet Neural Network [86].

2.4.4.4 Modified Retinal Analysis and Grading Network ($RAG - Net_{v2}$)

In 2020, Taimur Hassan, Muhammad Usman Akram, Naoufel Werghi, and Muhammad Noman Nazir introduced a deep retinal analysis and grading framework (RAG-FW) that is a hybrid convolutional framework that extracts multiple retinal lesions from OCT scans and utilizes them for lesion-influenced grading of retinopathy as per the clinical standards [87]. RAG-FW performs simultaneous pixel-level segmentation and scan-level classification. However, this network has been found to have limitations in discriminating similar texture objects, like retinal boundaries, when their transitional variations are small.

By considering its limitation, in 2021, Hina Raja and his three colleagues proposed a RAG-Net modified version, dubbed $\text{RAG} - \text{Net}_{v2}$, whereby the contextual-aware unit is built upon atrous convolutions (also known as dilated convolution) [88]. The first RAG-Net possesses kernels with smaller receptive fields (field of view) that do not retain accurately the contextual information of small and similar textural regions. The modified RAG-Net (see also Fig. 19) comes up with atrous convolutions. The atrous convolutions are formulated in a residual fashion which greatly enhances the kernels' receptive field to perform more broad and context-aware filtering while maintaining the same spatial resolution.

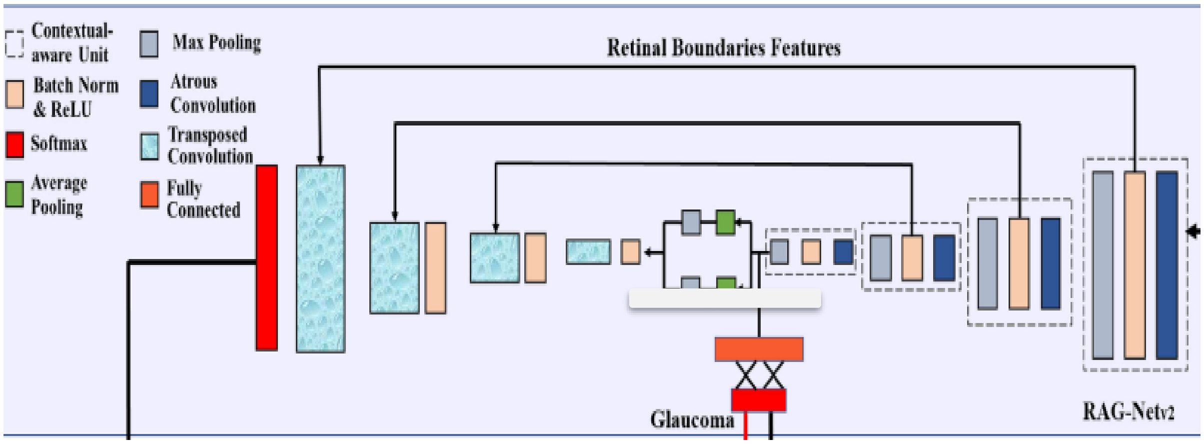


Figure 19: Hybrid Convolutional Network [88].

CHAPTER THREE

LITERATURE REVIEW

3.1 Introduction

Appropriate management of glaucoma depends on the clinician's ability to diagnose. In order to assist and improve the diagnosis process, automatic detection of glaucoma using different modalities has been tried in several studies. There are related works done on glaucoma diagnosis using AI and image processing techniques based on different imaging based diagnosis modalities. This chapter provides a review of the existing works.

3.2 Machine Learning Applications

In one of the previous studies, Tehmina Khakil et al. conducted research on the issue of improvement of automated detection of glaucoma based on fundus images. The study proposed a reliable computer-aided glaucoma diagnosis system based on a combination of hybrid structural and textural features. Under the hybrid structural feature (HSF) set, the study used different structural glaucoma conditions for classifying a sample with the help of a support vector machine (SVM). This module mainly included cup-to-disc ratio (CDR) computation, vertical rim-to-disc ratio (RDR) calculation, and cup shape analysis from the enhanced (Enh) image. Whereas, the hybrid texture feature (HTF) module classified and analyzed the samples using a total of 115 texture and intensity-based features with the help of SVM. The system classifies the features separately. Then based on those input feature vectors, both modules used SVM independently and make a binary decision about glaucoma. The system then makes a final decision after correlating the results of both classifiers. If both modules categorize the input image as glaucomatous, then the system labels the input image as glaucoma. If both modules grade it as healthy, then it is classified as healthy. Else, if the results of both modules do not match, then it is considered as suspect [11].

In 2020, Tehmina Shehryar, Muhammad U. Akram and Samina Khalid presented an improved automated detection of glaucoma by correlating fundus and SD-OCT image analysis. For reliable diagnosis of glaucoma, they proposed a hybrid computer-aided-diagnosis (H-CAD) system that integrates both fundus and OCT imaging technologies. The fundus module analyzes the changes caused by glaucoma on the outer layer of the eye's posterior part by considering a variety of structural and textural features. Then after analyzing both features, they used SVM classifier for the classification and making decisions. Whereas, in the OCT module, the cup-to-disc ratio (CDR) is computed by examining the internal layers of the retina. Finally, based on a triple synthesis of three diverse systems (two from structural and textural features of the fundus module and one from the OCT module), the system makes decisions. They tested their system on two publically available datasets and one local dataset. Their results showed that the three modules, HSF, HTF and OCT, achieved an overall accuracy of 76%, 84%, and 95%, respectively. The fused results, which combined the HSF, HTF, and OCT modules, offered 100% sensitivity and 96% overall accuracy [89].

3.3 Deep Learning

Xiangyu Chen and his colleagues used a convolutional neural network (CNN) with deep learning (DL) architecture for automated glaucoma diagnosis. The net of the CNN contains six layers with weights: the first four are convolutional and the remaining two are fully connected. Then the output of the last fully-connected layer is fed to a soft-max classifier for glaucoma prediction. The study used extracted Region of Interest (ROI) of the funduscopy images as the input to the CNN. In order to evaluate the glaucoma diagnosis performance of their proposed system, the authors performed experiments on two glaucoma fundus image datasets. The study utilized the area under the curve (AUC) of the receiver operation characteristic curve (ROC) to evaluate the performance of glaucoma diagnosis. Based on the results they reported, the AUC values of their method on the first and second datasets were 0.831 and 0.887, respectively [90].

Two other studies were conducted by Jinho Lee et al. in April 2020 and Ko Eun Kim et al. in July 2020. Both papers introduced a deep learning model based on traditional OCT reports, consisting of deviation and thickness maps of the RNFL and ganglion cell-inner plexiform layer (GC-IPL) analysis taken from SD-OCT (Cirrus, version 6.0; Carl Zeiss Meditec Inc.). The first study, conducted by Jinho Lee et al. took the extracted bottleneck feature of RNFL and GCIPL thickness and deviation map, acquired from SD-OCT optic disk and macular scans. Then they used these measurements to evaluate the diagnostic efficacy of their DL model (Neural Search Architecture (NAS) Network) for distinguishing glaucomatous and normal images. Accordingly, their system achieved an AUC of 0.99 (95% CI) with 94.7% sensitivity and 100.0% specificity. The latter work, conducted by Ko Eun Kim et al., also developed and validated a DL system for glaucoma diagnosis using deviation and thickness map of the RNFL and GC-IPL analysis acquired by SD-OCT. The system also evaluated heat maps to visualize the diagnostic pattern of their DL (VGG16 model) pattern. Their results showed that the DL system using a combination of OCT maps, i.e. the RNFL and GC-IPL deviation maps, demonstrated the highest diagnostic performance (AUC = 0.979), While the results obtained using all four OCT maps, i.e. the deviation and thickness maps of the RNFL and the GC-IPL, were also comparable (AUC = 0.977). The least performance was obtained when using that of the RNFL and GC-IPL thickness maps (AUC = 0.964) [91][21].

In another study, Stefan Maetschke et al. in 2019 proposed a DL technique that distinguishes glaucomatous eyes from normal eyes by taking direct row, unsegmented OCT volumes of the optic nerve head (ONH) using a 3D CNN. The authors provided an observational study by comparing the accuracy of their proposed system, with various feature-based machine learning algorithms. Importantly, this study considered the demographic background such as gender and race distribution, Intraocular Pressure (IOP), Mean Field Defects (MD), and Glaucoma Hemifield Test (GHT). For the feature-based approach, they took a set of 22 measurements computed by the Cirrus OCT scanner to train different machine learning algorithms including Naïve Bayes (Gaussian), Logistic Regression, SVM, Random Forest, Gradient Boosting, and Extra Trees. Their results indicated that Logistic regression achieved the highest test result (AUC = 0.89) closely followed by linear SVM. On the other hand, for feature-agnostic approaches, down-sampled 3D-OCT volumes were inputted to the CNN. This DL approach achieved a substantially higher AUC value of 0.94. Moreover, this approach provided insight into which regions of an OCT volume (specifically the neuroretinal rim and the optic disc cupping as well as the lamina cribrosa (LC) and its surrounding areas) are important for glaucoma detection. In general, the study attempted to demonstrate the superiority of their proposed DL approach over the classical feature-based machine learning techniques [92].

Xi Wang et al. introduced Multi-centered glaucoma OCT image screening with semi-supervised joint structure and function multi-task learning. The study used both structure analysis and function regression to address the glaucoma OCT image screening problem. By bringing out some reasons, the authors deduced that training a B-scan-wise deep network using B-scan OCT images is feasible and reasonable. Their method consists of two major parts. In the first part, which comprises functional regression, the method used a semi-supervised learning technique to address the missing label problem of visual field measurement. In the second part, they trained a multi-task learning network to unify structural analysis and surrogate-driven function regression. Most importantly, they built the largest glaucoma OCT image dataset involving 975,400 B-scans. Then, by using another independent dataset (Stanford dataset) containing 246,200 B-scans, their method was tested by two large-scale datasets. Their results indicated that the proposed methods achieved AUC value of 0.977 on their dataset and 0.933 on the Stanford dataset. However, this method focused on changes on the optic disk for algorithm development and evaluation and their dataset is composed of large OCT B-scans from OCT volumes of the optic disc [93].

3.4 Summary

As shown in the literature review above, automatic glaucoma diagnosis using deep learning algorithms is an important aspect of reliable and effective glaucoma detection and diagnosis. Even though OCT imaging is gaining increasing popularity in measuring the structural changes of the eyes, an unsatisfactory number of DL based automated methods have been developed based on OCT images to screen glaucoma. Moreover, most of the studies conducted on OCT images are focused on the optic nerve head (ONH) scan and traditionally measure thickness, thickness maps, and deviation maps. However, recent studies have shown that different areas of the macula are important for detecting subtle, early abnormalities of glaucoma [12][24]. These studies found that the diagnostic ability of segmented RNFL and GC-IPL to discriminate between normal and glaucoma eyes was high for early diagnosis. Therefore, by considering the research gap and the necessity for automatic early detection of glaucoma, the current work introduces a deep learning algorithm for an early glaucoma detection system based on macular SD-OCT B-scans.

CHAPTER FOUR

MATERIALS AND METHODS

4.1 Introduction

This chapter provides information about the study population and the method used for OCT image segmentation and classification. It involves five main stages: data collection, image preprocessing & data augmentation, deep segmentation model development, future selection, and deep classification. Figure 20 presents the general block diagram showing the proposed automated glaucoma diagnosis scheme that uses macular SD-OCT images as inputs.

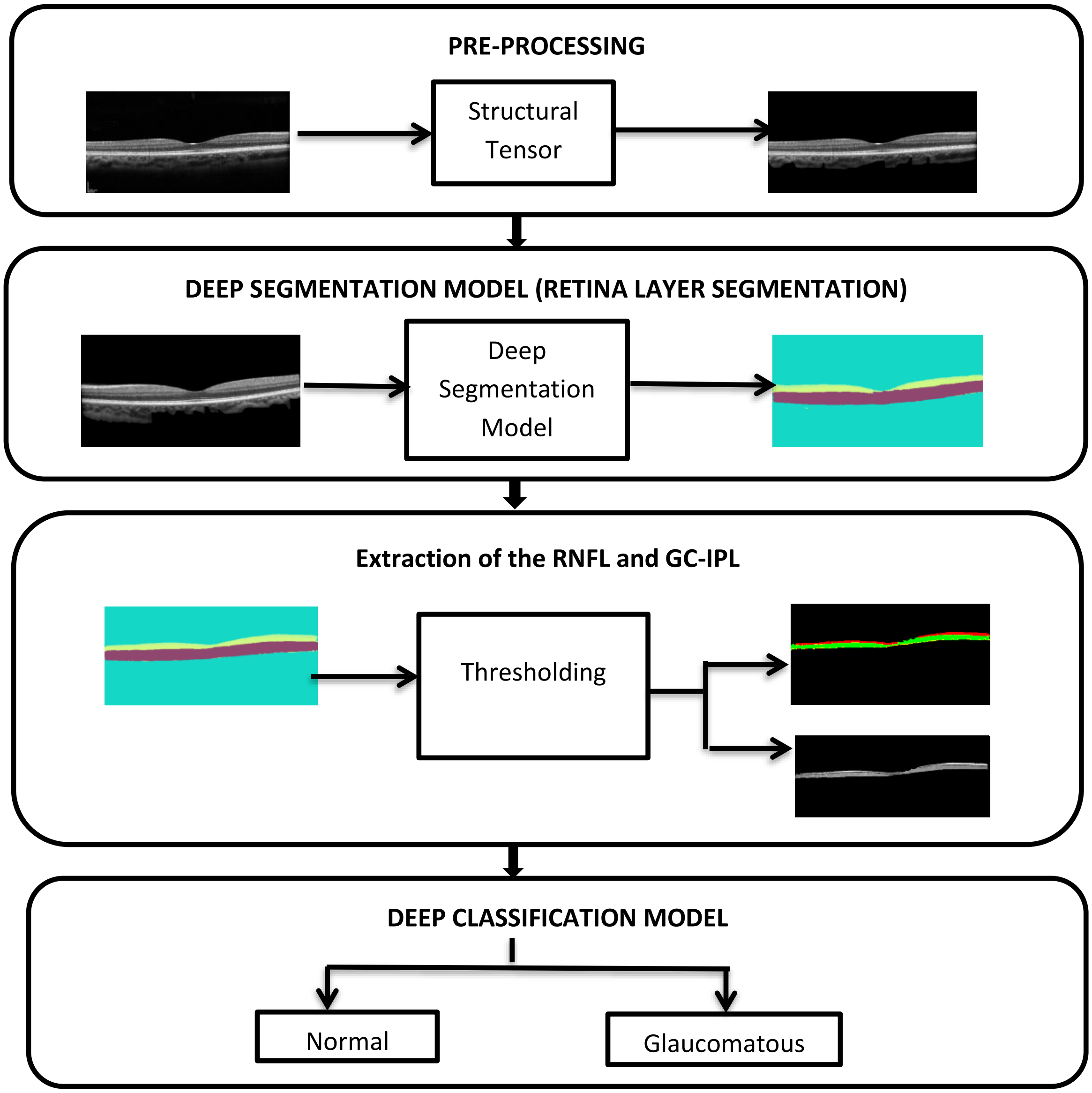


Figure 20: General block diagram of the proposed automated glaucoma diagnosis system.

First of all, the input OCT scan is preprocessed to separate the retinal area from the rest of the background. Afterwards, the preprocessed scan is passed to the deep segmentation model that segments the scan into four classes: RNFL, GC-IPL, the rest of the retina, and the background. Since only the GCC part (the RNFL and GC-IPL combined) is needed for classification, the feature selection method is applied to extract only the RNFL and GC-IPL. Finally, CNN-based classification is utilized to classify samples into normal or glaucomatous using the segmented GCC region as input.

The GCCs were segmented from four different regions of the macula: central, superior outside, inferior outside and inferior inside regions of the macula. It is believed that changes in these vulnerable regions are most important in glaucoma identification. Classification was performed using GCC regions taken from each of these regions separately and in combination.

The segmentation and classification models in the study were implemented in Python based on Keras with the Tensorflow backend. Experiments were performed on a PC with Intel(R) Core(TM) i7-6600U CPU @ 2.60GHz 2.81 GHz, installed RAM 8.00 GB, and system type 64-bit operating system, x64-based processor. The model's performance has been evaluated using several quantitative metrics. For both segmentation and classification, different pre-trained networks were compared. Segnet, PSPnet and $RAGNet_{v2}$ were used for segmentation and VGG16, VGG19 and ResNet50 were used for classification purpose.

4.2 Study Population

The macular SD-OCT image data was collected at Myungsung Christian Medical Center (Korea Hospital), Addis Ababa, Ethiopia between December 2021 and July 2022. The macular scans were performed on a Spectralis SD-OCT model (Heidelberd Engineering, 3.9µm axial resolution, 1.9mm scanning depth). Scanning was done using Posterior Pole horizontal macular scans protocol. The protocol includes the acquisition of 19 B-scans. The images included in this study had a scan quality > 25 and an automated real-time (ART) of 9 - 11.

A total of 1,262 macular SD-OCT B-scans were acquired and the current study focused on the four macular regions: central, superior outside, inferior outside and inferior inside regions of the macula. A total 306 (108 normal and 198 glaucomatous), 324 (108 normal and 216 glaucomatous), 325 (108 normal and 217 glaucomatous), and 307(108 normal and 199 glaucomatous) scans were taken from the central, superior outside, inferior outside and inferior inside macula regions, respectively. The data contained both glaucomatous and normal (cases and control groups) macula SD-OCT B-scans. Each group comprised both men and women with an age equal to or older than 18 years old. Cases were labeled (as glaucomatous or control) by a trained ophthalmologist with >2 years of experience.

The study's inclusion requirements for the control group included at least two or more of the followings:

- No VF defects: no defects deeper than $p < 2\%$ in the visual field;
- Not previously diagnosed with glaucoma and no disc changes for glaucoma;
- No treatment/no review story;
- All green OCT RNFL and OCT GC-IPL thickness analysis;
- Absence of macular pathology (which might affect the GCC region of the retina or any other layers of the retina).

The inclusion criteria for the case group were as follows:

- Patients with VF defects. However, based on other reliable diagnoses proof, some scans without VF defects are also included. If they are follow-up patients, their earlier scans before the VF defect are also taken.
- If there is no indication of normal tension glaucoma (NTG), only scans with intraocular pressure (IOP) higher than 21 mm Hg are included.
- The OND glaucomatous structural changes were described in the patient's card.
- Glaucomatous eyes with other macular pathology (which might affect the GCC region of the retina or any other layers of the retina) were excluded from the data.

Final labeled (as glaucomatous or control) data was generated according to the criteria mentioned above. In cases where labeling needed arbitration, a senior glaucoma specialist with >7 years of experience reviewed the cases and his diagnoses were considered final.

A total of 272 annotated cross-sectional images, taken from the four regions of the macula, are used to train, validate, and test the segmentation model. To create an accurate and precise ground truth, the dataset was annotated by an ophthalmologist using a labeling tool named 'Labelme'. This tool saves the annotated images in .json format. The .json files were converted into a .png format using the json_to_dataset command in Python. The raw data acquired from the OCT machine were in .jpg format and were converted to .png to fit the format of the annotated images.

Table 1 summarizes the study population considered in the current study for both segmentation as well as classification tasks.

Table 1: Summary of the study population used in the current study.

Model	**Region**	**Case (Glaucomatous /Normal)**	**# of Cross-sectional Images**	**Class**		
				Training	Validation	Test
Segmentation	Region 1 (Central)	Both	52	32	10	10
	Region 2 (Superior Outside Macula)	Both	65	40	13	12
	Region 3 (Inferior Outside Macula)	Both	65	40	13	12
	Region 4 (Inferior Inside Macula)	Both	90	54	18	18
Classification	Region 1 (Central)	Glaucomatous	198	158	40	-
		Normal	108	86	22	-
	Region 2 (superior Outside Macula)	Glaucomatous	216	173	43	-
		Normal	108	86	22	-
	Region 3 (Inferior Outside Macula)	Glaucomatous	217	174	43	-
		Normal	108	86	22	-
	Region 4 (Inferior Inside Macula)	Glaucomatous	199	159	40	-
		Normal	108	86	22	-

4.3 Pre-processing and Data Augmentation

4.3.1 Pre-processing

The purpose of the preprocessing step is to remove background artifacts and noisy content to obtain accurate extraction of RNFL, GC-IPL, and GCC regions and achieve better classification performance. The preprocessing is performed through a structure tensor. It is a second-moment matrix, which is computed from the gradients of the image data. It performs by highlighting the predominant orientations of the image gradients within the specified neighborhood of a pixel. The procedures are described as follows:

- The gradient of the image data (I) is computed and the individual elements of the tensor matrix are smoothed through convolution with a Gaussian filter.

  For an image I(x,y), the structure tensor S is given by:

$$S= \begin{pmatrix} W*(\nabla X.\nabla X) & W*(\nabla X.\nabla Y) \\ W*(\nabla Y.\nabla X) & W*(\nabla Y.\nabla Y) \end{pmatrix} \quad \text{or} \quad S= \begin{pmatrix} W*I_x^2 & W*I_xI_y \\ W*I_xI_y & W*I_y^2 \end{pmatrix} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(11)$$

where W is a smoothing kernel (typically a Gaussian), $*$ denotes convolution while ∇X and ∇Y represent the partial derivatives of the image gradients in the horizontal and vertical directions.

The image gradients ∇X and ∇Y are oriented at 0° and 90°, respectively. The individual elements of the tensor matrix are smoothed through convolution with a Gaussian kernel (W).

- Get a unique tensor of S and its coherent tensor: After computing the gradient of the image (S) and computing it for each pixel, it obtains three unique tensors. Then the one having the maximum coherency according to their norm is selected.
- Convert tensor to image: Afterward, the selected tensor is transformed into an 8-bit gray scale image.
- Extract the ILM and choroidal boundaries: Extraction of the ILM and choroidal boundaries is performed by detecting the first and last foreground-background transitions in each column of the scan. To avoid outliers, constrain the distance between consecutive pixels in the ILM and choroidal boundaries to be below a threshold τ (20) determined empirically [94].
- Interpolate gaps and smooth the ILM and Choroidal layers: Afterward, linear interpolation is performed to estimate the missing values in each layer and smoothed through median filtering.
- Multiply generated mask with the original scan: Finally, a retinal mask is generated which is multiplied with the original scan to isolate the retinal regions as shown in Fig. 21 below.

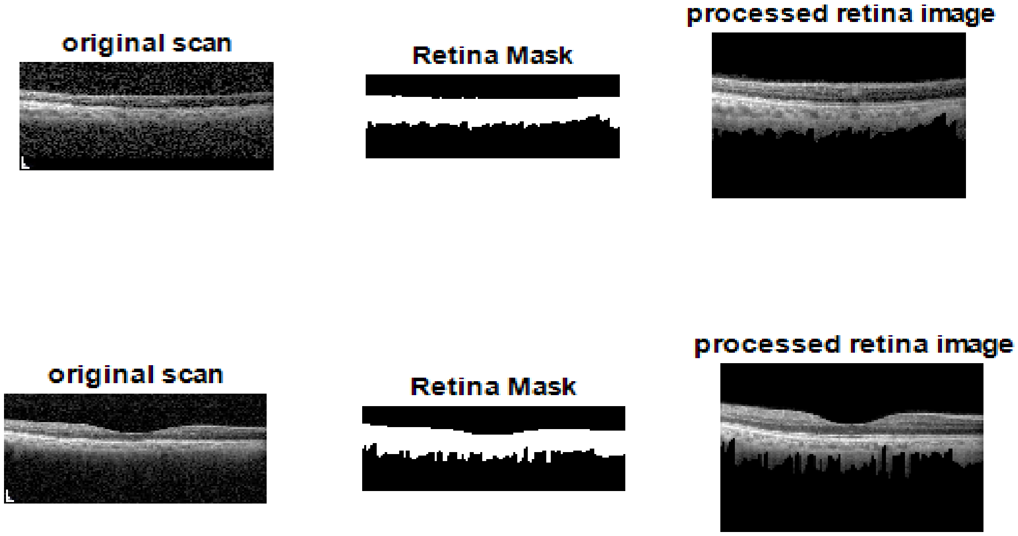


Figure 21: Preprocessing output: (top) taken from the superior outside and (bottom) central macular regions.

4.3.2 Data Augmentation

Image augmentation is a technique of altering the existing data to create more data for the model training process. Data augmentation is one of the most effective strategies for improving the model's performance and ability in image classifiers and segmentation. To compensate for the low number of collected training scans and balance the labeled data, augmentation is performed on available scans. Following augmentation, in order to fit the segmentation models, the images were resized to 768 x 576. The performed data augmentation is as follows:

- **Horizontal flip:** Flipping rearranges the pixels while protecting the features of the image. In the current research, all OCT scans were horizontally flipped.
- **Rotation:** After the horizontal flip, the scans were rotated between -3 to 3 degrees.
- **Noise addition:** Data augmentation was also performed through blurring the scans by adding noise. In the current research, a zero-mean white Gaussian noise with a variance of 0.01 was added to each scan to create a blurred version.

4.4 Deep Model for Retina Layer Segmentation

In the current study, segmentation of the macula SD-OCT scans done using three deep segmentation models: PSPNet, SegNet, and RAG-Net$_{v2}$. In order to choose a segmentation framework that can accurately segment the retinal layers (the RNFL, the GC-IPL and the rest of the retina), the study compared the performance of these three models.

A total of 272 cross-sectional images, taken from all four regions of the macula together with their annotations were used to train, validate and test the segmentation models. The dataset was annotated by an expert ophthalmologist. 60% of the dataset is used for training and the rest 40% was used for testing (20%) and validation (20%). To compensate for the low number of training scans, fine-tuning the weights of the pre-training segmentation models and augmentation of the training scans were performed. Figure 22 presents a typical pre-processed SD-OCT image, the annotation made by the expert ophthalmologist as well as the annotated image overlaid on the pre-processed image to show discrepancy between the two. This is basically the output of labelme annotator used by the ophthalmologist.

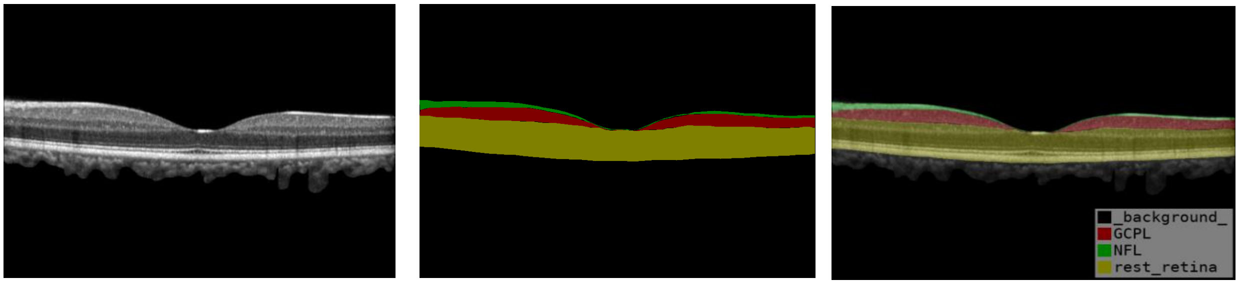


Figure 22: Pre-processed SD-OCT image (left), ophthalmologist annotated image (middle), and annotated image overlaid on pre-processed image (right).

The training is conducted for 30 epochs where each epoch lasted for 248 iterations. The optimization during the training is performed through ADADELTA [95] having a default learning rate of one with a decay factor of 0.95. Figure 23 presents a pre-processed SD-OCT scan inputted to the network and the final segmented image using SegNet model.

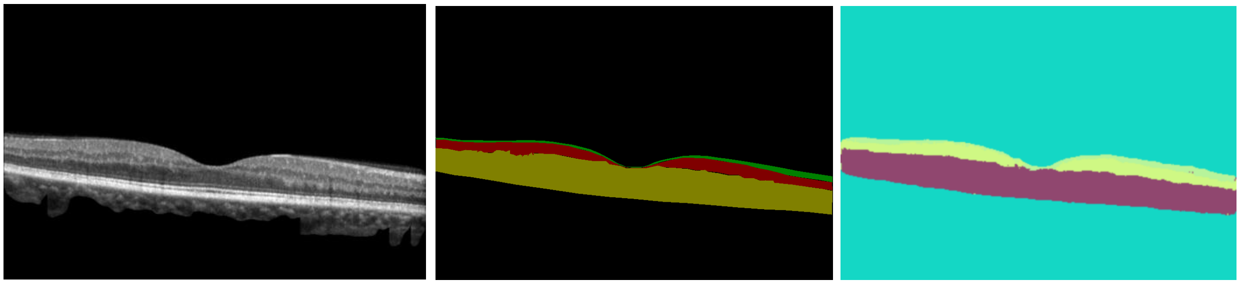


Figure 23: Test input for SegNet segmentation model (left), ophthalmologist annotated image (middle), and output of SegNet segmentation model (right).

The segmentation model segmented three regions of the retina from the back ground: the RNFL at the top, followed by the GC-IPL, and the rest of the retina at the bottom. Note that as the color difference between the RNFL region and the background is difficult to see. We will see later that this color difference will be more visible after RNFL and GC-IPL extraction is applied.

4.5 Extraction of the RNFL and GC-IPL

Glaucoma only affects the GCC region of the macula and the GCC region contains the RNFL and the GC-IPL. Hence, the final segmented image should contain only the GCC part before it is inputted to the classifier and that requires omission of the rest of the retinal part from the segmentation output. A simple thresholding was applied for this purpose. Figure 24 presents the output of this thresholding step. The resulting segmented image that contains only the RNFL and the GC-IPL was inputted into the classifier in two different ways. In the first case, the segmented image (see for example the third column in Fig. 24) is directly inputted to the classifier to allow it learn just the thickness information from extracted GCC images. In the second case, the segmented image was used as a mask to extract the RNFL and the GC-IPL regions from the original (raw) image and that is inputted into the classifier (see the fourth column in Fig. 24, for example).

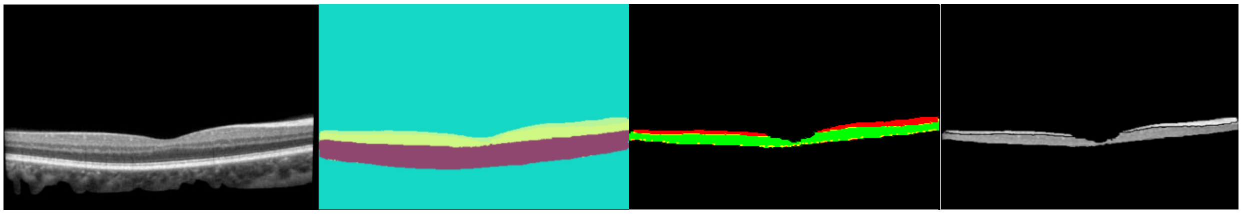


Figure 24: Pre-processed original image, segmented image, segmented GCC, masked GCC on the original image respectively (Left to Right).

4.6 Deep Classification Models

The segmented (and the masked) images are inputted into the classifier to categorize the cases into normal and glaucoma classes. There are several neural network architectures often used in the literatures, however, three popular ones, namely VGG16, VGG19, and ResNet50 are used in this study. These three CNN architectures were imported into python. Input image size of 224×224×3 was selected for the three models. For each model, the fully connected layer was replaced and the number of output layers was changed to 2 (normal or glaucomatous). The parameters of CNN layers are frozen and the fully connected layers for the classification task are trained by using glaucoma detection training data. The classification layer was also modified accordingly.

For each neural network architecture, training and testing were performed in 26 different ways: 3 instances from the superior outside macula, 3 from the inferior outside macula, 3 from inferior inside macula, another 3 from paired inferior outside and superior outside macula and 1 additional from the central macular (on the fovea) regions. This was done using both the segmented images and the masked images separately, making a total of 26 instances. As shown in Fig. 25, glaucoma highly affects the inferior inside macula, inferior outside macula, and superior outside macula regions. For each of the 26 instances, except the central region, training was done by varying the number of scans taken per region. Accordingly, 2, 3, and 4 scans per region are checked to find the optimum number of scans for each region that offers the best classification performance. For the central part, only 3 scans were considered. From Fig. 25, it is obvious that this region is less venerable to changes due to glaucoma compared to the inferior and superior regions.

Splitting the data into training and validation is required before starting the classification. In this regard, the images were randomly divided into two groups: 80% used for training and the remaining 20% for validation.

Training Parameter of the Classification Network

In order to choose the set of training parameters, a series of tests were carried out. SGDM and Adam optimizers were tested and it was observed that Adam performed much better than the SGDM optimizer hence the Adam optimizer was selected. Three learning rates: 0.01, 0.001 & 0.0001 were also tried during the training. The learning rate of 0.001 was chosen as the network achieved the highest accuracy than 0.01 & 0.0001 in the training and validation sets. The number of epochs was initially set at 30 but the classification performance didn’t show significant improvement after epoch number 25. Hence the number of the epoch was set at 25. Also, the categorical cross-entropy loss function was used in the classification layer as it is more popular in the current literature. Softmax activation function was chosen for training since it shows superior performance than other activation functions including ReLu. All CNN architectures employed pre-trained parameters based on ImageNet during initialization, and all the layers are fine-tuned during the training.

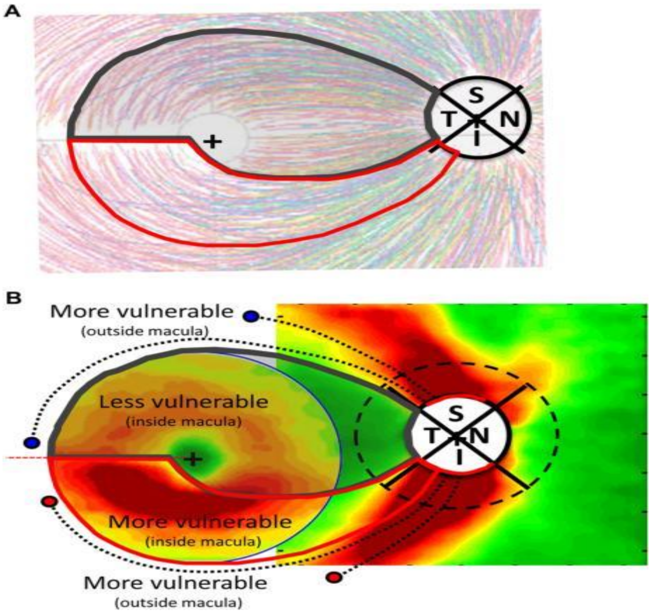


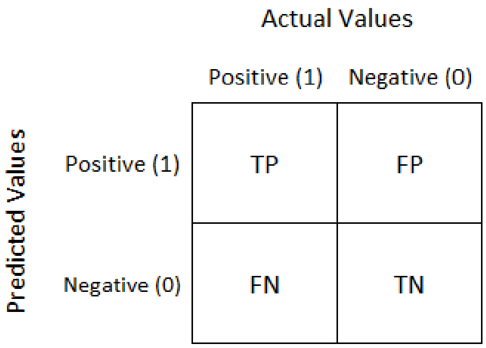
Figure 25: A schematic model of RNFL projections (top) and glaucomatous RGC+ and RNFL damage (bottom) [96].

4.7 Evaluation Metrics

The segmentation and classification performance of all the proposed models was evaluated using different evaluation metrics. Accuracy, Recall, F1 Score, Precision, Mean Dice Coefficient, Mean_IoU, and Class_Wise_IoU are used to evaluate the segmentation model [97]. On the other hand, the classification model's performance has been evaluated using accuracy, precision, recall, F1-score and AUC value.

➢ Confusion Matrix

A confusion matrix is a performance metric used to measure classification models where the output is binary or multiclass (see also Fig. 26).



		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figure 26: Confusion Matrix table [61].

Predicted values indicate values that are predicted by the model whereas, actual values imply values that are coming from the ground truth. In the current study, positive points belong to the normal eyes class and Negative points to the glaucomatous class.

TP stands for True Positive and indicates values that are actually positive and predicted positive.

FP stands for False Positive which implies values that are actually negative but predicted to be positive.

FN stands for False Negative and implies values that are actually positive but predicted to be negative.

TN stands for True Negative and indicates values that are actually negative and predicted to be negative.

➢ Accuracy

Accuracy is a metric that measures the number of predictions that are correct as a percentage of the total number of predictions that are made.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Number of total predictions}} = \frac{\text{TP+TN}}{\text{TP+ FP + TN + FN}} \quad \text{.....(12)}$$

- **Precision, Recall and F1 score**

Precision, Recall, and F1-score are the most common metrics that take into account class imbalance. Precision and Recall are the foundations of the F1 score**.**

Precision: Within everything that has been predicted as a positive, precision counts the percentage that is correct.

$$\text{Precision}= \frac{\text{TP}}{\text{TP+FP}} \quad \text{....(13)}$$

Recall: Recall tells us, within everything that actually is positive, how many did the model succeed to find.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad \text{....(14)}$$

F1 score (Dice Coefficient (DC)): F1 score is the harmonic mean of the positive predictive value and sensitivity. Its goal is to combine the precision and recall metrics into a single metric.

$$\text{F1 score} = \frac{\text{2 TP}}{\text{2TP + FP +FN}} \quad \text{....(15)}$$

OR

$$\text{F1 score} = 2*\frac{\text{Precision}*\text{Recall}}{\text{Precision+Recall}} \quad \text{....(16)}$$

In case of segmentation, the F1-score (or the dice coefficient (DC)) measures how well the proposed framework segments the RNFL, GC-IPL, and rest retinal regions as compared to their ground truths. In another word, the Dice Coefficient is 2 * the Area of Overlap divided by the total number of pixels in all classes**.** And it is computed through:

$$\text{F1 Score} =\frac{2 * \text{Area of Overlap}}{\text{Total Pixels Combined}} \quad \text{....(17)}$$

In this case TP indicates the correct extraction of positives (RNFL, GC-IPL, and rest retina regions), FP indicates the misclassified background pixels, and FN denotes those positive pixels which have been missed by the proposed segmentation framework. Afterwards, the mean dice coefficient (μDC) is computed by taking an average DC score computed scan-wise across the whole dataset.

Intersection Over Union (IoU, Jaccard Similarity Index or Overlap Index):

IoU is the area of overlap between the predicted segmentation and the ground truth divided by the area of union between the predicted segmentation and the ground truth [98].

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \qquad \ldots\ldots(19)$$

OR

$$\mathrm{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \qquad \ldots\ldots(20)$$

The value of this metric ranges from 0 to 1 (0 to 100%), with 0 signifying no overlap and 1 signifying perfect overlap. Mean IoU is a common evaluation metric for image segmentation, which first computes the IOU for each image to be segmented and then computes the average over the entire dataset.

ROC (Receiver operating characteristics) and AUC (Area Under the Curve):

The ROC curve is a graphical representation that shows the cutoff between the false positive and false negative rate for each potential cutoff. AUC illustrates the degree or measure of separability and is used as a summary of ROC curve [99].

CHAPTER FIVE

RESULTS AND DISCUSSION

The deep segmentation and classification models have been thoroughly evaluated on the locally collected dataset for the GCC-aware diagnosis of glaucoma. The experimental results presented and discussed below were acquired in both segmentation and binary classification results using different models. First, segmentation results are present that evaluated different segmentation models. Afterwards, a comparison of the three classification models is presented.

5.1 Segmentation Results

This section involves choosing the segmentation framework that can accurately extract the retinal regions (RNFL, GC-IPL and rest of the retina) to perform the classification effectively. For this purpose, the study compared the performance of three deep segmentation models.

5.1.1 Results Using Modified RAG-Net ($\mathbf{RAG-Net}_{v2}$)

$\mathrm{RAG-Net}_{v2}$ segmentation model achieved 95.5% validation accuracy. The test data was used to further evaluate the model's performance. A summary of the result is shown in Table 2, where we can observe that RAG-Netv2 achieved 66.6% Recall, 71% F1-score, 76.6% Precision, 75.5% Mean dice-coefficient, and 65.6% Mean_IoU. Although it performed well on segmentation of background and rest-retina sections (96.7% and 81.7% IoU, respectively), class-wise IoU tells that $\mathrm{RAG-Net}_{v2}$ segmentation performance of GC-IPL and RNFL segmentation is very poor (47.8% and 36.2% IoU, respectively).

Table 2: Results Using modified RAG-Net (RAG-Net _v2).

Recall	**F1 Score**	**Precision**	**Mean Dice Coefficient**	**Mean_IoU**	**Class_Wise_IoU**			
					Background	GC-IPL	RNFL	Rest-retina
66.6%	71%	76.6%	75.5%	65.6%	96.7%	47.8%	36.2%	81.7%

5.1.2 Results Using PSPnet

The PSPnet segmentation model achieved 97.5% validation accuracy. A summary of the results based on test data is shown in Table 3, where we can observe that PSPnet achieved 86.8% Recall, 86% F1-score, 85.3% Precision, 87.1% Mean dice-coefficient, and 77.2% Mean_IoU. In terms of Class-wise IoU, the model achieved 98.5%, 67.8%, 59%, and 88.4% segmentation performance for background, GC-IPL, RNFL, and Rest-retina sections, respectively, showing a significantly better performance than the modified RAG-Net.

Table 3: Results Using PSPnet.

Recall	**F1 Score**	**Precision**	**Mean Dice Coefficient**	**Mean_IoU**	**Class_Wise_IoU**			
					Background	GC-IPL	RNFL	Rest-retina
86.8%	86%	85.3%	87.1%	77.2%	98.5%	67.8%	59%	88.4%

5.1.3 Results Using SegNet

The SegNet segmentation model achieved 97.9% validation accuracy. Table 4 presents evaluation matrices computed for the performance of the SegNet model. Accordingly, this method offered 87% Recall, 87.5% F1-score, 88% Precision, 89% Mean dice coefficient, and 81% Mean_IoU. Moreover, the SegNet model offered Class wise IoU of 98.8% for background, 79.6% for GC-IPL, 72.2% for the RNFL, and 90.1% for the Rest-retina regions, showing its superior performance over the previous two models.

Table 4: Results Using SegNet.

Recall	**F1 Score**	**Precision**	**Mean Dice Coefficient**	**Mean_IoU**	**Class_Wise_IoU**			
					Background	GC-IPL	RNFL	Rest-retina
87%	87.5%	88%	89%	81%	98.8%	79.6%	72.2%	90.1%

5.2 Classification Results

The thinness of the RNFL, GC-IPL, and GCC profiles highlights the degeneration of RGCs, which directly reflects glaucomatous progression. The current study utilized the extracted GCC region of the SegNet segmentation model to perform RGC-aware classification of healthy and glaucomatous subjects. As mentioned in the previous chapter, the GCC extraction was performed in two ways. In the first case, the segmented GCC is directly inputted to the classifier. In the second case, the GCC region from the raw SD-OCT scan obtained using the segmentation output as a mask is considered as input to the classifier. In both cases, the rest of the retinal region as well as the background are both set to zero (dark pixels). The classification performances derived using the three CNN architectures described in the previous chapter are summarized below.

5.2.1 VGG16

5.2.1.1 Central Macula Region (Passes Through Fovea)

Three OCT B-scans that pass through the fovea are taken as different images to build the training and testing data. The confusion matrix in Fig. 27 and the performance metrics displayed in Table 5 indicate that the VGG16 model resulted in 84.3% accuracy, 88.24% precision, 88% recall, 87.98% f1-score, and 87.99% AUC during model validation.

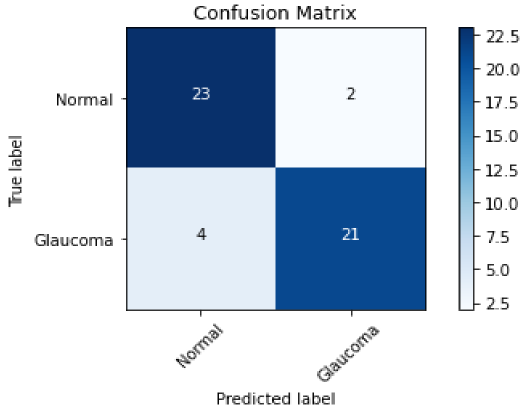


Figure 27: Confusion matrix computed for VGG16 applied on three scans taken from the central macula region

Table 5: Classification Accuracy, Precision, Recall, F1-score and AUC values computed for the VGG16 model applied on the central macula region.

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
84.3	88.24	88.00	87.98	87.99

5.2.1.2 Superior Outside Macula Region

As shown in Table 6, the training and validation accuracies of VGG16 model on the superior outside macula region indicate that the optimum number of images that should be taken from this region is three. Accordingly, further evaluation of the VGG16 model applied on the superior outside macula region is done by considering three B-scans of the region as three different images. The model offered 96% training and 94% validation accuracies when the segmented images are directly inputted while it achieved 96% training and 92.8% validation accuracies when the masked images are used instead.

Table 6: Classification results using VGG16 applied on superior outside macula region.

No of scans taken per region	**Segmented Image**		**Original Masked Image**	
	Training accuracy	Validation accuracy	Training accuracy	Validation accuracy

Two images	94.8%	82.8%	96%	83.9%
Three images	96%	94%	96%	92.8%
Four images	92.8%	85.6%	92%	87%

To further evaluate the classification performance of the VGG16 model applied on the superior outside macula region, the confusion matrix was generated that allows compute performance matrices. Figure 28 and Table 7 present the generated confusion matrix and the evaluation metrics, respectively. Accordingly, the model offered 93.3% Precision, 91.7% Recall, 91.8% F1-score, 91.7% AUC value, and 94.3% overall accuracy.

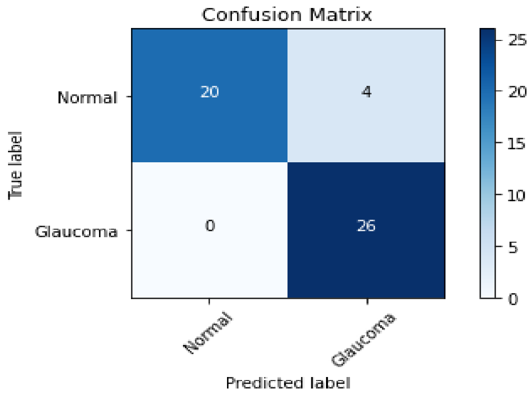


Figure 28: Confusion matrix computed for VGG16 applied on three scans taken from the superior outside macula region.

Table 7: Classification Accuracy, Precision, Recall, F1-score and AUC values computed for the VGG16 model applied on the superior outside macula region.

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
94.3	93.3	91.7	91.8	91.7

5.2.1.3 Inferior Outside Macula Region

Table 8 shows that the training and validation accuracies computed from this region were almost similar when using the segmented images and the masked images as inputs to the VGG16 model. In both cases, using three scans offered the best performance.

Table 8: Classification results using VGG16 applied on inferior outside macula region.

No of scans taken per region	**Segmented Image**		**Original Masked Image**	
	Training accuracy	Validation accuracy	Training accuracy	Validation accuracy
Two images	96%	83.9%	96%	82%
Three images	96%	85.5%	95.6%	85.6%
Four images	95%	82%	92%	84.8%

The binary classification performance of the VGG16 model when applied on the inferior outside macula region (3 scans per region taken as three different images) achieved 85.5% Accuracy, 90.9% Precision, 88.7% Recall, 89.5% F1-score, and 88% AUC value. Figure 29 presents the confusion matrix while Table 9 displays the computes performance metrics.

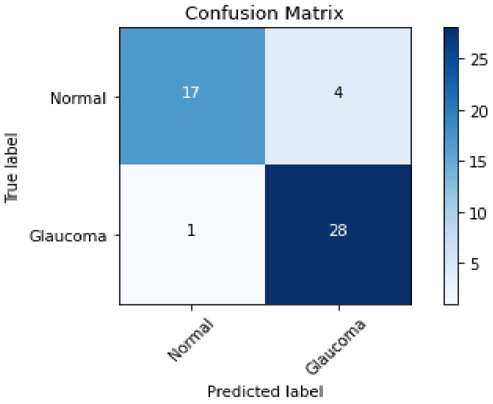


Figure 29: Confusion matrix computed for VGG16 applied on three scans taken from the inferior outside macula region.

Table 9: Classification Accuracy, Precision, Recall, F1-score and AUC values computed for the VGG16 model applied on the Inferior outside macula region.

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
85.5%	90.9	88.7	89.5	88

5.2.1.4 Inferior Inside Macula Region

As shown in Table 10, the validation accuracy of VGG16 model when applied on the inferior inside macula region indicates that the optimum number of images that should be taken from this region is three. Accordingly, further evaluation of the VGG16 model applied on the inferior inside macula region is carried out by considering three B-scans of the region as three different images. The performance of the model was generally higher when using the segmented images directly (89.9% validation accuracy) than when using the masked images (87.7% validation accuracy).

Table 10: Classification results using VGG16 applied on inferior inside macula region.

No of scans taken per region	**Segmented Image**		**Original Masked Image**	
	Training accuracy	Validation accuracy	Training accuracy	Validation accuracy
Two images	95.8%	71.6%	98%	67.8%
Three images	94.7%	89.8%	96%	87.7%
Four images	94%	87%	93%	83%

To further evaluate the classification performance of the VGG16 model applied on the inferior inside macula region, the confusion matrix was generated that allows compute performance matrices. Figure 30 and Table 11 present the generated confusion matrix and the evaluation metrics, respectively. Accordingly, the model offered 87% Precision, 87% Recall, 87% F1-score, 87% AUC value, and 89.84% overall accuracy.

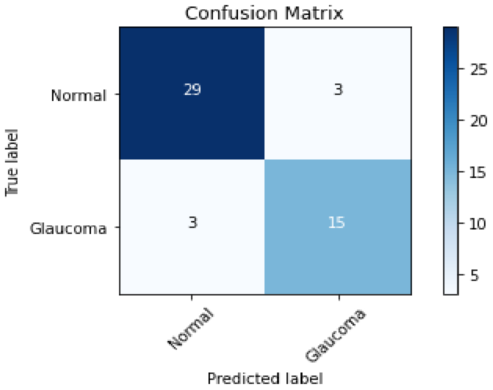


Figure 30: Confusion matrix computed for VGG16 applied on three scans taken from the inferior inside macula region

Table 11: Classification Accuracy, Precision, Recall, F1-score and AUC values computed for the VGG16 model applied on the inferior inside macula region.

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
89.8%	87	87	87	87

5.2.1.5 Paired Inferior and Superior Outside Macula Region

As shown in Table 12, the highest VGG16 classification performance was achieved by taking two segmented images from both the inferior and superior outside regions of the macula. The confusion matrix and the results of different evaluation metrics are presented in Fig. 31 and Table 13, respectively. The results indicate that the VGG16 model achieved 90.7% Accuracy, 90% Precision, 84.09% Recall, 84.98% F1-score, and 84.09% AUC values during the model validation stage.

Table 12: Classification results using VGG16 applied on paired inferior and superior outside macula regions.

No of scans taken per region	**Segmented Image**		**Original Masked Image**	
	Training accuracy	Validation accuracy	Training accuracy	Validation accuracy
Two images	92.5%	90.7%	88.6%	87.4%
Three images	92.4%	82.8%	94.1%	81.4%
Four images	91%	88.6%	90.6%	86.5%

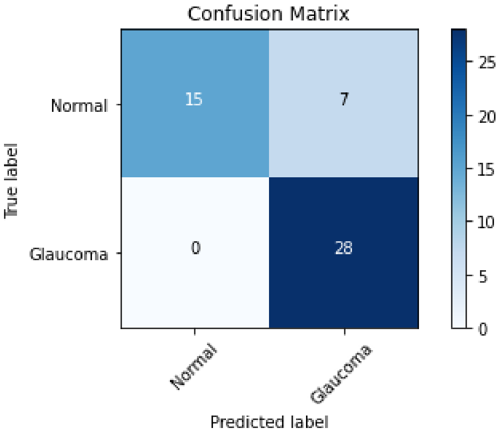


Figure 31: Confusion matrix computed for VGG16 model applied on three scans taken from the paired inferior and superior outside macula regions.

Table 13: Classification Accuracy, Precision, Recall, F1-score and AUC values computed for the VGG16 model applied on the paired inferior and superior outside macula regions.

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
90.7%	90.00	84.09	84.98	84.09

5.2.1.6 Summary of the VGG16 Performances Applied on the Four Macula Regions

Table 14 and Fig. 32 present summary of the performances of the VGG16 model when applied on the four macular regions deemed important in glaucoma detection. Overall, the VGG16 model applied on three scans taken from the superior outside macula region achieved the best accuracy of 94.3%. The other performance metrics, Precision, Recall, F1-score, and AUC values also indicate the same results. Figure 33 presents the ROC curves used to compute the corresponding AUC values.

Table 13: Best results computed using the VGG16 model applied on four different regions of the macula.

Regions	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
Central Macula	84.3	88.24	88.00	87.98	88
Superior Outside Macula	94.3	93.3	91.7	91.8	91.7
Inferior Outside Macula	88	88.3	87.8	87.9	88
Inferior Inside macula	89.8	87	87	87	87
Paired Inf-Sup Outside Macula	90.7	90.00	84.09	84.98	84.09

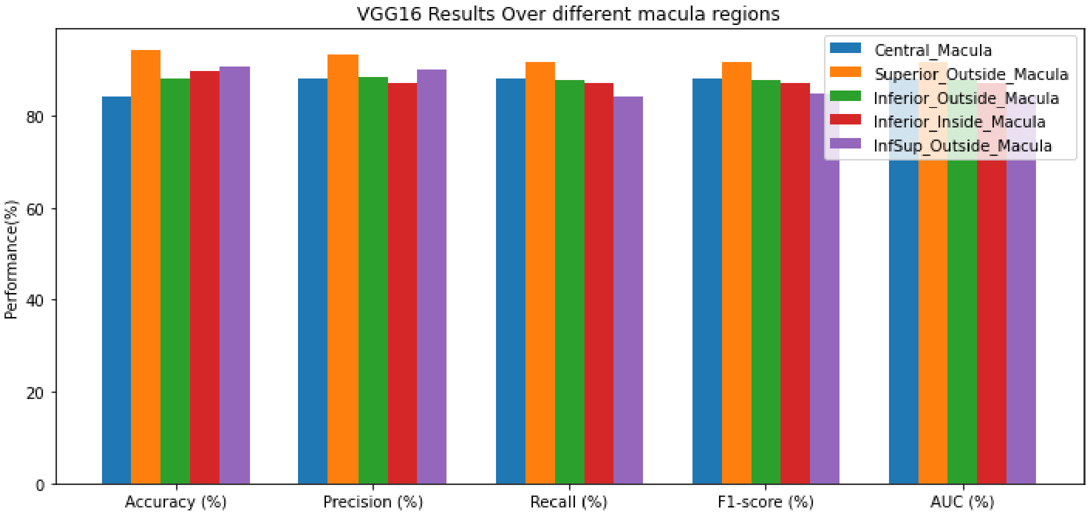


Figure 32: Bar plot showing evaluation metrics computed for the VGG16 model applied on four different regions of the macula.

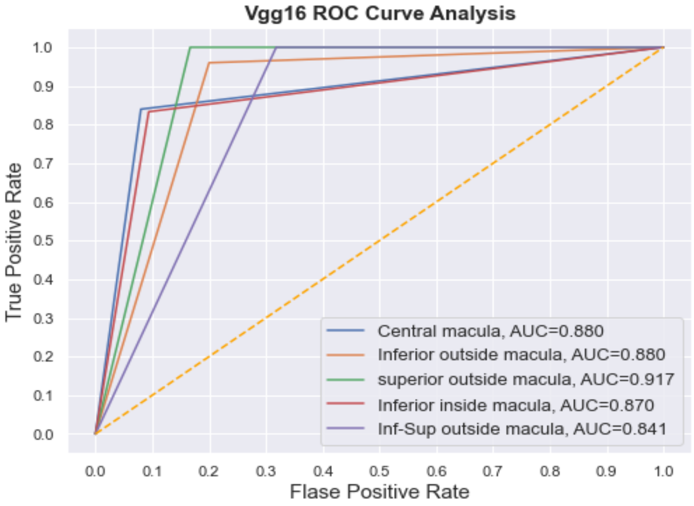


Figure 33: ROC curves and respective AUC values computed for the VGG16 model applied on four different regions of the macula.

5.2.2 VGG19

5.2.2.1 Central Macula Region (Passes Through Fovea)

Three OCT B-scans that pass through the fovea are taken as different images to build the training and testing data. The confusion matrix in Fig. 34 and the performance metrics displayed in Table 15 indicate that the VGG19 model achieved 82.4% Accuracy, 86.2% Precision, 86.4% Recall, 85.9%, F1-score, and 86.3 AUC value during model validation.

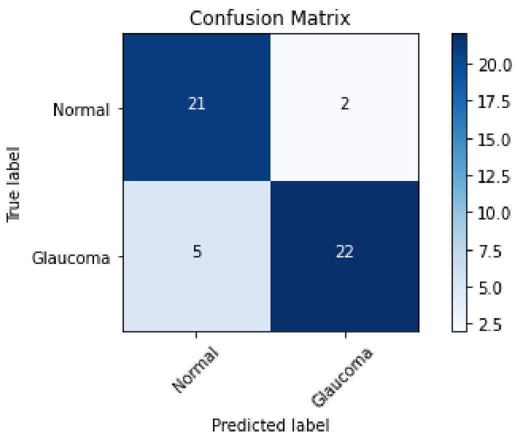


Figure 34: Confusion matrix computed for VGG19 applied on three scans taken from the central macula region

Table 14: Classification Accuracy, Precision, Recall, F1-score and AUC values computed for the VGG19 model applied on the central macula region.

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
82.4	86.2	86.4	85.9	86.3

5.2.2.2 Superior Outside Macula Region

As shown in Table 16, the validation accuracies of the VGG19 model on the superior outside macula region indicate that the optimum number of images that should be taken from this region is three. Accordingly, further evaluation of the VGG19 model applied on the superior outside macula region is done by considering three B-scans of the region as three different images. As a result, the model offered 85.6% validation accuracy when the segmented images are directly inputted while it achieved 67% validation accuracy when the masked images are used instead.

Table 15: Classification results using VGG19 applied on the superior outside macula region.

No of scans taken per region	**Segmented Image**		**Original Masked Image**	
	Training accuracy	Validation accuracy	Training accuracy	Validation accuracy
Two images	89.6%	79.7%	86%	77%
Three images	88%	85.6%	89%	67%
Four images	86.9%	80.3%	84.9%	78%

To further evaluate the classification performance of the VGG19 model for binary classification applied on the superior outside macula region, the confusion matrix was generated that allows calculation of performance metrics. Figure 35 and Table 17 present the generated confusion matrix and the evaluation metrics results. Accordingly, the model offered 83.51% Precision, 80.5% Recall, 81.08% F1-score, 80.50 AUC value and 85.6% overall accuracy.

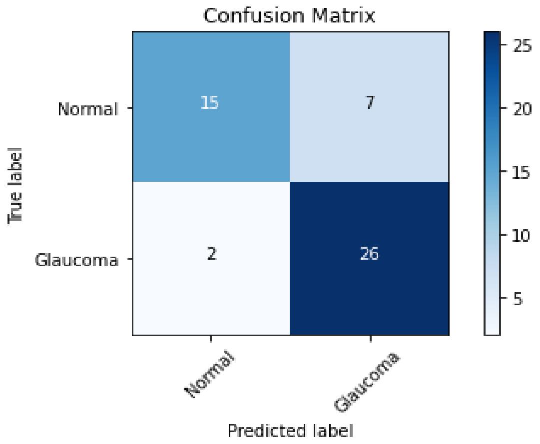


Figure 35: Confusion matrix computed for VGG19 applied on three scans taken from the superior outside macula region.

Table 16: Classification Accuracy, Precision, Recall, F1-score and AUC values computed for the VGG19 model applied on the superior outside macula region.

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
85.6	83.51	80.50	81.08	80.50

5.2.2.3 Inferior Outside Macula Region

Table 18 indicates that for highest validation accuracy of the VGG19 model, the optimum number of images that should be taken from the inferior outside macular region is two and using direct segmented images instead of the masked images. Accordingly, further evaluation of the VGG19 model applied on the inferior outside macula region was done by considering two B-scans of the region as two different images.

Table 17: Classification results using VGG19 applied on inferior outside macula region.

No of scans taken per region	**Segmented Image**		**Original Masked Image**	
	Training accuracy	Validation accuracy	Training accuracy	Validation accuracy
Two images	88.8%	85.7%	87.4%	78.3%
Three images	88.2%	75%	89%	75%
Four images	89%	83.7%	89%	79%

To further evaluate the classification performance of the VGG19 model applied on the inferior outside macula region, the confusion matrix was generated that allows compute performance matrices. Figure 36 and Table 19 present the generated confusion matrix and the evaluation metrics, respectively. Accordingly, the scheme achieved 85.7% Accuracy, 82.05% Precision, 82% Recall, 81.99% F1-score, and 82% AUC value.

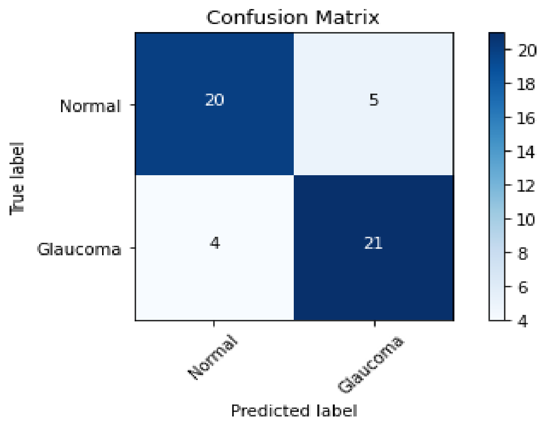


Figure 36: Confusion matrix computed for VGG19 applied on two scans taken from the inferior outside macula region

Table 18: Classification Accuracy, Precision, Recall, F1-score and AUC values computed for the VGG19 model applied on the inferior outside macula region.

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
85.7%	82.05	82	81.99	82

5.2.2.4 Inferior Inside Macula Region

As shown in Table 20, the validation accuracy of the VGG19 model when applied on the inferior inside macula region indicates that the optimum number of images that should be taken from this region is three. Accordingly, further evaluation of the VGG19 model applied on the inferior inside macula region is carried out by considering three B-scans of the region as three different images. The performance of the model was generally higher when using the segmented images directly (82% validation accuracy) than when using the masked images (80% validation accuracy).

Table 19: Classification results using VGG19 applied on inferior inside macula region.

No of scans taken per region	Segmented Image		Original Masked Image	
	Training accuracy	Validation accuracy	Training accuracy	Validation accuracy
Two images	88.9%	70%	88.6%	67%
Three images	86%	82%	85%	80%
Four images	83.8%	81%	84%	77%

To further evaluate the classification performance of the VGG19 model applied on the inferior inside macula region, the confusion matrix was generated that allows compute performance matrices. Figure 37 and Table 21 present the generated confusion matrix and the evaluation metrics, respectively. Accordingly, the model offered 82.03% Accuracy, 81.86% Precision, 81.49% Recall, 81.49% F1-score, and 81.49% AUC value.

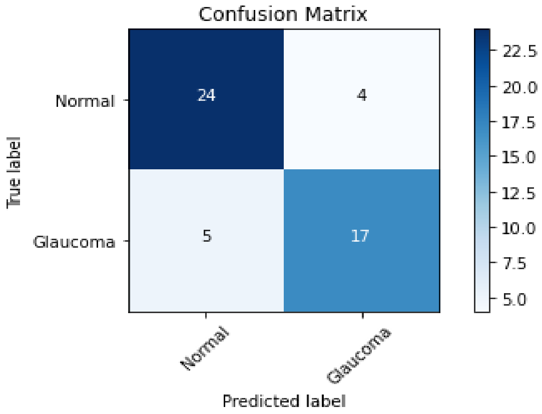


Figure 37: Confusion matrix computed for VGG19 applied on three scans taken from the inferior inside macula region.

Table 20: Classification Accuracy, Precision, Recall, F1-score and AUC values computed for the VGG19 model applied on the inferior inside macula region.

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
82.03%	81.86	81.49	81.49	81.49

5.2.2.5 Paired Inferior and Superior Outside Macula Region

As shown in Table 22, the highest VGG19 classification performance was achieved by taking two segmented images from both the inferior and superior outside regions of the macula combined. The confusion matrix and the results of different evaluation metrics are presented in Fig. 38 and Table 23, respectively. The results indicate that the VGG19 model achieved 88.12% Accuracy, 83.02% Precision, 83.02% Recall, 83.02% F1-score, and 83.02% AUC values during the model validation stage.

Table 21: Classification results using VGG19 applied on paired inferior and superior outside macula regions.

No of scans taken per region	**Segmented Image**		**Original Masked Image**	
	Training accuracy	Validation accuracy	Training accuracy	Validation accuracy
Two Images	87%	88%	85%	84%
Three Images	86.8%	81%	88%	75%
Four Images	84.8%	83%	85%	80.6%

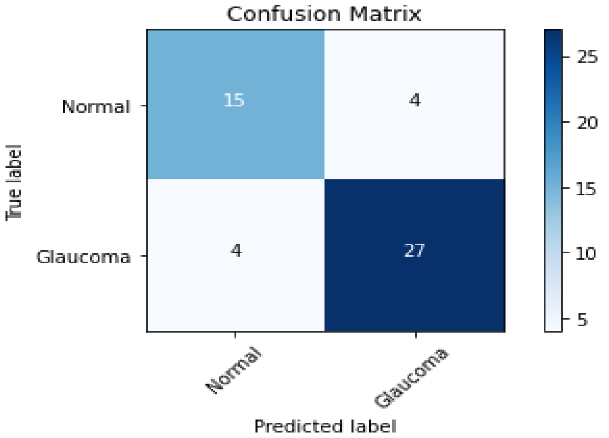


Figure 38: Confusion matrix computed for VGG19 applied on two scans taken from the paired inferior and superior outside macula regions.

Table 22: Classification Accuracy, Precision, Recall, F1-score and AUC values computed for the VGG19 model applied on the paired inferior and superior outside macula region.

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
88.12	83.02	83.02	83.02	83.02

5.2.2.6 Summary of the VGG19 Performances Applied on the Four Macula Regions

Table 24 and Fig. 39 present summary of the performances of the VGG19 model when applied on the four macular regions deemed important in glaucoma detection. Overall, the VGG19 model applied on three scans taken from the paired inferior-superior outside macula region achieved the best accuracy of 88.12%. However the other performance metrics, Precision, Recall, F1-score, and AUC values indicate that the best score of binary classification is achieved on the central macula region followed by the paired inferior-superior outside macula region. Figure 40 presents the ROC curves used to compute the AUC values for each of the different regions.

Table 23: Best results computed using the VGG19 model applied on four different regions of the macula.

Regions	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**AUC (%)**
Central Macula	85.6	86.2	86.4	85.9	86.39
Superior Outside Macula	85.6	83.51	80.50	81.08	80.5
Inferior Outside Macula	85.7	82.05	82.00	81.99	82
Inferior Inside macula	82.03	81.86	81.49	81.49	81.49
Paired Inf-Sup Outside Macula	88.12	83.02	83.02	83.02	83.02

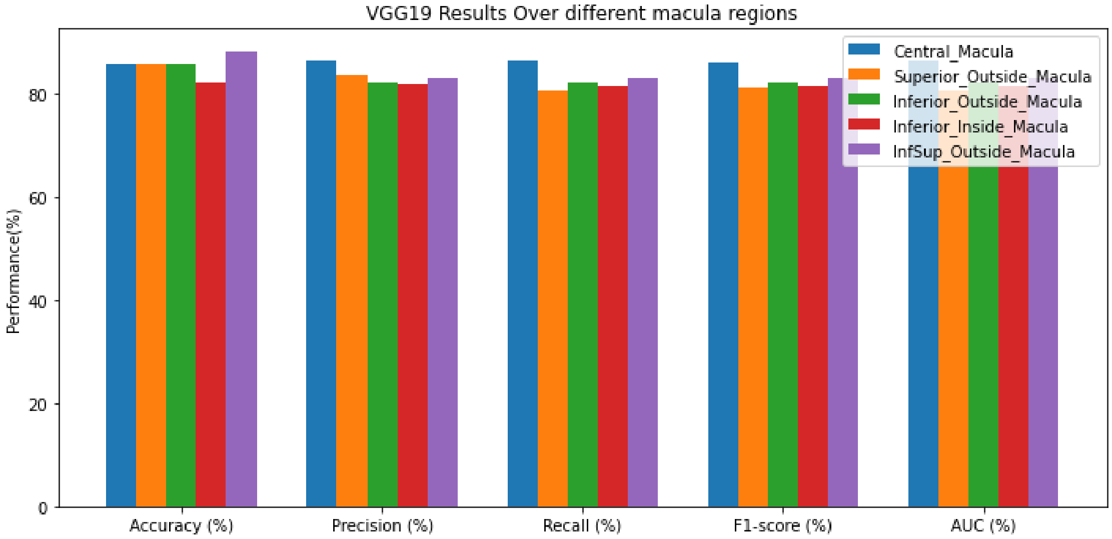


Figure 39: Bar plot showing evaluation metrics computed for the VGG16 model applied on four different regions of the macula.

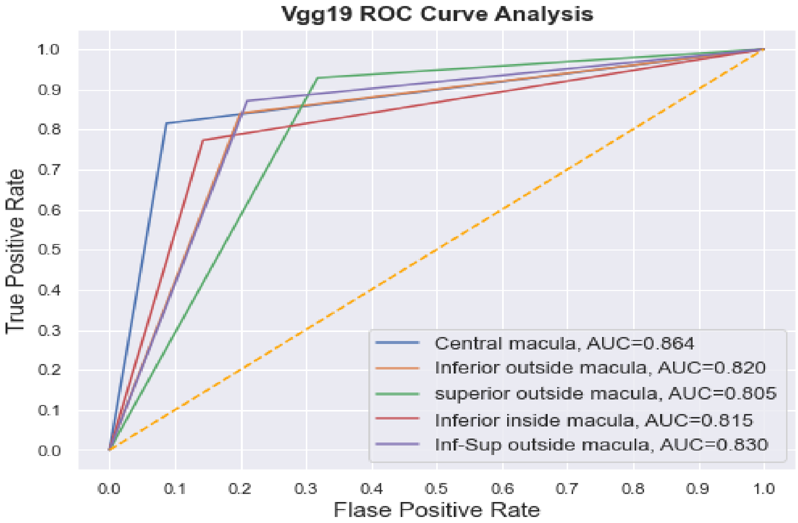


Figure 40: ROC curves and respective AUC values computed for the VGG19 model applied on four different regions of the macula.

5.2.3 ResNet50

5.2.3.1 Central Macula Region (Passes Through Fovea)

Three OCT B-scans that pass through the fovea are taken as different images to build the training and testing data. The confusion matrix in Fig. 41 and the performance metrics displayed in Table 25 indicate that the ResNet50 model resulted in 79.8% Accuracy, 83.4 Precision, 81.57 Recall, 81.6 F1-score, and 81.57 AUC values during model validation.

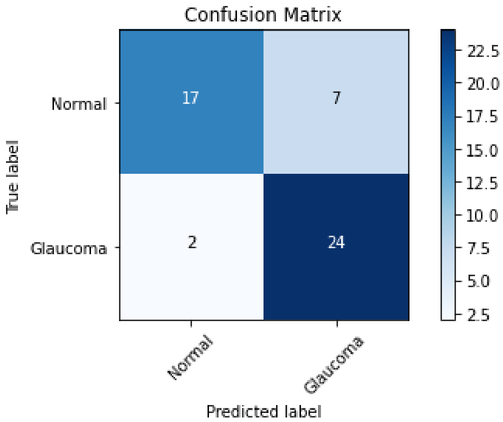


Figure 41: Confusion matrix computed for ResNet50 applied on three scans taken from the central macula region.

Table 24: Classification Accuracy, Precision, Recall, F1-score and AUC values computed for the ResNet50 model applied on the central macula region.

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
79.8	83.4	81.57	81.6	81.57

5.2.3 .2 Superior Outside Macula Region

As shown in Table 26, the validation accuracy of ResNet50 model when applied on the superior outside macula region indicates that the optimum number of images that should be taken from this region is two. Accordingly, further evaluation of the ResNet50 model applied on the superior outside macula region was done by considering two B-scans of the region as two different images. The model offered 77.6% validation accuracy when the segmented images are directly inputted while it achieved 71.9% validation accuracy when the masked images are used instead.

Table 25: Classification results using ResNet50 applied on superior outside macula region

No of scans taken per region	**Segmented Image**		**Original Masked Image**	
	Training accuracy	Validation accuracy	Training accuracy	Validation accuracy
Two images	84.2	77.6	85.6	71.9
Three images	88.9	69.7	77.9	69.7
Four images	87.7	59.5	80.5	59.4

Figure 42 and Table 27 present the generated confusion matrix and the evaluation metrics, respectively. Accordingly, the model offered 84.46% Precision, 85.83% Recall, 83.90% F1-score, 85.83% AUC value, and 77.6% overall accuracy.

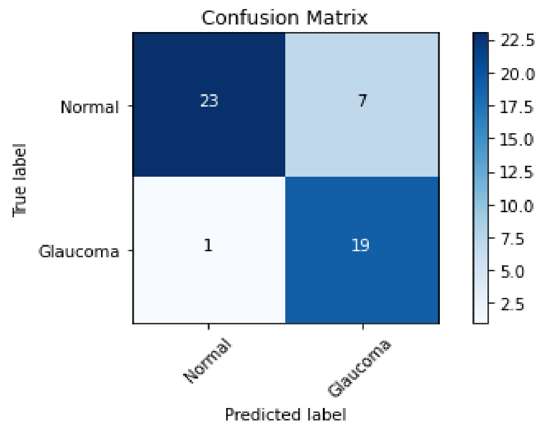


Figure 42: Confusion matrix computed for ResNet50 applied on two scans taken from the superior outside macula region

Table 26: Classification Accuracy, Precision, Recall, F1-score and AUC values computed for the ResNet50 model applied on the superior outside macula region.

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
77.6	84.46	85.83	83.90	85.83

5.2.3.3 Inferior Outside Macula Region

As shown in Table 28, the validation accuracy of ResNet50 model on the inferior outside macula region indicates that the optimum number of images that should be taken from this region is two. Accordingly, further evaluation of the ResNet50 model applied on the inferior outside macula region was done by considering two B-scans of the region as two different images. The model offered 81.7% validation accuracy when the segmented images are directly inputted while it achieved 75.6% validation accuracy when the masked images are used instead.

Table 27: Classification results using ResNet50 applied on inferior outside macula region.

No of scans taken per region	**Segmented Image**		**Original Masked Image**	
	Training accuracy	Validation accuracy	Training accuracy	Validation accuracy
Two images	91%	81.7%	92%	75.6%
Three images	84.9%	77.7%	88.9%	79.9%
Four images	75.7%	63.3%	80.7%	53.3%

The binary classification performance of the ResNey50 model when applied on the inferior outside macula region achieved 81.80% Accuracy, 78.04% Precision, 78.00% Recall, 77.9% F1-score, and 78% AUC value. Figure 43 presents the confusion matrix while Table 29 displays the computed performance metrics.

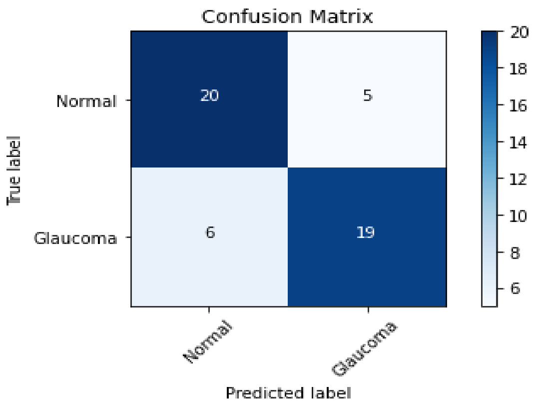


Figure 43: Confusion matrix computed for ResNet50 applied on two scans taken from the interior outside macula region.

Table 28: Classification Accuracy, Precision, Recall, F1-score and AUC values computed for the ResNet50 model applied on the inferior outside macula region.

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
81.8	78.04	78.00	77.9	78

5.2.3.4 Inferior Inside Macula Region

As shown in Table 30, the validation accuracy of ResNet50 model when applied on the inferior inside macula region indicates that the optimum number of images that should be taken from this region is four and the segmented images were more important than the masked images. Accordingly, further evaluation of the ResNet50 model applied on the inferior inside macula region was carried out by considering four B-scans of the region as four different images.

Table 29: Classification results using ResNet50 applied on inferior inside macula region.

No of scans taken per region	**Segmented Image**		**Original Masked Image**	
	Training accuracy	Validation accuracy	Training accuracy	Validation accuracy
Two images	86.3%	61%	80.7%	79.1%
Three images	82.9%	57.3%	80.9%	71.4%
Four images	89.7%	79.1%	79.8%	76.6%

To further evaluate the classification performance of the ResNet50 model applied on the inferior inside macula region, the confusion matrix was generated. Figure 44 and Table 31 present the generated confusion matrix and the evaluation metrics, respectively. Accordingly, the model offered 79.1% Accuracy, 65.87% Precision, 65.94% Recall, 65.88% F1-score, and 65.94% AUC values during model validation.

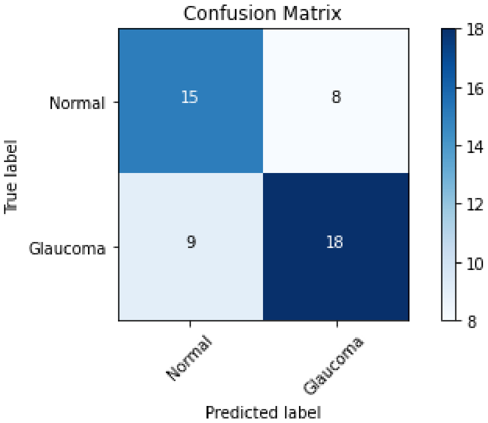


Figure 44: Confusion matrix computed for ResNet50 applied on four scans taken from the inferior inside macula region.

Table 30: Classification Accuracy, Precision, Recall, F1-score and AUC values computed for the ResNet50 model applied on the inferior inside macula region.

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
79.1	65.87	65.94	65.88	65.94

5.2.3.5 Paired Inferior and Superior Outside Macula Region

As shown in Table 32, the highest ResNet50 classification performance was achieved by taking four segmented images from both the inferior and superior outside regions of the macula. The confusion matrix and the different evaluation metrics are presented in Fig. 45 and Table 33, respectively. The results indicate that the ResNet50 model achieved 71.79% Accuracy, 83.64% Precision, 79.33% Recall, 79.17% F1-score, and 79.32% AUC values during the model validation stage.

Table 31 Classification results using ResNet50 applied on paired inferior and superior outside macula regions.

No of scans taken per region	**Segmented Image**		**Original Masked Image**	
	Training accuracy	Validation accuracy	Training accuracy	Validation accuracy
Two Images	52%	50%	61%	51%
Three Images	78%	70%	77.2%	62%
Four Images	78.1%	71.79%	78.1%	61.9%

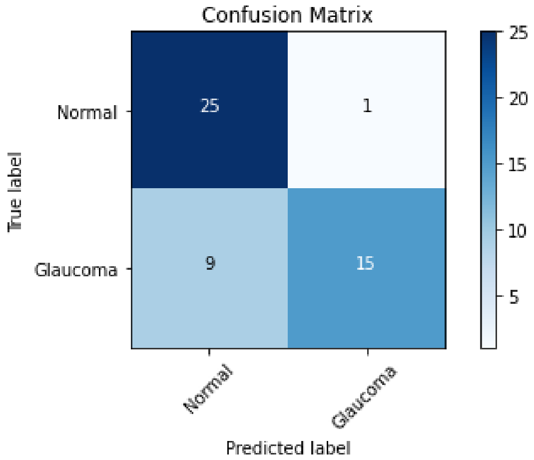


Figure 45: Confusion matrix computed for ResNet50 applied on four scans taken from the paired inferior and superior outside macula regions.

Table 32: Classification Accuracy, Precision, Recall, F1-score and AUC values computed for the ResNet50 model applied on the paired inferior and superior outside macula regions.

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
71.79	83.64	79.33	79.17	79.32

5.2.3.6 Summary of the ResNet50 Performances Applied on the Four Macula Regions

Table 34 and Fig. 46 present summary of the performances of the ResNet50 model when applied on the four macular regions deemed more important in glaucoma detection. Overall, the ResNet50 model applied on two scans taken from the inferior outside macula region achieved the best accuracy of 81.7%. However the other performance metrics, Precision, Recall, F1-score, and AUC values indicate that the best score of binary classification is achieved on the superior outside macula region. Figure 47 presents the ROC curves used to compute the AUC values for each of the different regions.

Table 33: Best results computed using the ResNet50 model applied on four different regions of the macula.

Regions	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**AUC (%)**
Central Macula	79.8	83.4	81.57	81.6	81.57
Superior Outside Macula	77.6	84.46	85.83	83.90	85.83
Inferior Outside Macula	81.7	78.04	78.00	77.9	78
Inferior Inside macula	79.8	65.87	65.94	65.88	65.94
InfSup Outside Macula	71.79	83.64	79.33	79.17	79.32

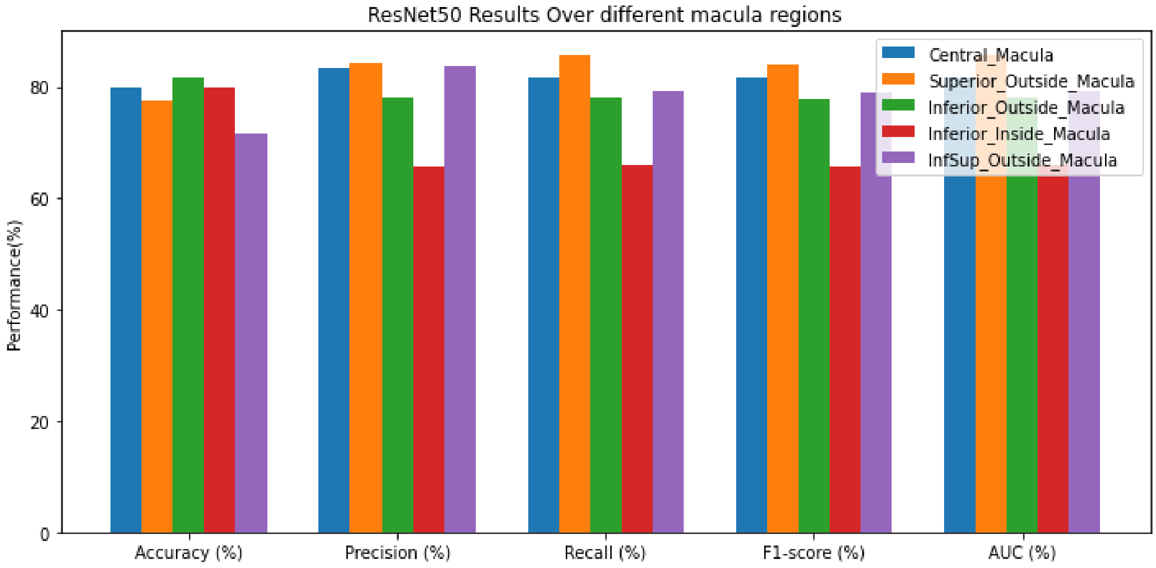


Figure 46: Bar plot showing evaluation metrics computed for the ResNet50 model applied on four different regions of the macula.

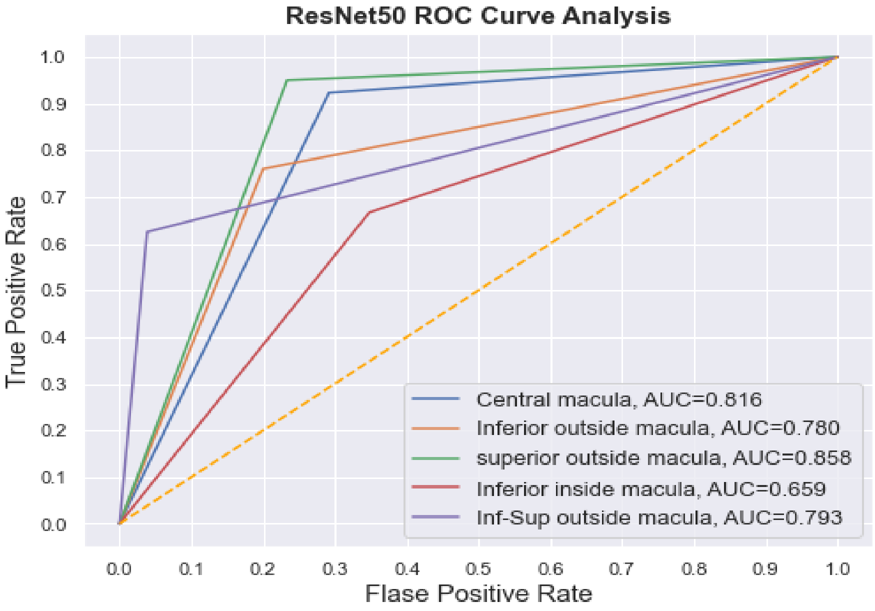


Figure 47: ROC curves and respective AUC values computed for the ResNet50 model applied on four different regions of the macula.

5.3 Discussion

5.3.1 Discussion on the Segmentation Results

The current study compared the performance of the popular state-of-the-art deep segmentation schemes namely PSPNet, and SegNet, as well as the recently proposed framework $\mathrm{RAG}-\mathrm{Net}_{v2}$. The segmentation performance is shown in Table 35 and Figure 48, where we can observe that

SegNet achieved the overall best μDC score of 89% followed by PSPnet. If we look at the performance of each model for extracting individual regions in terms of class-wise IoU, we can see that the best score for segmenting the RNFL, GC-IPL, rest retina, and the background is achieved by SegNet. The μIoU also shows the highest performance is achieved by SegNet. In terms of validation accuracy, the overall best performance is also achieved by SegNet at 97.89% as shown in Table 35 followed by PSPNet and $\mathrm{RAG-Net}_{v2}$. Also in terms of precision, F1-score, and Recall, SegNet achieved the best performance than the other two models. Overall, SegNet outperforms PSPnet and $\mathrm{RAG-Net}_{v2}$ with a larger margin in segmenting the RNFL, GC-IPL, and the rest-retina regions. Accordingly, the results from the SegNet model were used for further image classification.

Table 34: Overall segmentation model result.

Model	**Accuracy(%)**	**Recall (%)**	**F1 Score (%)**	**Precision (%)**	**Mean Dice Coefficient (%)**	**Mean-IoU (%)**	**'Class_Wise_IoU(%)**			
							Background	GC-IPL	RNFL	Rest-retina
RAGNet $_{v2}$	95.5	66.6	71	76.6	75.5	65.6	96.7	47.8	36.2	81.7
PSPnet	97.5	86.8	86	85.3	87.1	77.2	98.5	67.8	54	88.4
SegNet	97.9	87	87.5	88	89	81	98.8	75	60.2	90.1

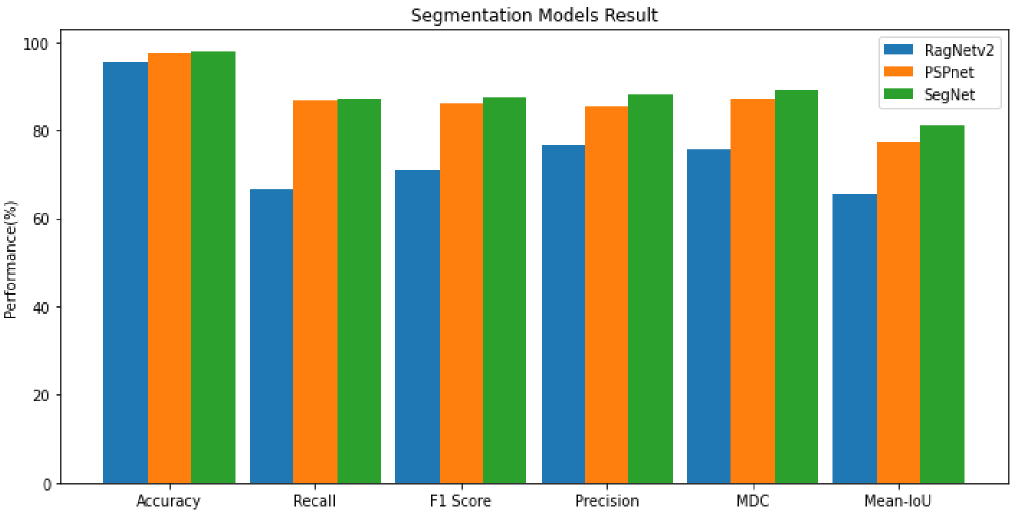


Figure 48: Bar plot showing evaluation metrics computed for the different segmentation models.

5.3.2 Discussion on the Classification Results

Three different classifiers (VGG16, VGG19, and ResNet50) were used for differentiating healthy and glaucomatous subjects. The performance of each classifier for screening glaucoma is measured through standard metrics, namely Precision, Recall, F1-score, AUC value and overall Accuracy. All the three models were trained and validated on images taken from four macula regions known to be vital for glaucoma detection. The performance of the models was compared by taking each model's results generated from each region separately. In one instance, a combination of information coming from two different regions of the macula was also considered. Overall, the VGG16 model offered the best classification performance than the other two models when it is applied on any of the four macular regions. The only exception is when the central macular region is taken where VGG19 performed slightly better than VGG16 in terms of overall accuracy while in terms of the other metrics (Precision, Recall, F1-score and AUC), the VGG16 was better.

Though the four macular regions are considered important for glaucoma detection, the proposed automated deep learning tool in this study showed that it is the superior region, particularly the outside superior region that is more important than the inferior macular region. The paired inferior-superior region also offered promising results. The central region offered the least overall accuracy. This is supported by studies that demonstrated that the superior and inferior areas of the optic nerve are most commonly affected by glaucoma [45]. Note that the deep models were also tested for their efficacy in classifying normal and glaucomatous subjects by pairing any two of the four macular regions (also a combination of all four regions). However, the results were not as satisfactory and those results are not reported in this thesis.

Table 35: Overall classification results obtained using the different deep learning models.

Regions	**Model**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**AUC (%)**
Central Macula	VGG16	84.3	88.24	88.00	87.98	88
	VGG19	85.6	86.2	86.4	85.9	86.39
	ResNet50	79.8	83.4	81.57	81.6	81.57
Superior Outside Macula	VGG16	94.3	93.3	91.7	91.8	91.7
	VGG19	85.6	83.51	80.50	81.08	80.5
	ResNet50	77.6	84.46	85.83	83.90	85.83
Inferior Outside Macula	VGG16	88%	88.3	87.8	87.9	88
	VGG19	85.7%	82.05	82.00	81.99	82
	ResNet50	81.8	78.04	78.00	77.9	78
Inferior Inside macula	VGG16	89.8%	87	87	87	87
	VGG19	82.03%	81.86	81.49	81.49	81.49
	ResNet50	79.8	65.87	65.94	65.88	65.94
Paired Inf-Sup Outside Macula	VGG16	90.7%	90.00	84.09	84.98	84.09
	VGG19	88.12	83.02	83.02	83.02	83.02
	ResNet50	71.79	83.64	79.33	79.17	79.32

CHAPTER SIX

CONCLUSION AND FUTURE WORKS

6.1 Conclusion

The main aim of this study was to detect glaucoma earlier using macula SD-OCT B-scans and Deep Convolutional Neural Networks. In this work, a fully automated system for the segmentation and classification of glaucoma from macula SD-OCT scans are presented. Unlike other frameworks that rely on cup-to-disc ratios and ONH SD-OCT scan for screening glaucoma, the proposed system analyzes the pathological changes related to the degeneration of RGCs through the RNFL, GC-IPL, and GCC profiles.

The study also assessed the most informative macula region for glaucoma detection. To achieve the objective, the study attempted to find the most suitable deep segmentation and classification models that could accurately segment relevant retina layers and classify the scans as glaucomatous and normal. A series of experiments was performed to find the most informative macula region and the optimum number of scans per region that result in best classification outputs. The experiments have shown the superiority of VGG16 in screening glaucoma cases as compared to the VGG19 and ResNet50 evidenced by different evaluation metrics. The VGG16 model achieved the highest performance on superior outside macula scans. Although two GCC extraction methods were used, using direct segmented images offered better results almost in every trained model than the masked images. The primary assumption when directly inputting the segmented images to the classifier was to allow it learns just the thickness information from extracted GCC images.

6.2 Limitations of the Study and Feature Works

The number of hospitals having OCT machine is very limited in Ethiopia and most of them use optic nerve head (ONH) scans for the diagnosis of glaucoma. This limited the size of the primary data that needed to be collected to develop and test the segmentation and classifications tools. Although different imaging modalities are available for use in glaucoma detection, the scope of the current research work was limited to those captured on SD-OCT machine. Due to the absence of publicly available data, the study only considered those acquired locally at the Myungsung Christian Medical Center (Korean Hospital) in Addis Ababa. Collecting data from different sites (hospitals) will make the data more representative and might improve the performance of the proposed deep learning segmentation and classification schemes. That is essential if this work has to be clinically applicable for GCC-aware glaucoma detection. Studying effects of patient demographic information such as age, race, and gender on the proposed glaucoma detection tool could also be further investigated. By preparing such normative data, it may be possible to enhance the model performance and also perform glaucoma grading.

REFERENCE

[1] K. S. Yadav, R. Rajpurohit, and S. Sharma, "Glaucoma: Current treatment and impact of advanced drug delivery systems," *Life Sci.*, vol. 221, no. February, pp. 362–376, 2019, doi: 10.1016/j.lfs.2019.02.029.

[2] OMS, "Global initiative for the elimination of avoidable blindness: action plan 2006-2011," 2007. [Online]. Available: https://apps.who.int/iris/handle/10665/43754

[3] A. Tanna *et al.*, "Glaucoma," in *Clinical Evaluation and Imaging of the Anterior Segment*, 2021, pp. 53–57.

[4] K. Allison, D. Patel, and O. Alabi, "Epidemiology of Glaucoma: The Past, Present, and Predictions for the Future," *Cureus*, vol. 12, no. 11, 2020, doi: 10.7759/cureus.11686.

[5] A. . Quigley, H.A. and Broman, "The number of people with glaucoma worldwide in 2010 and 2020," *Br. J. Ophthalmol.*, vol. 90(3), pp. 262–267, 2006, [Online]. Available: https://bjo.bmj.com/content/90/3/262.short

[6] C. C. Tham YC, Li X, Wong TY, Quigley HA, Aung T, "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, doi: 10.1016/j.

[7] S. Kyei, J. Aberor, F. Assiamah, and M. A. Kwarteng, "Optical coherence tomography indices in the diagnosis and discrimination of stages of primary open-angle glaucoma in an African population," *Int. Ophthalmol.*, vol. 41, no. 3, pp. 981–990, 2021, doi: 10.1007/s10792-020-01652-6.

[8] M. T. Tegegn, A. K. Assaye, and G. A. Mersha, "Proportion, causes and associated factors of blindness among adult patients attending tertiary eye care and training center in Ethiopia," *Clin. Optom.*, vol. 13, pp. 83–91, 2021, doi: 10.2147/OPTO.S295626.

[9] A. K. Schuster, C. Erb, E. M. Hoffmann, T. Dietlein, and N. Pfeiffer, "The diagnosis and treatment of glaucoma," *Dtsch. Arztebl. Int.*, vol. 117, no. 13, pp. 225–234, 2020, doi: 10.3238/arztebl.2020.0225.

[10] M. . Wu, Y., Szymanska, M., Hu, Y., Fazal, M.I., Jiang, N., Yetisen, A.K. and Cordeiro, "Measures of disease activity in glaucoma," *Biosens. Bioelectron.*, vol. 196, p. 113700, 2022.

[11] T. Khalil, M. U. Akram, S. Khalid, and A. Jameel, "Improved automated detection of glaucoma from fundus image using hybrid structural and textural features," *IET Image*

Process., vol. 11, no. 9, pp. 693–700, 2017, doi: 10.1049/iet-ipr.2016.0812.

[12] A. Kamalipour and S. Moghimi, “Macular optical coherence tomography imaging in glaucoma,” *J. Ophthalmic Vis. Res.*, vol. 16, no. 3, pp. 478–489, 2021, doi: 10.18502/jovr.v16i3.9442.

[13] J. Wu *et al.*, “GAMMA Challenge:Glaucoma grAding from Multi-Modality imAges,” pp. 1–43, 2022, [Online]. Available: http://arxiv.org/abs/2202.06511

[14] R. da R. R. & G. de O. R. Lucas M. Ceschini, Lucas M. Policarpo, “Aiding Glaucoma Diagnosis from the Automated Classification and Segmentation of Fundus Images,” *Brazilian Conf. Intell. Syst.*, vol. 13654, pp. 343–356, 2022, doi: https://doi.org/10.1007/978-3-031-21689-3_25.

[15] A. A. &Hira R. Hina Raja, Muhammad Usman Akram, Taimur Hassan, Aneeqa Ramzan, “Glaucoma Detection Using Optical Coherence Tomography Images: A Systematic Review of Clinical and Automated Studies,” *IETE J. Res.*, pp. 1–21, [Online]. Available: https://doi.org/10.1080/03772063.2022.2043783

[16] A. Akashi, A. Kanamori, M. Nakamura, M. Fujihara, Y. Yamada, and A. Negi, “The ability of macular parameters and circumpapillary retinal nerve fiber layer by three SD-OCT instruments to diagnose highly myopic glaucoma,” *Investig. Ophthalmol. Vis. Sci.*, vol. 54, no. 9, pp. 6025–6032, 2013, doi: 10.1167/iovs.13-12630.

[17] C. Zheng, T. V. Johnson, A. Garg, and M. V. Boland, “Artificial intelligence in glaucoma,” *Curr. Opin. Ophthalmol.*, vol. 30, no. 2, pp. 97–103, 2019, doi: 10.1097/ICU.0000000000000552.

[18] B. Prabhakar, R. K. Singh, and K. S. Yadav, “Artificial intelligence (AI) impacting diagnosis of glaucoma and understanding the regulatory aspects of AI-based software as medical device,” *Comput. Med. Imaging Graph.*, vol. 87, p. 101818, 2021, doi: 10.1016/j.compmedimag.2020.101818.

[19] H. Chen, Z., Wollstein, G., Schuman, J.S. and Ishikawa, “AI and Glaucoma. In Artificial Intelligence in Ophthalmology,” *Artif. Intell. Ophthalmol.*, pp. 113–125, 2021.

[20] G. Guidoboni *et al.*, “Precision medicine and glaucoma management: how mathematical modeling and artificial intelligence help in clinical practice,” *Expert Rev. Ophthalmol.*, vol. 17, no. 5, pp. 299–302, 2022, doi: 10.1080/17469899.2022.2130249.

[21] K. E. Kim, J. M. Kim, J. E. Song, C. Kee, J. C. Han, and S. H. Hyun, “Development and

validation of a deep learning system for diagnosing glaucoma using optical coherence tomography," *J. Clin. Med.*, vol. 9, no. 7, pp. 1–14, 2020, doi: 10.3390/jcm9072167.

[22] A. . Mangione, C.M., Barry, M.J., Nicholson, W.K., Cabana, M., Chelmow, D., Coker, T.R., Davis, E.M., Donahue, K.E., Epling, J.W., Jaén, C.R. and Krist, "Screening for primary open-angle glaucoma: US Preventive Services Task Force recommendation statement.," *JAMA Netw.*, vol. 327, no. 20, pp. 1992–1997, 2022, [Online]. Available: https://jamanetwork.com/journals/jama/article-abstract/2792609

[23] J. Chen and R. Kardon, "Avoiding clinical misinterpretation and artifacts of optical coherence tomography analysis of the optic nerve, retinal nerve fiber layer, and ganglion cell layere," *J Neuroophthalmol.*, vol. 36, no. 4, p. 417, 2016, doi: https://doi.org/10.1097/WNO.0000000000000422.

[24] R. Hood, D.C., Raza, A.S., de Moraes, C.G.V., Liebmann, J.M. and Ritch, "Glaucomatous damage of the macula.," *Prog. Retin. Eye Res.*, no. 32, pp. 1–21, 2013.

[25] A. . Khurana, "Anatomy and Physiology of Eye," in *Comprehensive Ophthalmology*, Jaypee Brothers Medical Publishers, 1999.

[26] C. W. Oyster, "Retina III: Regional Variation and Spatial Organization (Chapter 15)," *Hum. Eye Struct. Funct.*, pp. 649–700, 1999.

[27] R. A. Jmarchn., "Schematic diagram of the human eye." https://commons.wikimedia.org/wiki/File:Schematic_diagram_of_the_human_eye_en.svg

[28] J. . Grossniklaus, H.E., Geisert, E.E. and Nickerson, "Introduction to the Retina," *Prog. Mol. Biol. Transl. Sci.*, vol. 134, pp. 383–396, 2015.

[29] A. . Hildebrand, G.D. and Fielder, "Anatomy and physiology of the retina," in *In Pediatric retina*, Springer, Berlin, Heidelberg, 2011, pp. 39–65.

[30] C. . Gupta, M.P., Herzlich, A.A., Sauer, T. and Chan, "Retinal Anatomy and Pathology," *Retin. Pharmacother.*, vol. 55, pp. 7–17, 2016, [Online]. Available: https://doi.org/10.1159/000431128

[31] Masland RH, "The fundamental plan of the retina," *Nat Neurosci*, vol. 4, pp. 877–886, 2001.

[32] B. B. Wassle H, "Functional architecture of the mammalian retina," *Physiol*, vol. 71, pp. 447–480, 1991.

[33] Kolb H., "The Organization of the Retina and Visual System," in *Gross Anatomy of the Eye*, Salt Lake City (UT: University of Utah Health Sciences Center, 1995.

[34] “Schematic diagram of the human eye”, [Online]. Available: schem.jpeg (570×500) (utah.edu)

[35] P. J. Foster, R. Buhrmann, H. A. Quigley, and G. J. Johnson, “Prevalence Surveys,” pp. 238–243, 2002.

[36] N. Zhang, J. Wang, Y. Li, and B. Jiang, “Prevalence of primary open angle glaucoma in the last 20 years: a meta-analysis and systematic review,” *Sci. Rep.*, vol. 11, no. 1, pp. 1–12, 2021, doi: 10.1038/s41598-021-92971-w.

[37] David Ardaya, “7 Risk Factors for Glaucoma,” *California Optometric Association*. https://www.aplaceformom.com/caregiver-resources/articles/risk-factors-for-glaucoma

[38] A. R. Rudnicka, S. Mt.-Isa, C. G. Owen, D. G. Cook, and D. Ashby, “Variations in primary open-angle glaucoma prevalence by age, gender, and race: A Bayesian meta-analysis,” *Investig. Ophthalmol. Vis. Sci.*, vol. 47, no. 10, pp. 4254–4261, 2006, doi: 10.1167/iovs.06-0299.

[39] N. Halawa, O.A., Jin, Q., Pasquale, L.R., Kang, J.H., Lorch, A.C., Sobrin, L., Miller, J.W., Li, Y., Eslami, M., Wang, M. and Zebardast, “Race and Ethnicity Differences in Disease Severity and Visual Field Progression Among Glaucoma Patients,” *J. Ophthalmol.*, vol. 242, pp. 69–76, 2022, [Online]. Available: https://doi.org/10.1016/j.ajo.2022.05.023

[40] V. V. Kapetanakis, M. P. Y. Chan, P. J. Foster, D. G. Cook, C. G. Owen, and A. R. Rudnicka, “Global variations and time trends in the prevalence of primary open angle glaucoma (POAG): A systematic review and meta-analysis,” *Br. J. Ophthalmol.*, vol. 100, no. 1, pp. 86–93, 2016, doi: 10.1136/bjophthalmol-2015-307223.

[41] J. C. Downs and C. A. Girkin, “Lamina cribrosa in glaucoma,” *Curr. Opin. Ophthalmol.*, vol. 28, no. 2, pp. 113–119, 2017, doi: 10.1097/ICU.0000000000000354.

[42] Edward DeHoog and James Schwiegerling, “Fundus camera systems: a comparative analysis,” *Appl. Opt.*, vol. 48, no. 2, pp. 221–228, 2009, [Online]. Available: https://doi.org/10.1364/AO.48.000221

[43] D. G. Jay S. Duker, Nadia K Waheed, *Handbook of Retinal OCT*. :Optical Coherence Tomography E-Book, Elsevier Health Sciences. [Online]. Available: https://books.google.com/books?hl=en&lr=&id=iQU7EAAAQBAJ&oi=fnd&pg=PP1&dq=OCT&ots=_zYb-NxIau&sig=tR5pcnek0E8bZ-Iv4hNTczwbae0

[44] M. S. Kim *et al.*, “Effect of Weiss ring on peripapillary retinal nerve fiber layer thickness

measurements using SD-OCT,” *Sci. Rep.*, vol. 12, no. 1, pp. 1–8, 2022, doi: 10.1038/s41598-022-22094-3.

[45] F. A. Medeiros, L. M. Zangwill, C. Bowd, R. M. Vessani, R. Susanna, and R. N. Weinreb, “Evaluation of retinal nerve fiber layer, optic nerve head, and macular thickness measurements for glaucoma detection using optical coherence tomography,” *Am. J. Ophthalmol.*, vol. 139, no. 1, pp. 44–55, 2005, doi: 10.1016/j.ajo.2004.08.069.

[46] C. F. Chauhan, B.C. and Burgoyne, “From clinical examination of the optic disc to clinical assessment of the optic nerve head: a paradigm change,” *Am. J. Ophthalmol.*, vol. 156, no. 2, pp. 218–227, 2013, doi: 10.1016/j.ajo.2013.04.016.

[47] Alyssa Schroer, “Artificial Intelligence,” *Built In National*, 2022. https://builtin.com/artificial-intelligence

[48] B. Ekaba, “Building machine learning and deep learning models on google cloud platform,” in *Building machine learning and deep learning models on google cloud platform*, New York: Apress Media, LLC, 2019, pp. 423–441.

[49] “The difference between artificial intelligence, machine learning, and deep learning.” https://www.bing.com/images/search?view=detailV2&ccid=mwv8tqYW&id=57014A8D69A0E1A22F29B1BA7CA85CEB1291353A&thid=OIP.mwv8tqYWSx4O9_p7PUZ_EgHaGZ&mediaurl=https%3A%2F%2Fblog.digitalogy.co%2Fwp-content%2Fuploads%2F2019%2F03%2FThe-Difference-between-Artificial

[50] “Convolutional Neural Network.” https://www.sciencedirect.com/science/article/pii/B9780128230145000089

[51] IBM Cloud Education, “Convolutional Neural Networks,” *IBM Cloud Learn Hub*, 2020. https://www.ibm.com/cloud/learn/convolutional-neural-networks

[52] “Diagram of Convolutional Neural Networks”, [Online]. Available: entropy-19-00242-g001.png (3402×1379) (mdpi.com)

[53] “2D convolution with filter size 3x3.” https://th.bing.com/th/id/OIP.F96jmLp1Ljx4UjtMmhlLmQHaFE?pid=ImgDet&rs=1

[54] “Dilated Convolution.” https://iq.opengenus.org/dilated-convolution/#:~:text=Dilated convolution%2C also known as Atrous Convolution or,which in turn skips some of the points

[55] “The Dilated or Atrous Convolution”, [Online]. Available:

https://iq.opengenus.org/content/images/2021/08/dilated-convolution.gif

[56] "CNN | Introduction to Pooling Layer," 2022, [Online]. Available: https://www.geeksforgeeks.org/cnn-introduction-to-pooling-layer/

[57] "Types of pooling", [Online]. Available: https://editor.analyticsvidhya.com/uploads/597371_KQIEqhxzICU7thjaQBfPBQ.png

[58] "Example of flattening before feeding to the FC", [Online]. Available: https://sds-platform-private.s3-us-east-2.amazonaws.com/uploads/73_blog_image_1.png

[59] "Illustration of a fully connected layer", [Online]. Available: https://builtin.com/sites/www.builtin.com/files/styles/ckeditor_optimize/public/inline-images/3_fully-connected-layer_0.jpg

[60] "Softmax function", [Online]. Available: https://en.wikipedia.org/wiki/Softmax_function

[61] Thomas Wood, "Softmax Function," *DeepAI*, [Online]. Available: https://deepai.org/machine-learning-glossary-and-terms/softmax-layer

[62] "How ReLU works in convolutional neural network," *Knowl. Transf.*, 2022, [Online]. Available: https://androidkt.com/how-relu-works-in-convolutional-neural-network/

[63] O. P. Dilmurod Khasanov , Ma'ruf Tojiyev, "Gradient descent in machine learning," *2021 Int. Conf. Inf. Sci. Commun. Technol.*, pp. 1–3, 2021, doi: 10.1109/ICISCT52966.2021.9670169.

[64] S. Ruder, "An overview of gradient descent optimization algorithms," pp. 1–14, 2016, [Online]. Available: http://arxiv.org/abs/1609.04747

[65] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986, doi: 10.1038/323533a0.

[66] L. Bottou, "Stochastic gradient descent tricks," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7700 LECTU, no. 1, pp. 421–436, 2012, doi: 10.1007/978-3-642-35289-8_25.

[67] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.

[68] "Batch normalization", [Online]. Available: https://en.wikipedia.org/wiki/Batch_normalization

[69] NVS Yashwanth, "What is batch normalization?," *Towar. Data Sci.*, 2020, [Online]. Available: https://towardsdatascience.com/what-is-batch-normalization-46058b4f583

[70] Shipra Saxena, “Introduction to Batch Normalization,” *Anal. Vidhya*, 2021, [Online]. Available: https://www.analyticsvidhya.com/blog/2021/03/introduction-to-batch-normalization/

[71] Kizito Nyuytiymbiy, “Parameters and Hyperparameters in Machine Learning and Deep Learning,” *Towar. Data Sci.*, 2020, [Online]. Available: https://towardsdatascience.com/parameters-and-hyperparameters-aa609601a9ac

[72] M. Claesen and B. De Moor, “Hyperparameter Search in Machine Learning,” pp. 10–14, 2015, [Online]. Available: http://arxiv.org/abs/1502.02127

[73] Jason Brownlee, “A Gentle Introduction to Transfer Learning for Deep Learning,” *Deep Learn. Comput. Vis.*, 2017, [Online]. Available: https://machinelearningmastery.com/transfer-learning-for-deep-learning/

[74] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–14, 2015.

[75] Great Learning Team, “Introduction to VGG16 | What is VGG16?,” *Gt. Learn.*, 2022, [Online]. Available: https://www.mygreatlearning.com/blog/introduction-to-vgg16/

[76] “The architecture of the VGG-16 model”, [Online]. Available: https://www.researchgate.net/profile/Bibo_Shi/publication/323440752/figure/download/fig1/AS:739814685032448@1553396974148/The-architecture-of-VGG-16-model-To-represent-different-depth-levels-convolutional.jpg

[77] “The architecture of the VGG-19 model”, [Online]. Available: https://www.researchgate.net/profile/Mohammed_Y_Kamil/publication/344398328/figure/fig1/AS:940416480321539@1601224169795/Modified-VGG-19-model-architecture.ppm

[78] A. Abrol, M. Bhattarai, A. Fedorov, Y. Du, S. Plis, and V. Calhoun, “Deep residual learning for neuroimaging: An application to predict progression to Alzheimer’s disease,” *J. Neurosci. Methods*, vol. 339, 2020, doi: 10.1016/j.jneumeth.2020.108701.

[79] “The architecture of ResNet50 model”, [Online]. Available: https://www.researchgate.net/profile/Mohammed_Y_Kamil/publication/344398328/figure/fig1/AS:940416480321539@1601224169795/Modified-VGG-19-model-architecture.ppm

[80] C. Han, Y. Duan, X. Tao, and J. Lu, “Dense Convolutional Networks for Semantic Segmentation,” *IEEE Access*, vol. 7, pp. 43369–43382, 2019, doi: 10.1109/ACCESS.2019.2908685.

[81] S. Patel, “Deep learning models for image segmentation,” *Proc. 2021 8th Int. Conf. Comput. Sustain. Glob. Dev. INDIACom 2021*, no. August, pp. 149–154, 2021, doi: 10.1109/INDIACom51348.2021.00027.

[82] “The FCN end-to-end dense prediction pipeline”, [Online]. Available: https://miro.medium.com/max/632/1*wRkj6lsQ5ckExB5BoYkrZg.png

[83] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, 2017, doi: 10.1109/TPAMI.2016.2644615.

[84] “SegNet Neural Network”, [Online]. Available: http://deepmachinelearningai.com/wp-content/uploads/2020/02/SegNet-Neural-Network.png

[85] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid Scene Parsing Network Hengshuang,” *Photogramm. Rec.*, vol. 37, no. 178, pp. 248–249, 2022, doi: 10.1111/phor.1_12412.

[86] “PSPnet Neural Network”, [Online]. Available: https://meetshah.dev/images/blog/ss/pspnet.png

[87] Taimur Hassan; Muhammad Usman Akram; Naoufel Werghi; Muhammad Noman Nazir, “RAG-FW: A Hybrid Convolutional Framework for the Automated Extraction of Retinal Lesions and Lesion-Influenced Grading of Human Retinal Pathology,” *IEEE J. Biomed. Heal. Informatics*, vol. 25, no. 1, pp. 108–120, 2021, doi: 10.1109/JBHI.2020.2982914.

[88] H. Raja, T. Hassan, and M. U. Akram, “Clinically Verified Hybrid Deep Learning System for Retinal Ganglion Cells Aware Grading of Glaucomatous Progression,” *IEEE Trans. Biomed. Eng.*, vol. 68, no. 7, pp. 2140–2151, 2021.

[89] T. Shehryar *et al.*, “Improved automated detection of glaucoma by correlating fundus and SD-OCT image analysis,” *Int. J. Imaging Syst. Technol.*, vol. 30, no. 4, pp. 1046–1065, 2020, doi: 10.1002/ima.22413.

[90] J. L. Xiangyu Chen, Yanwu Xu, Damon Wing Kee Wong, Tien Yin Wong, “Glaucoma detection based on deep convolutional neural network,” *2015 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 715–718, 2015, doi: 10.1109/EMBC.2015.7318462.

[91] J. Lee, Y. K. Kim, K. H. Park, and J. W. Jeoung, “Diagnosing Glaucoma with Spectral-

domain Optical Coherence Tomography Using Deep Learning Classifier," *J. Glaucoma*, vol. 29, no. 4, pp. 287–294, 2020, doi: 10.1097/IJG.0000000000001458.

[92] S. Maetschke, B. Antony, H. Ishikawa, G. Wollstein, J. Schuman, and R. Garnavi, "A feature agnostic approach for glaucoma detection in OCT volumes," *PLoS One*, vol. 14, no. 7, pp. 1–12, 2019, doi: 10.1371/journal.pone.0219126.

[93] X. Wang *et al.*, "Towards multi-center glaucoma OCT image screening with semi-supervised joint structure and function multi-task learning," *Med. Image Anal.*, vol. 63, 2020, doi: 10.1016/j.media.2020.101695.

[94] A. R. Shelke Roscoe, J. A. , Morrow, G. R. , Colman, L. K. , Banerjee, T. K. , & Kirshner, J. J., "Thickness Profiles of Retinal Layers by Optical Coherence Tomography Image Segmentation," *Bone*, vol. 23, no. 1, pp. 1–7, 2008, doi: 10.1016/j.ajo.2008.06.010.Thickness.

[95] M. D. Zeiler, "ADADELTA: An Adaptive Learning Rate Method," 2012, [Online]. Available: http://arxiv.org/abs/1212.5701

[96] "A schematic model of RNFL projections and glaucomatous RGC+ and RNFL damage", [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/instance/3529818/bin/nihms-419007-f0011.jpg

[97] Ekin Tiu, "Metrics to Evaluate your Semantic Segmentation Model," *Towar. Data Sci.*, 2019, [Online]. Available: https://towardsdatascience.com/metrics-to-evaluate-your-semantic-segmentation-model-6bcb99639aa2

[98] Johannes S. Fischer, "Intersection over Union (IoU)," 2021.

[99] Sarang Narkhede, "Understanding AUC - ROC Curve," *Towards Data Science*, 2018. https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5