ARAB ACADEMY FOR SCIENCE, TECHNOLOGY AND MARITIME TRANSPORT

College of Engineering and Technology

Computer Engineering Department

A Hybrid Computer Aided Medical Diagnosis System Integrating Machine Learning and Automatic Region of Interest Segmentation: A Case-Study on Diagnosis of Optical Coherence Tomography Retinal Disorders

By

Mohamed Ramzy Ibrahim

B.Sc. Computer Engineering AASTMT, 2018

A thesis submitted to AASTMT in partial

Fulfillment of the requirements for the award of the degree of

MASTER'S OF SCIENCE

in

COMPUTER ENGINEERING

Supervisors

Prof. Dr. Sherin Moustafa Youssef

Computer Engineering Department

College of Engineering and Technology

Dr. Karma Mohamed Fathalla

Computer Engineering Department

College of Engineering and Technology

Arab Academy for Science, Technology and Maritime Transport

Alexandria

2020

بسم الله الرحمن الرحيم

DECLARATION

I certify that all the material in this thesis that is not my own work has been identified, and that no material is included for which a degree has previously been conferred on me.

The contents of this thesis reflect my own personal views and are not necessarily endorsed by the University.

Name Mohamed Ramzy Ibrahim

Signature ………………………………………………………………………………….

Date ....

We certify that we have read the present work and that in our opinion it is fully adequate in scope and quality as a thesis towards the partial fulfillment of the Master's Degree requirements in

Computer Engineering Department

Mohamed Ramzy Ibrahim

From
College of Engineering and Technology
AASTMT
Date: 13/12/2020

Supervisors:

Name:	**Prof. Sherin Moustafa Youssef**
Position:	Professor of Computer Engineering, College of Engineering and Technology, Arab Academy for Science, Technology and Maritime Transport, Alexandria
Signature:	……………………………………………………………………
Name:	**Dr. Karma Mohamed Fathalla**
Position:	Assistant Professor of Computer Engineering, College of Engineering and Technology, Arab Academy for Science, Technology and Maritime Transport, Alexandria
Signature:	……………………………………………………………………

Examiners:

Name:	**Prof. Saleh Abdel Shacour El Shehaby**
Position:	Professor of Computer Engineering, Institute of Graduate Studies and Research, Alexandria University
Signature:	……………………………………………………………………
Name:	**Prof. Aliaa Abdel Halim Abdel Razk Youssif**
Position:	Professor of Computer Science and Information Technology, College of Computing and Information Technology, Arab Academy for Science, Technology and Maritime Transport, Smart Village
Signature:	……………………………………………………………………
Name:	**Prof. Sherin Moustafa Youssef**
Position:	Professor of Computer Engineering, College of Engineering and Technology, Arab Academy for Science, Technology and Maritime Transport, Alexandria
Signature:	……………………………………………………………………

ACKNOWLEDGMENT

I respectfully grab this opportunity to acknowledge many people who deserve special mentions for their varied contributions in assorted ways that helps me during my research and the making of this thesis.

First, I give thanks to GOD for protection and power to finish this work.

I would like to express my sincere gratitude to my supervisor Prof. Dr. Sherin Youssef for the continuous support of my thesis study and research, for her patience, motivation, enthusiasm, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my master study.

My sincere thanks and deepest gratitude also go to my second supervisor Dr. Karma Fathalla for her support, help, interest, and valuable hints. Her insightful suggestions, comments and encouragement helped me in the time of writing and research of this thesis.

Besides my supervisors, I would like to thank my Mom for her everlasting support and devotions that encouraged me to pursue my academic career. I'm grateful for her patience, attention and continuous understanding that helped me move forward and chase my dreams no matter how impossible they seem.

I also acknowledge my family for showing faith in me and for their selfless love, care and dedicated efforts which contributed a lot for completion of my thesis. Thanks for always being there for me encouraging me to be better.

I would like to thank Dr Eman Elbanna for her time and efforts to support me and my work.

I consider myself the luckiest in the world to have such a lovely and caring family, standing beside me with their love and unconditional support.

I wish to thank all my close friends who I considered my forever support system for always being there whenever I needed them and encouraging me to finish this work and understanding how important it is for me.

I wish to thank all the staff of the computer engineering department for their support and their tutoring that helped me a lot through my masters.

DEDICATION

In memory of my father

To my mother and my family
With love and eternal appreciation

ABSTRACT

Optical Coherence Tomography (OCT) is a noninvasive recent imaging technique that has been increasingly used to diagnose and manage a variety of retinal diseases and glaucoma. OCT examinations can aid in the detection of many retina disorders in early stages that could not be detected in traditional fundography retinal images. Also, it can be used to identify and quantify retinal thickness changes in response to therapy.

Retinal disorders are often diagnosed and treated by an ophthalmologist. However, to accurately assess a retinal disease, ophthalmologist would need time consuming qualitative and quantitative analysis of the disease. Moreover, the disproportionate increase in workload relative to workforce leads to overloading medical experts resulting in missed critical cases. Also, there is a criticality for early detection of retinal disorders for better prognosis and to avoid the occurrence of complications that may lead to vision loss. For the previous reasons, there should be an automated computer aided diagnosis (CAD) system to assist ophthalmologists for accurate early diagnosis of different retinal disorders.

In this thesis, a new hybrid computer-aided OCT diagnostic system (HyCAD) is proposed for classification of retinal disorders: Diabetic Macular Edema (DME), Choroidal Neovascularization (CNV) and drusen disorders, while separating them from Normal retinal structure using OCT images. The proposed HyCAD hybrid learning system assimilates a range of techniques including RoI localization based on deep learning, hybrid feature extraction, followed by classification and diagnosis. An effective feature fusion phase has been introduced for combining the OCT image features, extracted by Deep Convolutional Neural Network (CNN), with the features extracted from the RoI segmentation phase. This fused feature set is used to predict multiclass OCT retina disorders. The proposed segmentation phase of retinal RoI regions adds substantial contribution as it draws attention to the most significant areas that are candidate for diagnosis. A new modified deep learning architecture (Norm-VGG16) is introduced integrating a kernel regularizer. Norm-VGG16 is trained from scratch on a large benchmark dataset and used in RoI localization and segmentation.

Various experiments are carried out to validate the effective performance of the proposed system on Large Dataset of Labeled Optical Coherence Tomography (OCT) v3 benchmark dataset compared with others in literature. The experimental results show that the proposed model achieves relatively high-performance in terms of accuracy, sensitivity and specificity. Experimental evaluation showed outstanding performance compared to others in literature, the proposed (HyCAD) model achieves a high classification sensitivity to urgent cases. The HyCAD models' results are compared to the results of its backbone CNN (Norm-VGG16) and other famous deep learning architectures. It achieved a significant increase in accuracy, sensitivity and specificity. The experimental results reflect that the fusion phase can effectively improve the identification ratio of the urgent patients' diagnostic images.

In this thesis, HyCAD a highly performing general computer aided diagnosis system architecture is introduced. The proposed architecture can be trained and applied on similar problems, since no underlying specific assumptions were made that would hinder its generalization.

LIST OF PUBLICATIONS DERIVED FROM THE THESIS

M. Ibrahim, K. Fathalla, and S. Youssef, "HyCAD-OCT: A Hybrid Computer-Aided Diagnosis of Retinopathy by Optical Coherence Tomography Integrating Machine Learning and Feature Maps Localization," Applied Sciences, vol. 10, pp. 4716, 07/08, 2020.

M. Ibrahim, S. Youssef and K. Fathalla, "Abnormality Detection & Intelligent Severity Assessment of Human Chest Computed Tomography Scans using Deep Learning: A Case Study on SARS-COV-2 Assessment" Journal of Ambient Intelligence and Humanized Computing, Springer, 2020. (under publication)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

AMD	Age-related macular degeneration
ANN	Artificial Neural Network
AE	Auto-Encoders
BM3D	Block Matching 3D filter
BM	Bruch's Membrane
CC	Choriocapillaris
CNV	Choroidal Neovascularization
CS	Choroidal Stroma
CT	Computed Tomography
CAD	Computer Aided Diagnosis
CRD	Cone-Rod Dystrophy
CNN	Convolution Neural Network
DBN	Deep Belief Network
DME	Diabetic Macular Edema
ELM	External Limiting Membrane
FD-OCT	Fourier domain OCT
GCL	Ganglion Cell Layer
GAP	Global Average Pooling
Grad-CAM	Gradient Computed Activation Map
HOG	Histogram of Oriented Gradients
IPL	Inner Plexiform Layer
ILM	Internal limiting Membrane
LSTM	Long-Short Term Memory
MRI	Magnetic Resonance Imaging
OCT	Optical Coherence Tomography
ONL	Outer Nuclear Layer
OPL	Outer Plexiform Layer
PR	Photoreceptor Layer
ReLU	Rectified Linear Units

RNN	Recurrent Neural Network
RoI	Region of Interest
ResNet	Residual Network
RBM	Restricted Boltzmann Machine
RNFL	Retinal Nerve Fiber Layer
RPE	Retinal Pigment Epithelium
SD-OCT	Spectral Domain Optical Coherence Tomography
SVM	Support Vector Machine
TD-OCT	Time Domain Optical Coherence Tomography
VGG	Visual Graphics Group
WHO	World Health Organization

Chapter One

Introduction

1 INTRODUCTION

In this chapter, an overview of Computer Aided Diagnosis (CAD) systems and the availability of different retinal imaging techniques is presented. Retinal disorders will be introduced in this chapter as well. Moreover, the problem statement, the motivation and the main objectives of the thesis will be provided. A brief description of thesis layout is given at the end of this chapter.

1.1 Overview

Computer Aided Diagnosis (CAD) systems were introduced in the 1990s [1]. It could assist medical experts in different number of ways, ranging from providing quantified image metrics to calculated probabilities of various diagnoses. The integration of machine learning and deep learning in development of an automatic CAD systems leads to solving different problems in medical imaging fields. CAD systems can be used in identification of regions of interest (ROIs), which could lead to significant time-savings for medical experts [1]. CAD systems help in detecting different diseases and disorders in various medical images modalities. The development of multimodal medical imaging techniques in radiology lead to accurate explanations of different complementary structures and function information of human bodies. Medical imaging modalities such as X-rays, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Optical Coherence Tomography (OCT), etc… play an important role in early detection and diagnosis of numerous diseases [2]. It provides direct visualization to see through the human bodies and produce and informative medical images to the minute anatomical changes and biological processes that helps in early detection of diseases. The two major advancements in medical images which assist both scientists and medical physicians are the improvement in the depth and clarity at which they can view tissues. Therefore, The ultrasound, CT, and MRI can penetrate inside the body with high penetration depth; however, they do not have sufficient resolution to capture cellular detail [3]. Electron microscopy can pick up extremely fine details; however, it is not able to view living samples within the body [3]. Recently, OCT has attracted much attention in many clinical and basic research fields due to its high sensitivity for noninvasive high-resolution imaging at cellular level. The diagnostic capability of OCT has revolutionized many fields such as ophthalmology, dermatology, cardiology, etc. In terms of other medical imaging devices, OCT is the one best offered currently [3]. OCT is becoming rapidly an important biomedical tool for imaging tissues and engineered tissues. Figure 1-1 shows the comparison of OCT with presently existing clinical imaging modalities.

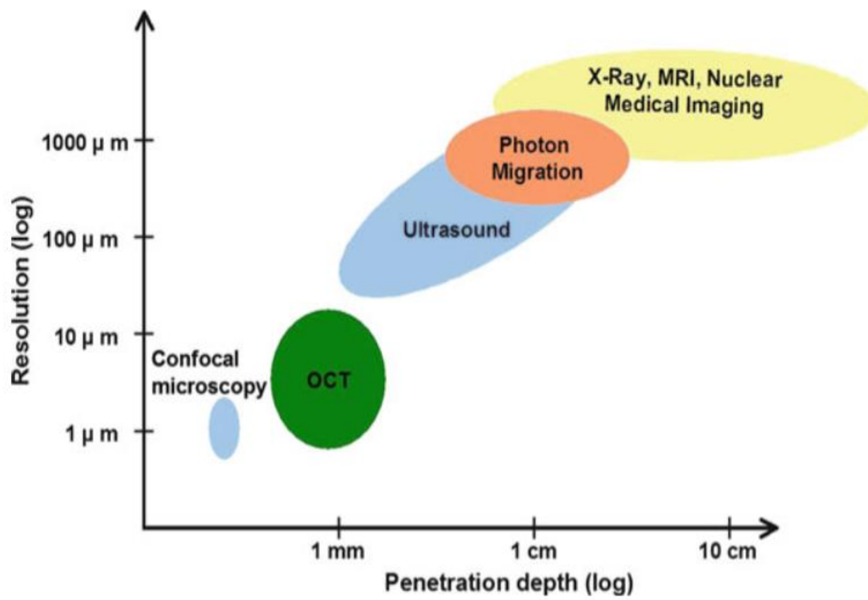


Figure 1-1. Comparison of different clinical imaging modalities in terms of their resolution and penetration depth

Medical imaging usually requires experienced medical doctors to best interpret the information revealed in the images [2]. However, because of various subjective factors as well as limited analysis time and tools, it is common that early detection of different diseases could be hard and subjective to medical experts experiences [2]. Recently, developing a computer-aided diagnosis (CAD) has become a major research field, applied widely in the detection and differential diagnosis of many types of abnormalities in medical images [4]. Recently, CAD has become a part of the routine clinical work for detection of different diseases from medical images.

1.2 Retinal disorders case study

According to the World Health Organization (WHO), at least 2.2 billion people all over the world suffer from vision impairment or blindness [5]. At least 1 billion of them have a vision impairment that could have been prevented or has yet to be addressed. Retinal diseases are from the most common causes of losing eyesight at an early age [5]. Most of these diseases affects a thin layer of tissue inside the back wall of the eye which is called retina. Optical Coherence Tomography (OCT) imaging is a type of optical biopsy that is used to view and capture small changes that occurs to the retina. It generates cross sectional 3D images by measuring the echo time delay or the magnitude of back reflected light [6]. Compared to traditional regular retinal fundus examinations, OCT has major advantages in identifying the presence of various ocular pathologies. A wide range of macular diseases such as Macular Edema, Choroidal Neovascularization, Macular Hole, Pigment Epithelium Detachment and Central Serous Retinopathy can be detected by OCT [7]. OCT examinations can aid the detection of the previously mentioned disorders in an early stage.

1.3 Problem Statment

The main problem with visually diagnosing medical images is that diagnosis is restricted to experts as it is based on their experiences. CAD models can automatically learn the patterns differences and provide efficient and fast diagnosis of abnormalities with high accuracy. These models can be used as a supportive decision system to the physicians in the early detection of edemas and abnormalities. The early detection of tumors and abnormalities can accelerate in serving the severe cases. Moreover, it is important to provide a framework that is highly performing, understandable and has a generalization ability to meet the requirements of medical applications.

1.4 Motivation

First, Most of CAD systems are developed using earlier fundography that was a widely used technique to detect retinal diseases. It can detect retinal disorders but it’s subjective and insensitive to small retinal thickness changes, macular breaks and Retinal Pigment Epithelium

(RPE) elevation. Major symptoms of Macular Edema do not appear in early stages even the patient himself does not know about the disease at that time[7]. On the other hand, cystic and sub retinal swellings are overt in early stages in OCT imaging, while they remain undetectable in retinal fundus images [7]. Moreover, it can also give an objective evaluation of macular and ocular pathology as it shows a cross sectional region of macula and optic disc [7]. In addition, OCT is noninvasive and quantifiable without pupil dilation. Hence, OCT scanning leads to a breakthrough in the screening, diagnosis and assessment of the necessity and efficacy of treatment for particular diseases.

Second, a major limitation in existing segmentation techniques of most of the deployed retinal disorders CAD systems is that they need the intervention of medical experts in the segmentation of retinal layers or in retina flattening before detecting different diseases. Finally, different CAD models were proposed in literature to detect abnormalities in OCT images, but they suffered from

- Focusing on detecting a specific retinal disorder.
- Lack of understandability (interpretation).

Therefore, our motive is to develop a model that automatically detects different retinal disorders in early stages from OCT images that overcomes the limitations of current systems. Moreover, to build a framework that is highly performing, understandable and has a generalization ability to meet the requirements of medical applications (not only retinal disorders).

1.5 Challenges

First challenge imposed by the images acquired from OCT imaging technique is the noise accompanied with the cross-sections of SD-OCT. The noise generated from the OCT capturing devices must be filtered for correct segmentation of retinal layers.

The second challenge is the segmentation of retinal layers. Most of techniques depends on experts manual flattening and segmentation. Medical knowledge was needed to correctly detect retinal layers and/or changes in layers' thickness and/or assist in retinal flattening. Such dependence on experts' availability gives rise to another concern, which is having poor generalization to other domains and puts a heavy load of work on medical systems.

The final challenge is that some of the retinal disorders need an immediate medical intervention to prevent any further complication that may lead to blindness. Accordingly, these retinal disorders need to be detected accurately and marked as critical at their early stages to prevent further complications.

1.6 Objective

The objective of the thesis is to design a robust computer-aided diagnostic (CAD) framework capable of diagnosing different diseases from different modalities of medical images. The framework shall provide high performance, which would be customizable according to the specific problem under study. The proposed CAD system shall provide a segmentation technique to segment the Region of Interest (RoI) for an efficient and accurate classification which will bring about a change in the roles of radiologists and other clinicians alike.

- Radiologists will be able to spend less time screening images and concentrate on decision-making.
- Non-radiologist physicians will use CAD system to have digital assistance to interpret medical images, making them less reliant on hospital radiology departments.

The CAD model shall present partially explainable results and can be generalized to meet requirements of different medical applications. Finally, the proposed CAD model needs to have a competitive performance to state-of-art and different deep learning architectures to prove its accuracy and robustness in classification of different diseases.

1.7 Contibution

The contribution of the paper can be summarized as

- A hybrid computer aided diagnosis (HyCAD) system used for retinal disorders detection from OCT images.
- HyCAD integrates automatic features extraction of deep learning and hand-crafted features to attain high accuracy classifications and provide relative insight into the classification decision.

- A modified deep learning architecture is proposed and used in automatic detection of Region of Interest (RoI). It achieved an outstanding training time performance (compared to other CNNs).
- HyCAD provides automated RoI localization using Grad-CAM which helps advance the diagnosis process.

The proposed models is validated on Large Dataset of Labeled Optical Coherence Tomography (OCT) v3 [8] benchmark for multi class classification of different retinal disorders.

1.8 Thesis Layout

The thesis is organized as follows:

Chapter 2 provides the medical background about Optical Coherence Tomography (OCT) Principles and different retinal diseases.

Chapter 3 illustrates some of work presented in literature and surveys the common steps used in segmentation of retinal layers and classification of retinal disorders from OCT images. Moreover, it explains the concept of deep learning and explanation of different machine learning-based classifiers.

Chapter 4 introduces the proposed model. It gives a detailed description of the methods used in the proposed model. It contains an overlook on image preprocessing, automatic Region of Interest (RoI) segmentation, a new modified deep learning architecture, handcrafted feature extraction and classification.

Chapter 5 describes all the experiments conducted to validate the proposed model. It also illustrates the performance indicators used in evaluating the system’s performance and the experimental results. It provides also a comparison with state-of-the-art studies.

Chapter 6 illustrates the conclusion and future work.

Chapter two

Medical Background

2 MEDICAL BACKGROUND

Medical imaging produces visual pictures of different anatomical structures inside the body. It is used in clinical examination and subsequent medical interference. Medical Imaging reveals internal structures hid by the skin and bones, it is used as a diagnostic tool for several diseases and has a vital role in monitoring treatment effects and predicting outcomes [9]. Technological advances in medical imaging increase the sensitivity of different techniques such as MRI, CT, Ultrasound, etc… It lead to high quality images that increase the sensitivity of detecting abnormalities leading to early diagnosis of different diseases and saving thousands of lives [9]. Radiologists have been involved in these technological developments in different medical images and have been responsible for much of the evaluation of the strengths and weaknesses of different investigations [9]. In this chapter, a focus is given on a recent retinal imaging technique called Optical Coherence Tomography (OCT). In addition, a review is given on different retinal disorders that cause major problems in humans eyes reaching blindness.

There are different retinal disorders that could affect the vision sight and could cause more serious complications leading to vision loss. Different imaging techniques were used to assess the retinal layers to try to early detect these disorders to assist physicians to have a correct treatment procedure. Earlier, Fundography imaging was the widely used technique to detect these diseases. It can detect retinal disorders but it’s subjective and insensitive to small retinal thickness, macular breaks and Retinal Pigment Epithelium (RPE) elevation. Major symptoms of a disease called Macular Edema do not appear in early stages even the patient himself does not know about the disease at that time [7]. In OCT, cystic and sub retinal swellings are overt in early stages, while they remain undetectable in retinal fundus images [7]. Moreover, it can also give an objective evaluation of macular and ocular pathology as it shows a cross sectional region of macula and optic disc [7]. In addition, OCT is noninvasive and quantifiable without pupil dilation. Hence, OCT scanning leads to a breakthrough in the screening, diagnosis and assessment of the necessity and efficacy of treatment for diseases. Several properties of OCT are sated as follow [10]:

- OCT captures image with axial resolutions of 1 to 15 µm. This resolution allows tiny morphology and some cellular features to be resolved unlike other imaging techniques.

- Imaging can be performed in situ, without the need to excise a specimen. This enables imaging of structures in which biopsy would be hazardous or impossible.
- Imaging can be performed in real time, without the need to process a specimen as in conventional biopsy and histopathology. This allows pathology to be monitored on screen and stored on high-resolution video tape. Real-time imaging can enable real-time diagnosis, and coupling this information with surgery, it can enable surgical guidance.
- Finally, OCT is compact and portable, an important consideration for a clinically viable device.

2.1 OCT imaging principles

OCT performs high-resolution of cross-sectional imaging of the internal microstructure in retinal layers by measuring echoes of backscattered light [10]. OCT imaging technique is always compared to ultrasound imaging technique because of the similar working principles. Both medical imaging techniques direct waves to the tissue under examination, where the waves echo off the tissue structure [11]. The reflected waves are analyzed and their delay is used to measure the depth in which the reflection occurred [11].

OCT imaging technology mainly consists of OCT camera which uses a low coherence interferometry in which low coherence visible light is allowed to penetrate human retina and it is reflected back to interferometer producing a cross sectional image of retina [7]. The delays of the back reflected waves cannot be measured directly, so a reference measurement is used [11]. Interferometers are investigative tools used in many fields of science and engineering. They produce interference fringes by splitting a light beam into two. The interferometers is used to direct part of light to the sample and another portion is sent to a reference arm with a well-known length [11].

2.1.1 Time domain OCT (TD-OCT)

In the first implementation of OCT is time domain OCT (TD-OCT) shown in Figure 2-1

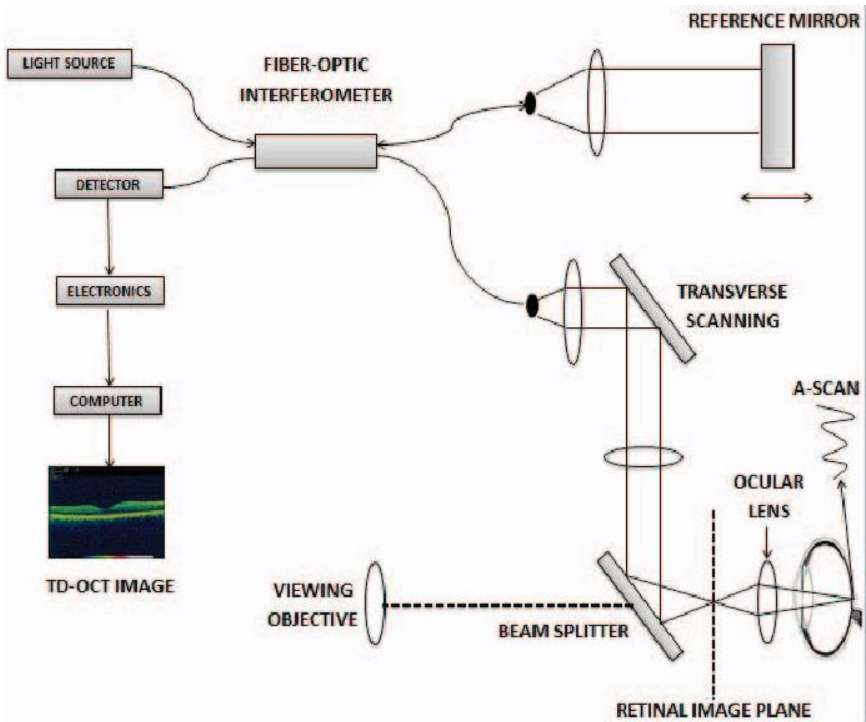


Figure 2-1 TD-OCT Imaging System based on Michelson interferometer [7]

The light of a low-coherence source is passed to the fiber-based interferometer. In a system using bulk optics the fiber coupler is replaced by a beam splitter [11]. The input beam is split into the sample beam and into the reference beam travelling to a mirror on a translational stage. The back-reflected light from each arm is combined and only interferes if the optical path lengths match and therefore the time travelled by the light is nearly equal in both arms. Modulations in intensity, also called interference fringe bursts, are detected by the detector [11]. For each sample point, the reference mirror is scanned in depth (z) direction and the light intensity is recorded on the photo detector [11]. As a result, a complete depth profile of the sample reflectivity at the beam position is generated, which in similar to ultrasound medical imaging technique that is called A-scan (amplitude scan) [11]. To create a cross-sectional image (or B-Scan), the sample beam is scanned laterally across the sample [11]. The scanned lateral sample is similar to the originated in ultrasound imaging and called B-Scan that means brightness scan.

2.1.2 Fourier domain OCT (FD-OCT)

Fourier domain OCT (FD-OCT) is the second generation of OCT technology and provides a more efficient implementation of the principle of low-coherence interferometry. In contrast to TD-OCT, FD-OCT uses spectral information to generate A-scans without the need for mechanical scanning of the optical path length [11]. As the speed of mechanical moving part is slow and each A-scan is captured and accumulated sequentially so the scan time of TD-OCT is slow [7]. The FD-OCT based on spectrometer, which is commonly referred to as spectral domain OCT (SD-OCT) was first proposed by Fercher et al. in 1995 [12]. It is used to capture all A-scans at all wavelengths in parallel using the spectrometer which replaces the point detector of the TD-OCT. The spectrometer uses a diffractive element to spatially separate the different wavelength contributions into a line image which is recorded by a high-speed line scan camera [11]. The SD-OCT is shown in Figure 2-2.

The advantage of SD-OCT over TD-OCT is that no need for mechanical moving reference which leads to high scanning speed of retinal layers. SD-OCT is 40–110 times faster than TD-OCT [13]. Moreover, It has higher sensitivity than TD-OCT [14].

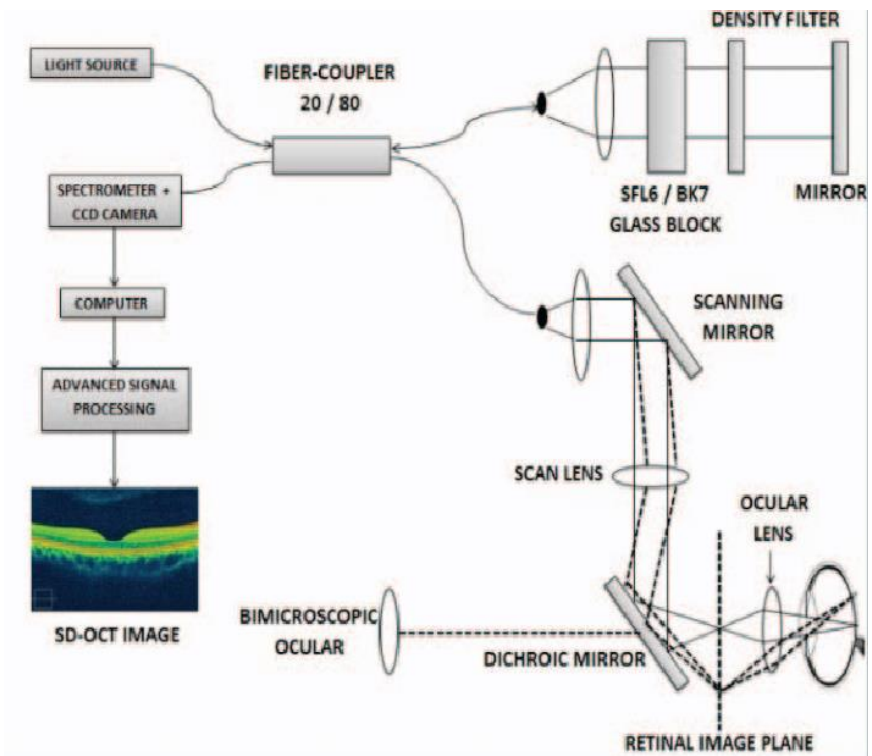


Figure 2-2 SD-OCT imaging system based on Michelson interferometer [7]

Regarding the previously stated advantages of SD-OCT, the images acquired for retinal disorder analysis are produced from SD-OCT retina scanning.

2.2 Retinal disorders

OCT enable the identification of a range of disorders with various severity levels through imaging the retina and its layers. Retina contains several layers with different thicknesses and intensities. Segmenting and measuring the thickness of each layer are considered as essential markers in assessing the health of the retina. The shape and width of the individual layers are considered two of the most important factors that reveal the current status of retina [13]. They may thicken or thin according to different diseases that indicate the current progress or status of a disease [13].

Many retinal diseases can be diagnosed from the irregularities in OCT images, such as Glaucoma, DME, CNV, Age-related macular degeneration (AMD), Cone-Rod Dystrophy (CRD), Retinitis Pigmentosa (RP) and Achromatopsia [13]. Any change in retinal thickness or absence of certain layer(s) in the retina can be used as a reference for diagnosis, disease progression and treatment monitoring [15]. Different studies are conducted, some of them define intraretinal layer while the rest of studies focuses on the most critical retina layers that are needed to identify a disease [13]. The minimum number of layers to be detected from OCT images are 13 1ayers which are Internal Limiting Membrane (ILM), Retinal Nerve Fiber Layer (RNFL), Ganglion Cell Layer (GCL), Inner Plexiform Layer (IPL), Inner Nuclear Layer (INL), Outer Plexiform Layer (OPL), Outer Nuclear Layer (ONL), External Limiting Membrane (ELM), Photoreceptor Layer (PR), Retinal Pigment Epithelium (RPE), Bruch's Membrane (BM), Choriocapillaris (CC) and Choroidal Stroma (CS) [16]. Figure 2-3 shows an OCT retinal image with its 13 distinctive layers.

Retinal Layers

Abbr.	Name
ILM	Internal Limiting Membrane
RNFL	Retinal Nerve Fibre Layer
GCL	Ganglion Cell Layer
IPL	Inner Plexiform Layer
INL	Inner Nuclear Layer
OPL	Outer Plexiform Layer
ONL	Outer Nuclear Layer
ELM	External Limiting Membrane
PR	Photoreceptor Layers
RPE	Retinal Pigment Epithelium
BM	Bruch's Membrane
CC	Choriocapillaris
CS	Choroidal Stroma

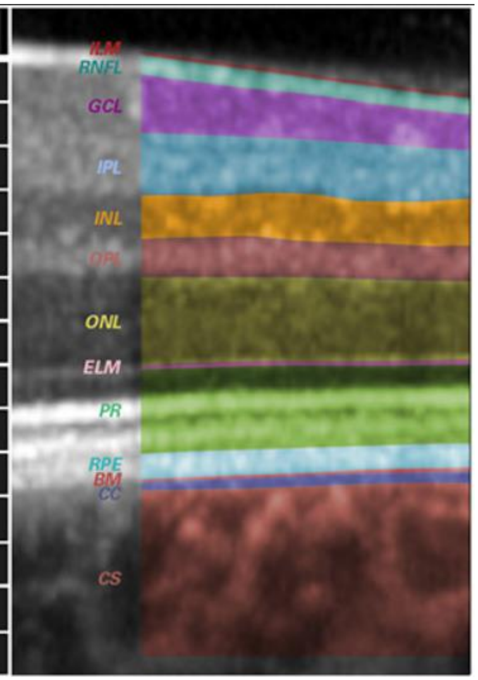


Figure 2-3 OCT retinal image with its distinctive 13 layers for a typical healthy person [13]

The 12 retinal layers detected in OCT retinal images can be used as a sign for healthy or diseased eyes. Recently, there are many studies, which are conducted to analyze the OCT retinal images to classify the healthy from diseased patients [13]. Each disease can be classified depending on some characteristics that can be absent or found in the OCT images. The morphological features, such as the shape and distribution of macular holes (MHs), cysts, drusen and blood vessels, can be visualized from OCT imaging and used as sign for different retinal disorders [13]. The next sections will discuss the characteristics of healthy eyes and diseased eyes. Moreover, the complications of each retina disorder are stated.

2.1.2.1 Normal healthy eye

OCT technology depicts tissue reflectivity. It is dependent on the tissue optical properties, microscopic refractive index of subcellular structures variations and the amount of light signal that the tissues absorb [13]. There is no standard number for retina layers for the OCT retina imaging. Different studies estimate different numbers, which are four, seven, ten, eleven, or twelve layers. On the other hand, normal retinal thickness differs from one device to another due to different characteristics, such as age, gender, race, and refraction. In addition, all measurements using SD-OCT have higher values than using TD-OCT due to the higher resolution [14]. Figure 2-4 represents an OCT retinal image for a normal person in retina macular region.

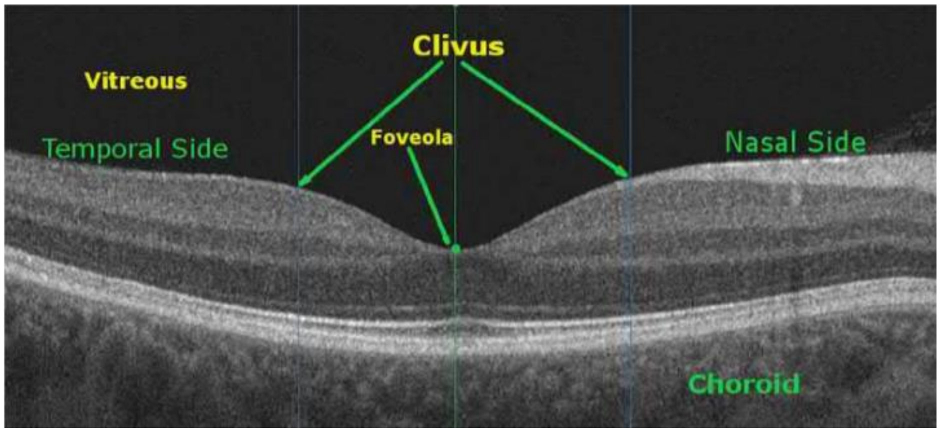


Figure 2-4 OCT normal retinal image [13]

2.1.2.2 Glaucoma

Glaucoma is a set of neurodegenerative eye diseases that leads to the loss of vision and blindness. It is the second leading cause of blindness in the world [17]. Both NFL thickness and the Euclidian distance between the Inner Limiting Membrane (ILM) and NFL can be used as sign for the presence of glaucoma disease. The glaucoma patient has a decreased NFL thickness as compared to the typical healthy subjects [18]. Recently, both of the choroid thickness and the measure of separation between Bruch’s Membrane (BM) and choroid can also be used as a sign of the presence of glaucoma disease [18]. Figure 2-5 shows an OCT image for a glaucoma patient.

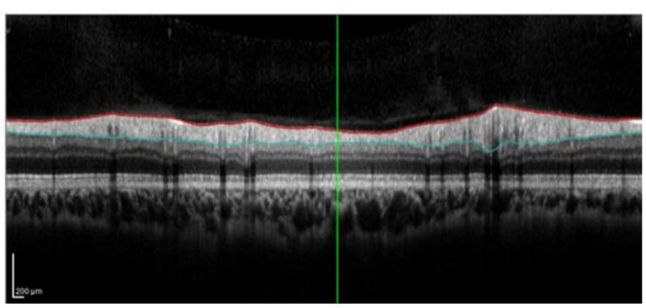


Figure 2-5 OCT glaucoma retinal image [18]

2.1.2.3 Central serous chorioretinopathy (CSC)

CSC is a chorioretinal illness that is not understood completely with systemic associations and it causes vision loss [13]. It has a multifactorial etiology with a very complicated pathogenesis [13]. Ophthalmoscopic indications of CSC range from mono- or paucifocal RPE lesions with a noticeable increase of the neurosensory retina by clear fluid to shallow detachments overlying large patches of irregularly depigmented RPE [13]. The irregular depigmented RPE results of a fluid that is accumulated under the retina distortions. Detecting the changes in both of fluid and RPE layer between normal retina and CSC disorder retina can help in diagnosing this disease [13]. Figure 2-6 shows an OCT retinal image for a CSC patient.

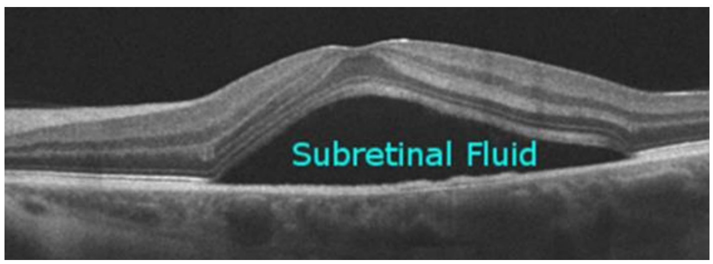


Figure 2-6 OCT CSC retinal image [13]

2.1.2.4 Unilateral anterior ischemic optical neuropathy (AION)

AION is causing damage to the optic nerve from inadequate blood supply that results in loss of vision. AION has two main types: Arteritic AION (AAION) and Non-arteritic AION (NAION) [13]. In AAION the retinal nerve fiber layer thickness. In NAION, the retinal NFL layer thickness is significantly increased in the acute stage. Then, it is significantly decreased in the resolving stage. Figure 2-7 shows an OCT retinal image for a NAION patient.

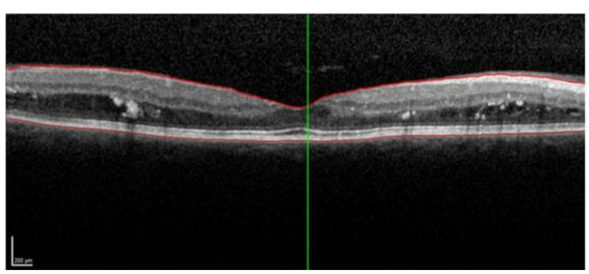


Figure 2-7 OCT NAION retinal image [13]

2.1.2.5 Choroidal Neovascularization (CNV)

It is the creation of new blood vessels in the choroid layer of the eye. Neovascular Degenerative maculopathy is commonly accompanied with extreme myopia, malignant myopic degeneration or age-related developments [19]. It is a major cause of visual loss.

An example of active myopic choroidal neovascularization with a neovascular membrane under the fovea together with subretinal fluid collection and destruction of underlying Bruch's membrane [19]. Figure 2-8 shows an OCT retinal image for a DME patient.

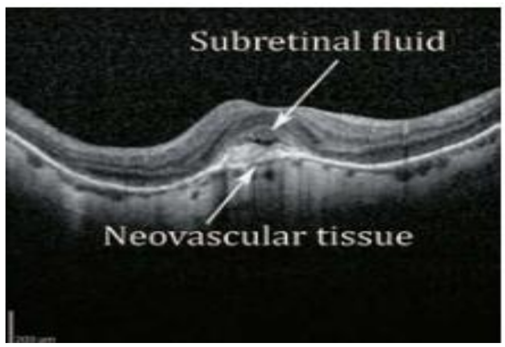


Figure 2-8 OCT CNV retinal image [19]

2.1.2.6 Age-related macular degeneration (AMD) and DRUSEN

AMD is a degeneration of the eye that is leading to severe visual impairment and visual loss for people who are 55 years old or older [13]. This disease is detected by searching for drusen, which is defined as abnormality between the basal lamina of RPE and the inner collagenous layer of BM [13]. Drusen is small yellow deposits of fatty proteins (lipids) that accumulate under the retina. They are tiny pebbles of debris that build up over time and can result in central vision loss [20]. Figure 2-9 shows an OCT retinal image for AMD patient with DRUSEN.

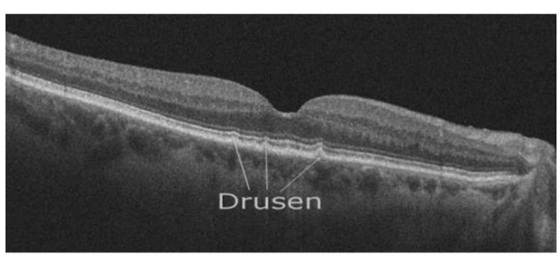


Figure 2-9 OCT Drusen retinal image

2.1.2.7 Diabetic Macular Edema (DME)

Macular Edema is the swelling of the central part of the retina [21]. It is considered one of the most common retinal disorders. Increased blood sugar levels, in people with diabetes, can damage the retinal blood vessels or result in tiny bulges protruding from the vessel walls, leaking or oozing fluid and blood into the retina. This leads to the serious eye complication called Diabetic Macular Edema (DME) [22].

DME causes cystoids called Cystoids Macular Edema. It affects the full thickness of the retinal tissue involving the anatomic fovea [23]. People over 60 years old are more prone to this retinal disorder [23]. It affects visual acuity and may lead to loss of vision or even blindness [22]. Figure 2-10 shows an OCT retinal image for a DME patient.

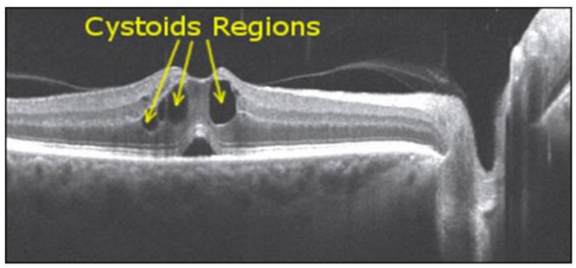


Figure 2-10 OCT DME retinal image

2.3 Summary

In this chapter, a medical background review on different medical images techniques and their importance was explained. A comparison between different retinal screening techniques such as fundography and OCT imaging techniques were depicted. According to literature, it was depicted that OCT is more sensitive to changes in retinal layers shape and thickness. Moreover, two different systems of OCT imaging were explained and compared. The result of that comparison shows that SD-OCT has improved the sensitivity and speed of capturing retinal layers images. Finally, a full description of the effect of different retinal disorders on the retinal layers were explained.

Chapter Three

Literature Survey

3 LITREATURE SURVEY

Machine learning and deep learning are used in many medical applications for diagnosing different diseases or disorders from different medical imaging techniques such as Magnetic Resonance Imaging, Computed Tomography, X-rays, Mammograms, Fundography, Ultrasound, Optical Coherence Tomography, etc….. This chapter provides a summary of the recent studies in retina segmentation. Moreover, it illustrates the machine learning concept and its importance in OCT retinal disorders early detection and classification. It provides a description of traditional machine learning based classifiers. Finally, deep learning concept and different architectures are explained.

3.1 Literature Review on Retina Segmentation

In particular, segmentation has received considerable attention recently as it is an important step in aiding diagnosis since it helps determine the RoI.

Different studies are conducted, some of them define intraretinal layer while the rest of studies focuses on the most critical retina layers that are needed to identify a disease [13]. Bagci et al. [24] were able to detect six different retinal layers, which are NFL, IPL + GCL, INL, OPL, ONL + PIS, and POS from normal healthy eyes. Lu et al. [25] identified and measured the thickness of six layers extracted from 3D images of healthy subjects. These layers were NFL, PIS, POS, retinal ganglion cell (RGC), IPL, and OPL.

Garvin et al. [26] proposed a method to segment five layers (NFL, GCL+IPL, INL+OPL, IS, OS) in OCT retinal images. The layers were identified using constructed geometric graph and computed a minimum cost closed set. This graph is constructed from the edge/regional information and a priori determined surface smoothness and interaction constraints.

Ghorbel et al. [27] proposed a method for the segmentation of eight retinal layers in Heidelberg spectralis SD-OCT images. Global segmentation algorithms such as active contours and Markov random fields are used. In addition, a Kalman filter was designed. It is used to model the approximate parallelism between the photoreceptor segments [27].

Shi et al. [28] were able to detect 10 retinal layers in 3D images of patients with retinal pigment epithelial detachments (PED) by using multi-resolution graph search based surface detection. Sugmk et al. [29] identified the retinal pigment epithelium (RPE) layer. This layer is important in detecting the shape of drusen. Moreover, the RPE layer is used to localize the retinal nerve fiber layer (RFL) and to detect a bubble of blood area in RFL complex.

In the work of Salarian et al. [30], a method that uses graph theory and the shortest path algorithm was presented to detect certain layers. The aim was to choose the RoI that could be used to distinguish normal cases from abnormal ones. They discovered that using changes in some parts, such as inner limiting membrane (ILM), retinal nerve fiber layer (RNFL) and retinal pigment epithelium (RPE), leads to separating these layers easily. The proposed technique was applied to all B-Scan images of 16 people, including low-quality images and some images with diseased eyes. The results were accurate according to manual segmentation of an expert.

A comparison between different edge detectors was conducted by Luo et al. [31]. They studied and compared the performance of canny edge detector [31], the two-pass method proposed by Bagci et al. [24] and Edgeflow technique [32]. All of these techniques were used with retinal OCT images to delineate the retinal layer boundaries. From the evaluation of the results, it was shown that the two-pass method outperforms the Canny detector and the EdgeFlow technique that is used with OCT images to delineate the retinal layer boundaries. In addition, the mean localization deviation metrics show that the smallest edge shifting problem is caused by the two-pass method. The study suggests that the two-pass method is the best one for retinal OCT boundary detectors.

ElTanboly et al. [33] used a joint Markov-Gibbs random field (MGRF) model of intensities and shape descriptors to detect 12 different layers, which are GCL, NFL, INL, IPL, OPL, ONL, external limiting membrane (ELM), myoid zone (MZ), ellipsoid zone (EZ), outer photoreceptor (ORP), interdigitating zone (IZ), and RPE layers of healthy retinas.

Dash et al. [34] proposed a graph-based segmentation technique for separating four layers which are ILM, RPE, INL and ONL. Dodo et al. [35] proposed fully automatic method for annotation of retinal layers in OCT images comprising of fuzzy histogram hyperbolisation (FHH) and graph cut methods to segment 7 retinal layers across 8 boundaries. The boundaries are (NFL-GCL, INL-OPL, IS-OS, RPE, ILM, IPL-INL, OPL- ONL, OS-RPE).

Table 3-1. summaries retina segmentation methods stated in the literature survey with stating their results and limitations.

Table 3-1. Retinal Layers segmentation literature survey

Author	Year	Techniques applied	Results	Limitations
Bagci et al. [24]	2007	Two pass edge detection algorithms to segment 6 layers (NFL, IPL + GCL, INL, OPL, ONL + PIS, and POS layers).	The RMSE for detection of boundaries ranged between 2.6 and 5.5 pixels, corresponding to 5.2 and 11.0 microns.	• Applied only on normal trace • Detect 6 layers only from retinal OCT
Garvin et al. [26]	2008	The five layers (NFL, GCL+IPL, INL+OPL, IS, OS) in OCT retinal images were identified using constructed geometric graph and computed a minimum cost closed set. This graph is constructed from the edge/regional information and a priori determined surface smoothness and interaction constraints.	Overall mean unsigned border positioning error of 2.9 microns. Inner retinal layer thickness for the affected eye was (21%) smaller on average than for the unaffected eye (verified by 3 medical experts)	• Applied only on unilateral anterior ischemic optic neuropathy (AION) • Detect 5 layers only from retinal OCT.
Lu et al. [25]	2011	The five layers (NFL, PIS, POS, retinal ganglion cell (RGC), IPL, and OPL) where identified. OCT image is first cut into multiple sections by the retinal blood vessels that are detected through an iterative polynomial smoothing procedure. The non-vessel OCT sections are then filtered. Finally, the layer boundaries of the filtered non-vessel OCT sections are detected, which are further clustered to differentiate retinal layers to determine the complete retinal layer boundaries.	The boundary positional error ranges between 1.78 ±0.79 and 4.89 ±2.39	• Limited number of subjects (only 4 patients) • Detect 5 layers only from retinal OCT.
Ghorbel et al. [27]	2011	proposed a method for the segmentation of seven retinal layers (RNFL, GCL+IPL, INL, OPL, ONL, IS+OS and RPE). Global segmentation algorithms such as active contours and Markov random fields are used. In addition, a Kalman filter was designed. It is used to model the approximate parallelism between the photoreceptor segments.	An overall sensitivity of 90.57% is achieved on Topcon database and 90.71% on Spectralis database.	• Applied only on normal trace • Detect 6 layers only from retinal OCT.

Table 3-1 (cont.). Retinal Layers segmentation literature survey

Author	Year	Techniques applied	Results	Limitations
Shi et al. [28]	2014	Proposed a model that was able to detect 10 retinal layers (NFL, GCL, IPL, INL, OPL, ONL+ISL, CL, OSL, VM and RPE) in 3D images of patients with retinal pigment epithelial detachments (PED) by using multi-resolution graph search-based surface detection	overall mean unsigned border positioning error for layer segmentation is 7.87±3.36 µm and is comparable to the mean inter-observer variability (7.81±2.56 µm).	• Limited number Pigment Epithelial detachments patients • Detect 10 layers only from retinal OCT
Sugmk et al. [29]	2015	Proposed an algorithm based on thresholding to segment the retinal pigment epithelium (RPE) layer. This layer is important in detecting the shape of drusen. Moreover, the RPE layer is used to localize the retinal nerve fiber layer (RFL) and to detect a bubble of blood area.	Segmentation is used for classification of DME and Normal classes which achieves 87.5%.	• Extracts only 2 layers which is not sufficient for detecting DME disease
Salarian et al. [30]	2015	proposed a method that uses graph theory, and the shortest path algorithm was presented to detect certain layers. They aim to choose the RoI that could be used to distinguish normal cases from abnormal ones. They discovered that using changes in some parts, such as inner limiting membrane (ILM), retinal nerve fiber layer (RNFL) and retinal pigment epithelium (RPE), leads to separating these layers easily.	The results were accurate according to manual segmentation of an expert. (no results are stated)	• No result values were stated (only depends on expert opinion)

Table 3-1 (cont.). Retinal Layers segmentation literature survey

Author	Year	Techniques applied	Results	Limitations
Luo et al. [31]	2017	They studied and compared the performance of canny edge detector, the two-pass method proposed by Bagci et al. 2007 and Edgeflow technique. All of these techniques were used with retinal OCT images to delineate the retinal layer boundaries.	The two-pass method achieved accuracy of 0.98 ± 0.0033 while canny edge detector and edgeflow technique achieves 0.98 ± 0.0026 and 0.97 ± 0.0027 respectively The mean localization of two-pass method, edge detector technique and edgeflow technique are 1.02 ± 0.2484, 1.27 ± 0.3340 and 1.05 ± 0.1831 respectively. The study suggests that the two-pass method is the best one for retinal OCT boundary detectors.	• The limitation of this comparison is that they uses a dataset with normal traces only. • The comparison is limited to segmentation of 6 layers only to have a fair comparison between these different techniques
ElTanboly et al. [33]	2017	Used a joint Markov-Gibbs random field (MGRF) model of intensities and shape descriptors to detect 12 different layers, which are GCL, NFL, INL, IPL, OPL, ONL, external limiting membrane (ELM), myoid zone (MZ), ellipsoid zone (EZ), outer photoreceptor (ORP), interdigitating zone (IZ), and RPE layers of healthy retinas.	Overall segmentation accuracy was 73.2 ± 4.46.	• Segmentation accuracy is not high • Some important layers like Bruch's membrane is not segmented. • Applied on normal traces only
Dash et al. [34]	2018	Proposed a graph-based segmentation technique for separating four layers (ILM, RPE, INL and ONL)	Overall segmentation sensitivity was 90%.	• Detect 4 layers only from retinal OCT
Dodo et al. [35]	2019	Proposed fully automatic method for annotation of retinal layers in OCT images comprising of fuzzy histogram hyperbolisation (FHH) and graph cut methods to segment 7 retinal layers across 8 boundaries. The boundaries are (NFL-GCL, INL-OPL, IS-OS, RPE, ILM, IPL-INL, OPL- ONL, OS-RPE)	RMSE of layers NFL, GCL+IPL, INL, OPL, ONL+IS, OS and RPE are 0.2688 ± (0.0185), 0.5762 ± (0.0590), 0.6307 ± (0.0785), 0.4839 ± (0.0410), 0.6596 ± (0.0823), 0.4401 ± (0.0362), 0.4369 ± (0.3291) respectively	• Applied only on normal trace • Detect 7 layers only from retinal OCT

3.2 Literature Review on Retina Disorder Classification Using OCT

Although the previous studies present effective segmentation techniques, there is a need to develop comprehensive diagnostic systems that emphasize RoI and perform disease classification. Such systems can help alleviate the burden on medical care systems. Current automated diagnostic systems either use traditional ML techniques of feature extraction, segmentation and classification or utilize deep learning architectures with unprocessed images.

An example of these automated systems with OCT imaging can be found in the work of Srinivasan et al. [36]. It described an automated method for detecting retinal diseases using features extracted by Histogram of Oriented Gradients (HOG) techniques and SVMs to classify three classes namely DME, dry AMD and Normal. Their method does not rely on the segmentation of inner retinal layers; however, retinal flattening is performed. The proposed model achieved 100% for both DME and dry AMD classes and 86.67% for normal class. Another example of traditional approach is that of Alsaih et al. [15]. They proposed a pipeline model that generated a large set of extracted features using HOG and Local Binary Pattern (LBP) at four levels of the multiscale Gaussian lowpass image pyramid. These features were either reduced using Principle Component Analysis (PCA) or directly passed to SVM classifier, which was used to classify two classes: DME and normal. Their best model achieved an accuracy of 81%. Naz et al. [6] addressed the problem of automatic classification of OCT images through SVM (leave one out) technique after image denoising and extraction of features from thickness profile and cyst spaces. This technique was based on extracting features from segmenting the retinal layer (ILM and choroid layer) and from the changes in thickness of these layers. Naz et al. [6] applied that technique for the identification of patients with DME versus normal subjects. The proposed model achieved an accuracy of 79.65%.

These traditional approaches [6, 15, 36], which employed hand-crafted feature for the classifiers, were shown to obtain promising results. However, these approaches shared a common disadvantage, which is requiring abundant expert knowledge. Medical knowledge was needed to correctly detect retinal layers and/or changes in layers' thickness [6] and/or assist in retinal flattening [15, 36]. Such dependence on experts' availability gives rise to another concern, which is having poor generalization to other domains [37]. Despite these limitations, hand-crafted features remain to provide intuitive information, which enables the needed analysis of the model [38]. Hence, the generation of hand-crafted features that are not heavily reliant on experts' knowledge would provide the required balance of providing explanation of the model without being dependent on an expert. This balance learning techniques were employed to perform embedded feature extraction implicitly. Awais et al. [39]used a deep learning architecture (VGG16) for extracting features from different layers of the network. Classifications were made using KNN and Decision Tree based on the extracted features. The model is used to classify two classes: DME and normal. The best model got an accuracy of 87.5%.

Perdomo et al. [40] proposed a new CNN architecture for OCT classification. The new CNN consists of two blocks, the first block contains four subblocks with convolutional layers and max pooling and the last block has two fully connected layers. The new CNN was evaluated with different learning rates, until it was able to correctly detect DME with a classification accuracy of 93.75% between two classes: DME and normal at learning rate 0.00001.

Motozawa et al. [41] developed a pipeline model consisting of two CNNs. The first is used to differentiate AMD class from normal class and the second CNN model is used to divide the AMD class into those with exudates and without exudates using transfer learning. The first CNN achieved an accuracy of 99% and the second one achieved an accuracy of 93.9%. Despite that the models attained high accuracies; the classification was limited to two classes which simplifies the problem. The work of Kermany et al. [42] established a diagnostic tool based on a deep learning framework for the screening of patients with common treatable blinding retinal diseases. They applied their approach on an OCT dataset [8] with four classes which are DME, CNV, DRUSEN, NORMAL. They used transfer learning inception model which achieved 96.1% average accuracy. Their approach demonstrated performance comparable to that of human experts in classifying age-related macular degeneration and diabetic macular edema. Using the same dataset, Li et al. [37] adopted a transfer learning VGG16 network that achieved an accuracy of 98.6%. Both studies applied deep learning approaches [37, 42] to classify four different classes including serious diseases (DME, CNV) and achieved promising results. However, the problem for deep learning models (CNNs) presented in the work of Kermany et al. [42] and Li et al. [37] is that its sensitivity to urgent referrals classes (CNV/DME) is unsatisfactory. This is the main argue point against the approaches in the work of Kermany et al. [42] and Li et al. [37]. Another common concern of deep learning architecture is the limited interpretability of the constructed models. The incorporation of hand-crafted features can provide an adequate solution to deep learning issues. Hand-crafted features can offer complementary information that increase the understandability of the classification decision [43, 44]. In addition, the use of hand-crafted features was shown to considerably improve the performance of pure deep learning approaches [45]. Recently, similar studies for image classification and recognition have shown that the integration of deep and traditional learning yield better performance. Examples of these studies include the work of Zhang et al. [46] for face recognition and Xie at al. [47] for lung nodule classification. Moreover, deep learning architectures in return can provide automated localization of the retinal RoI, which overcomes the need for an expert to aid in the specification of the RoI, required for traditional feature extraction.

Table 3-2. summaries retina disorders classification stated in the literature survey with stating their results and limitations.

Table 3-2. Retinal Disorders classification Literature Survey

Author	Year	Techniques applied	Results	Limitations
Srinivasan et al. [36]	2014	Proposed an automated method for detecting retinal diseases using features extracted by Histogram of Oriented Gradients (HOG) techniques and SVMs to classify three classes namely DME, dry AMD and Normal. Their method does not rely on the segmentation of inner retinal layers; however, retinal flattening is performed.	The proposed model achieved 100% for both DME and dry AMD classes and 86.67% for normal class.	• The disadvantage of it is requiring abundant expert knowledge • Medical experts were needed in retinal flattening
Alsaih et al. [15]	2017	Proposed a pipeline model that generated a large set of extracted features using HOG and Local Binary Pattern (LBP) at four levels of the multiscale Gaussian lowpass image pyramid. These features were either reduced using Principle Component Analysis (PCA) or directly passed to SVM classifier, which was used to classify two classes: DME and normal.	Their best model achieved an accuracy of 81%.	
Naz et al. [6]	2017	addressed the problem of automatic classification of OCT images through SVM (leave one out) technique after image denoising and extraction of features from thickness profile and cyst spaces. This technique was based on extracting features from segmenting the retinal layer (ILM and choroid layer) and from the changes in thickness of these layers. Naz et al. applied that technique for the identification of patients with DME versus normal subjects.	The proposed model achieved an accuracy of 79.65%.	• The disadvantage of it is requiring abundant expert knowledge • Medical experts were needed to correctly detect retinal layers or changes in layers' thickness
Awais et al. [39]	2017	Used a deep learning architecture (VGG16) for extracting features from different layers of the network. Classifications were made using KNN and Decision Tree based on the extracted features. The model is used to classify two classes: DME and normal.	The best model got an accuracy of 87.5%.	• The classification was limited to two classes which simplifies the problem. • Another common concern of deep learning architecture is the limited interpretability of the constructed models

Table 3-2 (cont.). Retinal Disorders classification Literature Survey

Author	Year	Techniques applied	Results	Limitations
Perdomo et al. [40]	2018	proposed a new CNN architecture for OCT classification. The new CNN consists of two blocks, the first block contains four subblocks with convolutional layers and max pooling and the last block has two fully connected layers. The new CNN was evaluated with different learning rates. The proposed model could separate between two classes: DME and Normal.	The proposed model achieved classification accuracy of 93.75% at learning rate 0.00001.	• The classification was limited to two classes which simplifies the problem. • Another common concern of deep learning architecture is the limited interpretability of the constructed models.
Motozawa et al. [41]	2019	developed a pipeline model consisting of two CNNs. The first is used to differentiate AMD class from normal class and the second CNN model is used to divide the AMD class into those with exudates and without exudates using transfer learning. Despite that the models attained high accuracies.	The first CNN achieved an accuracy of 99% and the second one achieved an accuracy of 93.9%.	• The classification was limited to two classes which simplifies the problem. • Another common concern of deep learning architecture is the limited interpretability of the constructed models
Kermany et al. [42]	2018	Established a diagnostic tool based on a deep learning framework for the screening of patients with common treatable blinding retinal diseases. They applied their approach on an OCT dataset with four classes which are DME, CNV, DRUSEN, NORMAL. They used transfer learning inception model.	The proposed model achieved an accuracy, sensitivity and specificity of 96.1%, 97.8% and 97.4% respectively.	• Limited interpretability of deep learning models. • No visualization of RoI
Li et al. [37]	2019	Established a diagnostic tool based on a deep learning framework for the screening of patients with common treatable blinding retinal diseases. They applied their approach on an OCT dataset with four classes which are DME, CNV, DRUSEN, NORMAL. They used transfer learning VGG16 model.	The proposed model achieved an accuracy, sensitivity and specificity of 98.6%, 97.8% and 99.4%, respectively.	• Limited interpretability of deep learning models. • No visualization of RoI

3.3 Review on Machine learning-based classification techniques

The following subsections describe different supervised machine learning classification algorithms where each input belongs to a predefined output class during learning phase.

3.3.1 Suppot Vector Machine (SVM)

Support Vector Machines (SVM) is a group of supervised learning algorithm that can be applied to classification or regression. Support Vector Machines (SVMs) are the newest supervised machine learning technique. It is theoretically well motivated algorithm: derived from Statistical Learning Theory by Vapnik and Chervonenkis since the 60s [48]. SVM works by taking a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. SVMs revolve around the notion of a "margin"—either side of a hyperplane that separates two data classes. The goal of SVM is maximizing the margin and thereby creating the largest possible distance between the separating hyperplane and the instances on either side of it has been proven to reduce an upper bound on the expected generalization error [49]. The SVM is has two phases training and testing phases. At training, a set of training data each labelled as depending on one of two categories. An SVM training algorithm builds a model that designates new examples into one category or the other. This method is a representation of the examples as points in space, mapped so that the examples of the different classes are divided by a clear gap that is as wide as possible [49]. During testing phase, new examples are then mapped into that same space and estimated according to a class based on which side of the gap they fall on [49]IN. SVM has empirically good performance and successful applications in many fields such as bioinformatics, text, pattern recognition and others.

3.3.2 Neural Networks

The basic idea of neural networks is the artificial neuron which is based on the functionality of the neuron inside the biological brain. Every neuron or node (nerve cell) is connected to each neuron in the next layer through a connection link. A nerve cell is made up of axon (output), dendrites (input), a node (soma), nucleus (activation function), and synapses (weights). The activation function in the artificial neuron acts as the nucleus in a biological neuron whereas the input signals and its respective weights model the dendrites and synapses respectively. Figure 3-1 illustrates the neuron structure.

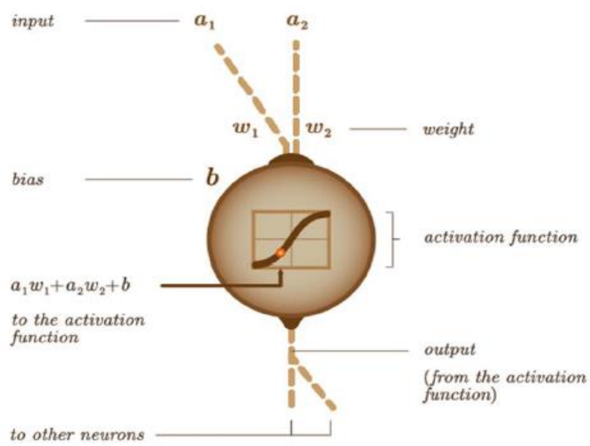


Figure 3-1 Artificial Neuron Structure

3.3.2.1 Perceptron Learning

In perceptron learning, if $\propto_1$ through $\propto_2$ are input feature values and w_1 through w_2 are connection weight vector (typically real numbers in the interval [−1, 1]), then perceptron computes the sum of weighted inputs: $\sum_i \propto_i . w_i$ and output goes through an adjustable threshold: if the sum is above threshold, output is 1; else it is 0 [49]. The most common way for learning from a batch of training instances is to run the algorithm repeatedly through the training set until it finds a prediction vector which is correct on all of the training set [49]. This prediction rule is then used for predicting the labels on the test set. The problem with perceptron is that it can only classify linearly separable sets of instances. If a straight line or plane can be drawn to separate the input instances into their correct categories, input instances are linearly separable, and the perceptron will find the solution [49]. Otherwise (if instances are not linearly separable) learning will reach to a point where all instances are correctly classified.

3.3.2.2 Artifical Neural Network (ANN)

An artificial neural network consists of large number of units (neurons) joined together. The model is realized with a structure that is made up of input, hidden, and output-layers, as shown in Figure 3-2.

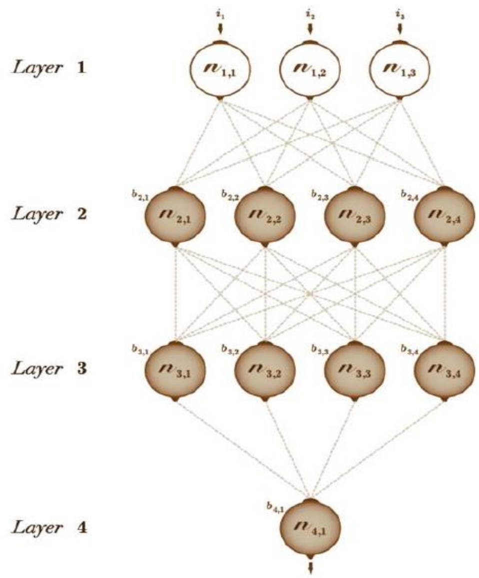


Figure 3-2 Artificial Neural Network Structure

Units in a network (shown in Figure 3-2) are usually segregated into three classes: input units, which receive information to be processed; output units, where the results of the processing are found; and units in between known as hidden units. the behavior of the ANN is defined by the current values of the weights [49]. The weights of the network to be trained are initially set to random values, and then instances of the training set are repeatedly exposed to the network. The values for the input of an instance are placed on the input units and the output of the network is compared with the desired output for this instance. Then, all the weights in the network are adjusted slightly in the direction that would bring the output values of the network closer to the values for the desired output [49]. ANNs have been applied to many real-world problems but still, their most striking disadvantage is their lack of ability to reason about their output in a way that can be effectively communicated [49]. For this reason many researchers have tried to address the issue of improving the comprehensibility of neural networks, where the most attractive solution is to extract symbolic rules from trained neural networks [49].

3.4 Review on Deep Learning classification techniques

Deep learning or Deep neural network (DNN) belongs to class of machine learning. It relies on the collection of machine learning algorithms which models high level abstractions in the data with multiple nonlinear transformations [50]. DNN is a subclass of neural networks. DNN takes learning algorithms and by continuously increasing the amounts of data, the efficiency of training processes can be improved [50]. The term “deep” represents the large number of hidden layers between the input and output layer. It is also known as hierarchical learning that consists of multiple layers which includes nonlinear processing units for the purpose of conversion and automatic feature extraction [50]. The hierarchical structure of deep Learning can be shown in Figure 3-3. Every subsequent layer takes the results from the previous layer as the input, where the higher-level features are defined from lower-level ones. The deep learning algorithms are highly motivated by artificial intelligence field which attempts to emulate the human brain’s ability to observe, analyze, learn, and make decisions. The deep learning approach is widely used in the fields of adaptive testing, big data, cancer detection, health care, document analysis and recognition, speech recognition, object detection, natural language processing, pedestrian detection, image classification and voice activity detection [50].

The main difference between traditional machine learning and deep learning is that traditional machine learning needs manual feature extraction from the collected dataset. After that, various types of automated algorithms that learn to model functions and predict future actions from data. In traditional machine learning, algorithms are directed by data analysis to examine specific variables in data sets. In deep learning algorithms, deep neural networks pass data through many processing layers to interpret automatic data features and relationships. Moreover, deep learning algorithms are self-directed on data analysis once they’re put into production. Figure 3-4 shows difference between traditional machine learning and deep learning.

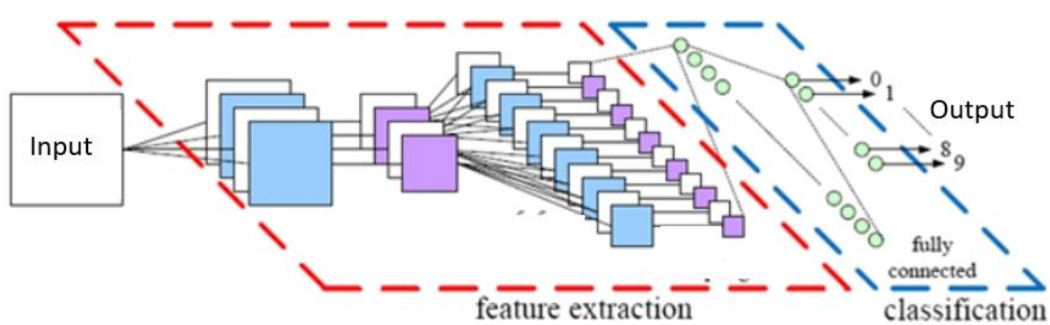
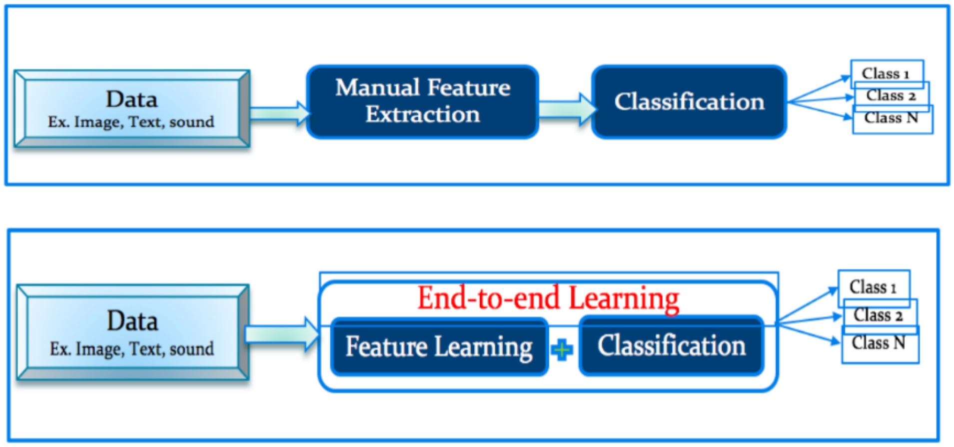


Figure 3-3 Hierarchical Deep learning networks



(a)

(b)

Figure 3-4 (a) Traditional machine learning approach and (b) Deep learning approach.

Unlike neural network, learning of DNN takes place not only as discriminative supervised learning but also generative unsupervised learning. The next subsections will describe different deep learning techniques of discriminative supervised deep networks and generative unsupervised deep networks.

3.4.1 Generative unsupervised deep networks

Generative unsupervised learning technique is used to capture correlation of observed data without availability of information about target class availability [51]. The correlation of data is useful for pattern analysis and synthesis purposes. From the generative unsupervised deep networks are Auto-Encoders (AE), Restricted Boltzmann Machine (RBMs) and Deep Belief Networks (DBNs).

3.4.1.1 Auto-Encoders (AE)

An Auto-encoder (AE) is a type of neural network which is based on unsupervised learning technique [50]. The network is trained to generate output that imitates inputs. AE architecture is composed of three main layers which are input layer, encoding layer and decoding layer. The AE tries to reconstruct/represent its input which allows the encoder layer to learn the best representation of the input. Its architecture consists of three layers which are an encoding network, a hidden layer, and a decoding network as shown in Figure 3-5. The encoder network is a feedforward neural network that tries to compress the input into a latent space and learn the best representations of the input. The hidden layer which is trying to generate a code which helps to represent the input. The decoder network is also a feedforward network, similar to encoder network, which tries to reconstruct the input back to its original dimensions.

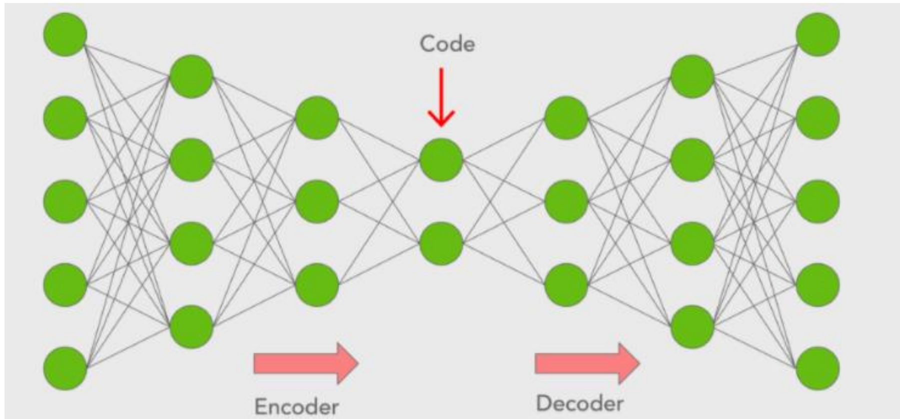


Figure 3-5 Structure of Auto Encoder Network

AE are similar to Principle Component Analysis (PCA), both of them tries to minimize the objective function. The advantage of AE over the PCA is that it has more flexibility that allow the representation of not only the linear transformation, as PCA, but also the nonlinear transformations. AEs are trained with a backpropagation algorithm that employs a metric known as the loss function [52]. The advantage of the backpropagation algorithm is that it minimizes the lost information from input reconstruction helping the AE network to produce a nearly identical reconstructed output sample [52]. From the examples of auto encoders are the denoising auto encoder network. In denoising auto encoder, part of the input dataset is elected randomly and corrupted by noise. Then parameters of the network are adjusted and the hidden layer inside the network will learn a code that help in separating the noise during reconstructing output sample [52].

3.4.1.2 Restricted Boltzmann Machines (RBMs) and Deep Belief Network (DBNs)

Boltzmann machine (BM) is a neural network of symmetrical connected neuron [51]. It consists of a number of visible layers and a number of hidden layers. All neurons inside networks are connected together. BM make stochastic decisions about whether to be on or off [51]. Restricted Boltzmann Machine (RBM) is a special type of BM that is consists of two-layer neural network with one visible layer and one hidden layer. RBM differ from BM that it has no hidden to hidden or visible to visible connections [51]. Figure 3-6 shows the connections of BMs and RBMs networks. Another special type of BM is Deep Restricted Boltzmann Machine (DBM) which consists of many hidden layers. Similar to RBMs, DBM has no connections between neurons of same layers [51]. Each layer captures complicated, higher-order correlations between the activities of hidden features in the layer below [51]. DBMs can learn more complex and highly desirable representations for solving recognition problems [51]. Further, the high-level representations can be built from a large supply of unlabeled inputs. Very limited labeled data (supervised) can be used to fine tune model for a specific task.

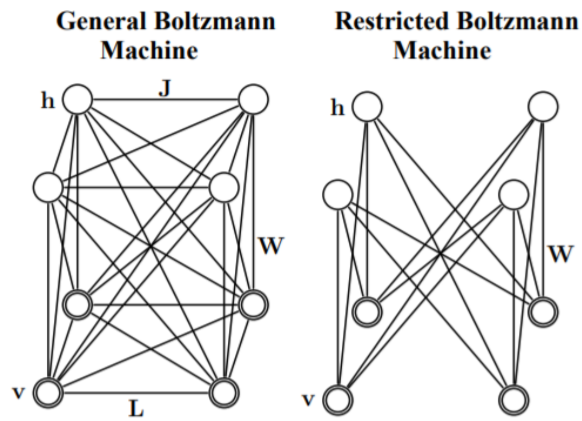


Figure 3-6 General Boltzmann Machine and Restricted Boltzmann Machine networks

Deep Belief Network (DBN) is a stacked layer architecture of several hidden layers. DBN is a stacked layer of RBMs. Each RBM network act as a layer inside DBN which is trained independently to encode the statistical dependencies of the units within the previous layer. DBN training starts by training the lower RBM x layer followed by the upper layer of h1, h2, etc… as shown in Figure 3-7 [53] that shows training of one input layer, x, and three hidden layers h1, h2, h3. From left to right, each layer of DBN is a RBM network. The bottom layers are intended to extract low-level features from the input data, while the upper layers are expected to gradually refine previously learned concepts [53].

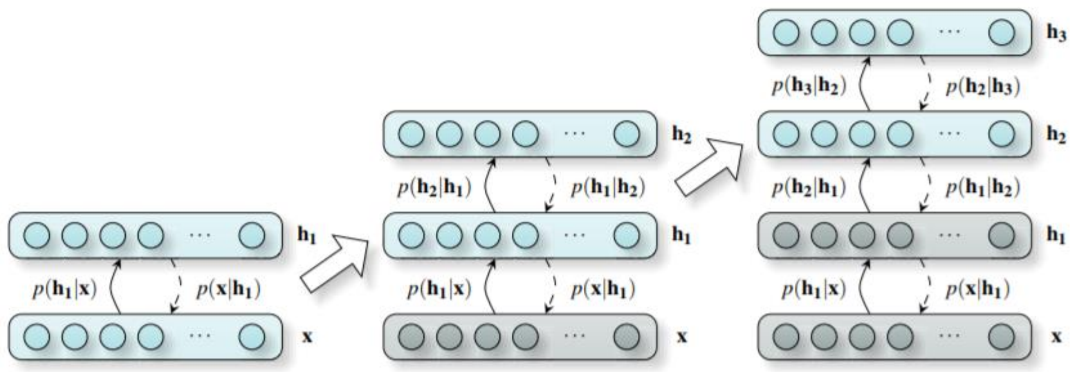


Figure 3-7 Training process of a Deep Belief Network (DBN)

3.4.2 Discriminative supervised deep networks

Discriminative supervised learning technique is used to categorize data where each input has a corresponding output target [51]. From the discriminative supervised deep networks are Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and Long-Short Term Memory (LSTM).

3.4.2.1 Convolutional Neural Network (CNN)

CNN is a neural network with multiple layers architecture [50]. CNN can extract automatic discriminative features which have some invariance properties (e.g. translation invariance) [54]. It consists of three main layers which are convolution layers, pooling layers and fully connected layers. The early convolution layers of the architecture are used for extracting local low-level features from the raw input while the deeper convolution layers of CNN are used for combining features together to generate global high-level features [50]. The pooling layers are used to down sample the dimensionality of the extracted feature. The fully connected layers form an ANN network where each neuron in the previous layer is connected to all the neurons in the current layer. The total number of fully connected neurons in the final layer determines the number of classes [52]. The advantages of CNNs include that they are well suited for end-to-end learning that generates automatic features from the raw data without any a priori feature selection. Moreover, CNNs scale well to large datasets. The disadvantages of CNNs include that they may output false predictions with high confidence, may require a large amount of training data, may take longer to train than simpler models, and involve a large number of hyper parameters such as the number of layers or tIhe type of activation functions, the limited interpretability of the constructed models. Some of famous CNN are as follow:

- ***AlexNet model:*** It was proposed by Alex Krizhevsky [55]. AlexNet CNN is a feedforward network with 8 layers. It contains five convolution layers and three fully connected layers [55]. Figure 3-8 shows the structure of AlexNet.

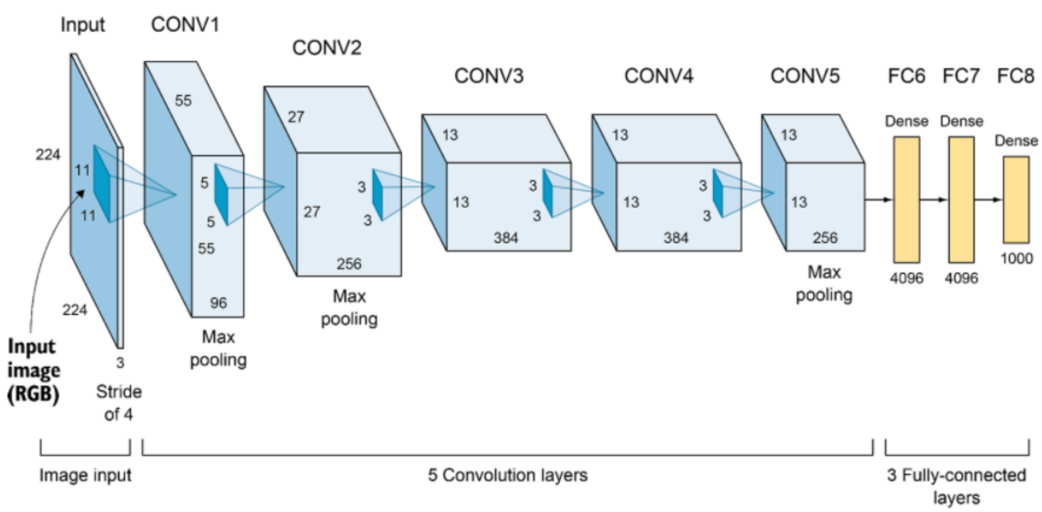


Figure 3-8. AlexNet Architecture

- **Visual Graphic Group Net (VGG Net) model:** This net was developed by the technicians at the Visual Graphics Group from the Oxford and is in pyramid shape. The model consists of the bottom layers which are wide and the top layers are deep [50]. There are 2 versions of VGG which are VGG-16 and VGG-19. VGG-16 is a combination 13 convolutional layers and Three fully connected layers as shown in Figure 3-9. The VGG-19 is a much deeper network with 16 convolutional layers and three fully connected layers [56].

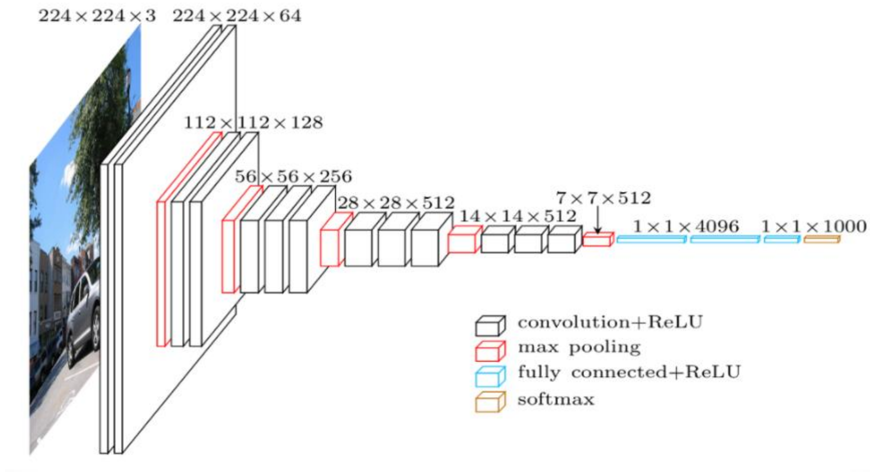


Figure 3-9. VGG-16 architecture [50]

- ***GoogleNet (Inception) model:*** They are models developed by Google [57]. It has different versions as Inception-v1, Inception-v2 and Inception-v3. All inception models are based on the idea inception block. The idea of inception block is based on not stacked convolutional layers, but stacked building blocks which them-selves consist of multiple convolutional layers. Here single layer carries multiple kinds of the feature extractors that help the network to perform better. The inception building block can be found in Figure 3-10. The difference between inception versions are the number of repetitions of the fundamental inception building block. For example, Inception-v1 has 22-layers deep, starts with three convolution layers, followed by 9 inception blocks, and ends with a fully connected layer while Inception-v3 consists of 48 layers. It starts with 6 convolution layers and followed by 10 inception blocks. The full architecture is described in detail in [57].

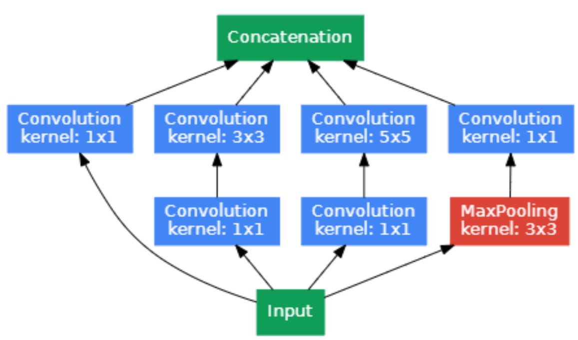


Figure 3-10. Inception Block

- **ResNet model:** It is a type of deep network based on residual learning. There are different versions of ResNet which are ResNet-18, ResNet-34, ResNet-50, ResNet-101 and ResNet-152. All of them has the same building units or residual blocks and formed by stacking the building residual blocks over each other [56, 58]. Any ResNet starts with block has a structure as shown in Figure 3-11. The residual blocks are divided into two types which are Identity shortcut and Projection shortcut. The first block shown in Figure 3-12(a) is the identity shortcut bottleneck block which is composed of a sequence of convolution layers of kernel size (1 × 1) and stride = 1 connected to a convolution layer with kernel (3 × 3) and stride = 1 followed by a convolution layer followed by kernel (1 × 1) and stride = 2 [59]. This block is used when the input and output of feature map are the same. The other block shown in Figure 3-12(b) is the projection shortcut bottleneck block which has the same sequence of layers with a newly added convolution layer in the projection shortcut which has a kernel size of (1 × 1) with stride = 2 [59]. It is applied when shortcuts go across the feature map of two sizes. In the two blocks all the convolution layers are followed by batch normalization and RELU activation function. The difference between different ResNet versions are the number of stacked residual blocks. For example, ResNet-18 with 22-layers are deep, starts with a convolution layer followed by 8 residual blocks and ends with a fully connected layer [59]. Another example is the ResNet-50 which has 16 residual blocks and ends with fully connected layer as shown in Figure 3-13.

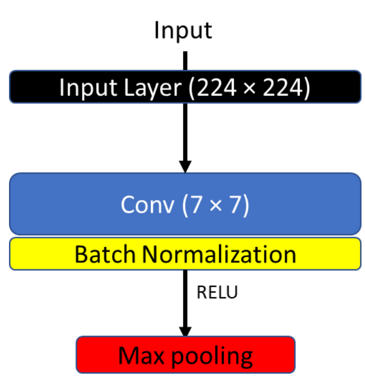


Figure 3-11. ResNet Starting Block

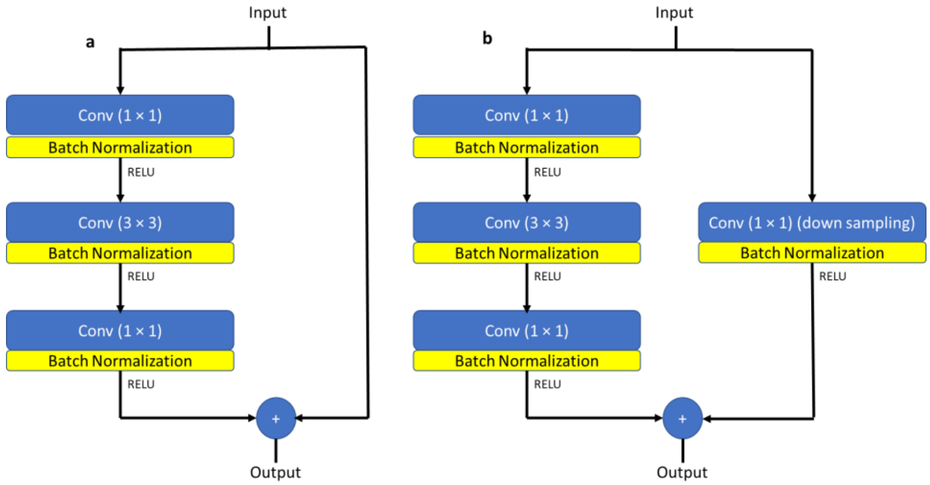


Figure 3-12. Residual Blocks (a) Identity Shortcut (b) Projection Shortcut

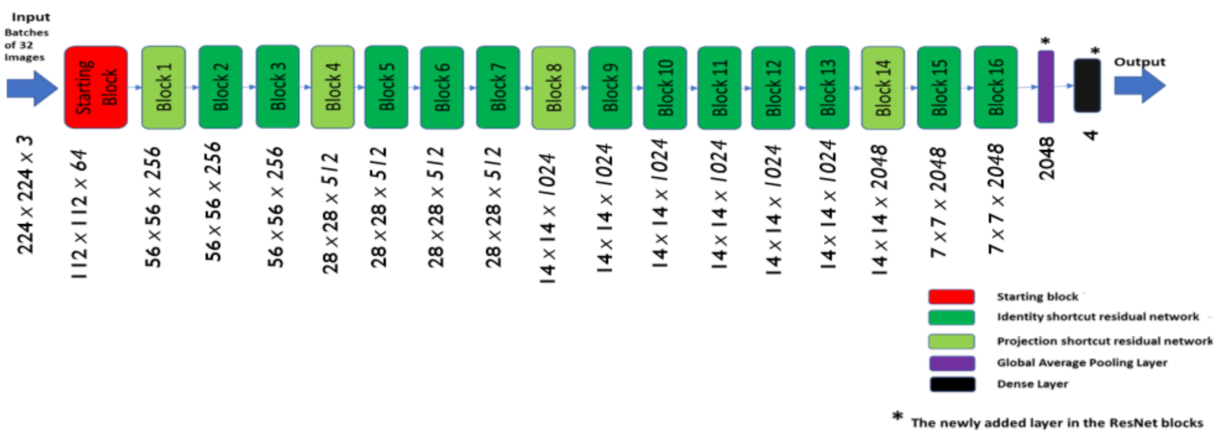


Figure 3-13. ResNet-50 architecture

- **Xception model:** It has 71 layers. It started by two convolution layers followed by depth separable convolution layers, four convolution layers and dense layer [60].
- **MobileNet model:** It is a light-weight network. It has 53 layers which are divide to 52 convolution layers and the last is the dense layer. The network start with 16 residuals and bottlenecks blocks and ends with one convolution layer followed by dense layer [61].

3.4.2.2 Recurrent Neural Network (RNN)

In contrast to the feed-forward network, the RNN employs a recursive approach (recurrent network) whereby the network performs a routine task with the output being dependent on the previous computation. Figure 3-14 shows the recurrent neural networks concept. This functionality is created with inbuilt memory. The most common type of RNN is the Long Short-Term Memory (LSTM) network. It has the capability to learn long-term dependencies.

The LSTM algorithm incorporates a memory block with three gates: the input, output, and forget gate. These gates control the cell state and decide which information to add or remove from the network. This process repeats for every input [52].

1 Input gate: decides what new information is to be stored and updated in the cell state.

2 Output gate: judges what information is used based on the cell state.

3 Forget gate: evaluates what information is redundant and discards it from the cell state.

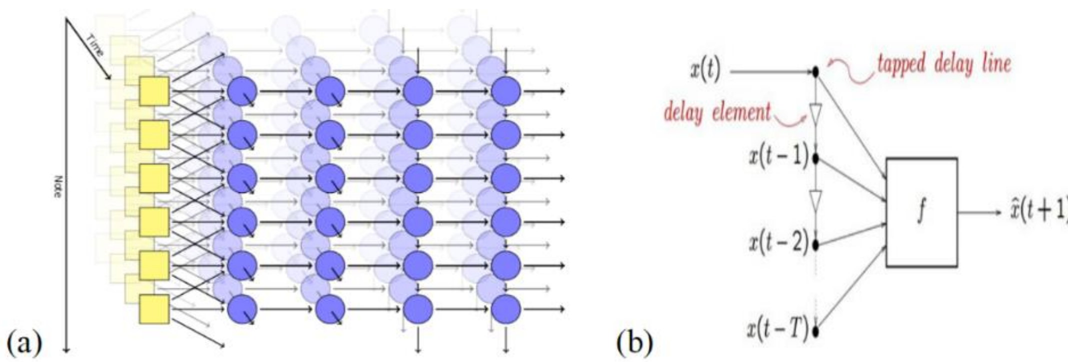


Figure 3-14. Recurrent neural networks concept: (a) RNN extends across time, (b) using the past to predict the future

The previously described deep learning architectures have demonstrated their potential by surpassing the performance of traditional machine learning techniques. Moreover, deep learning algorithms minimize the need for spatial crafted feature engineering. The big disadvantage is that these models is computationally intensive but nowadays, this problem is solved by using high-specification Graphics Processing Units GPU to accelerate the training of these models.\

3.5 Summary

In the first two sections of this chapter, the related work in the field of segmentation of retinal layers and classification of different retinal layers disorders was explained. It was depicted that the previously model needed manual segmentation before classification. The problem of manual segmentation that it needs medical experts and took a lot of time. Moreover, the automatic segmentation proposed in literature have limitation in number of layer segmented or was applied on normal trace or one retinal disorder. In the third section, two of the most famous machines leaning based classifiers are explained which are SVM and Neural networks. In the fourth section, an illustration of the deep learning concepts of generative unsupervised deep networks and discriminative supervised deep networks.

Chapter Four

Proposed Framework

4 HYCAD: HYBRID COMPUTER AIDED DIAGNOSIS PROPOSED FRAMEWORK

In this chapter, the proposed model architecture will be introduced, with an illustration of the phases' description. The proposed model utilizes modified variant of an existing deep learning architecture, namely VGG16, with OCT images for efficient diagnosis of retina disorders. The proposed deep learning architecture modifications aim at improving the performance of the standard basic architectures in terms of attained accuracy and training time requirements. The implemented architectures provide RoI localization and feature generation. Hand-crafted features are generated from the RoI that is extracted by the deep learning architecture. Afterwards, both CNN-based features and hand-crafted features are merged and input to the dense layer. Diagnosis is produced based on the classification output by the dense layer. The proposed system architecture is shown in Figure 4-1.

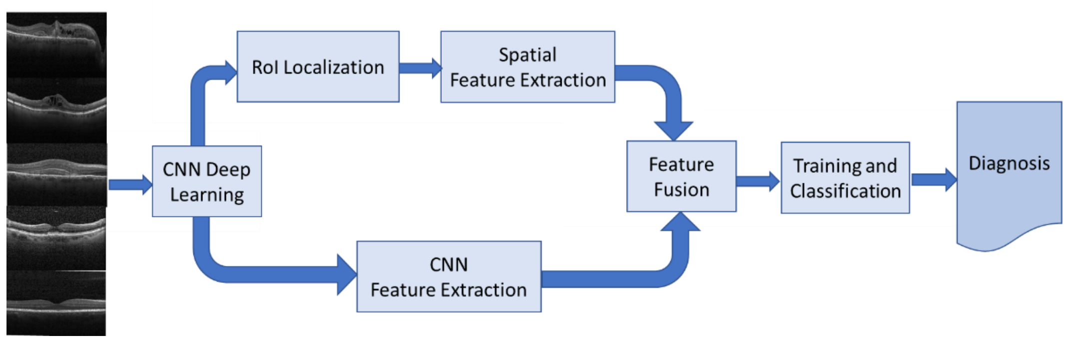


Figure 4-1 Proposed Framework diagram

4.1 Preprocessing Stage

A range of preprocessing steps are applied to the dataset to limit the processing system requirements and to increase the robustness of the models.

First, individual cross-sectional tomography (B-scans) in the SD-OCT volume are denoised using the sparsity-based block matching and 3D-filtering (BM3D) denoising method which reduces the speckle noise. It is the filter from the nonlocal means class that looks for local neighborhoods of similar shapes and puts them into a 3D matrix [62]. The applied BM3D filter noticeably reduces the speckle noise in images, as can be seen in Figure 4-2. Noise reduction helps in achieving good performance metrics.

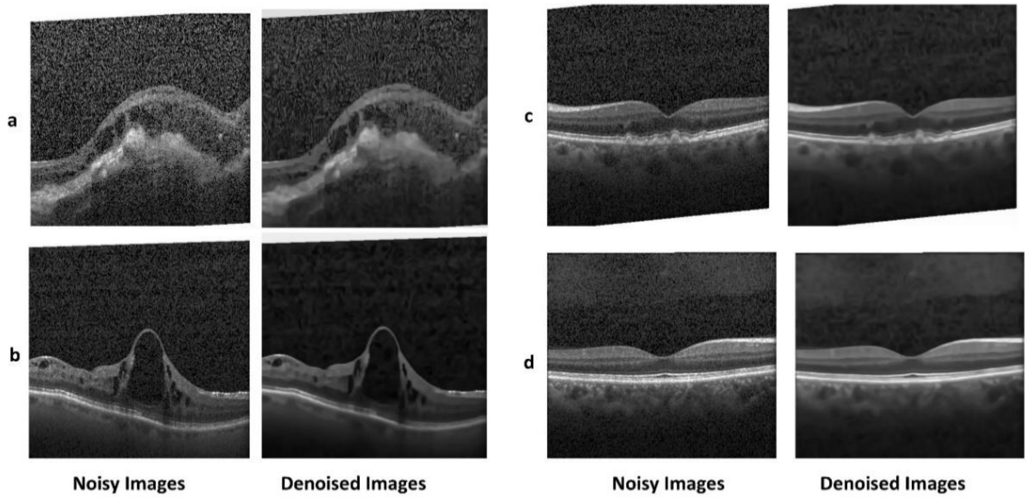


Figure 4-2 Noised vs. Denoised Images (a) CNV class (b) DME class (c) DRUSEN class (d) NORMAL class

BM3D filter and grouping is based on the idea of collaborative filtering that has 4 steps which are [63]:

1. Finding the image patches similar to a given image patch and grouping them in a 3D block
2. 3D linear transform of the 3D block
3. Shrinkage of the transform spectrum coefficients
4. Inverse 3D transformation

This 3D filter therefore filters out simultaneously all 2D image patches in the 3D block. By attenuating the noise, collaborative filtering reveals even the finest details shared by the grouped patches. As shown in Figure 4-3 filtered patches are then returned to their original positions. Since these patches overlap, many estimates are obtained which need to be combined for each pixel [63]. Aggregation is a particular averaging procedure used to take advantage of this redundancy. The first collaborative filtering step is much improved by a second step using Wiener filtering. This second step mimics the first step, with two differences. The first difference is that it compares the filtered patches instead of the original patches [63]. The second difference is that the new 3D group (built with the unprocessed image samples but using the patch distances of the filtered image) is processed by Wiener filtering instead of a mere threshold. The final aggregation step is identical to those of the first step [63].

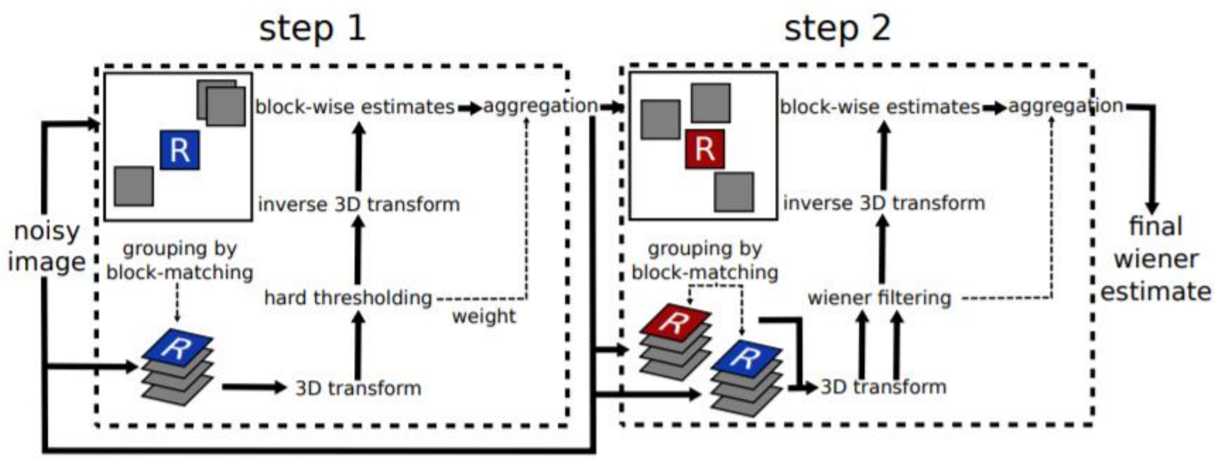


Figure 4-3. Scheme of the BM3D algorithm

Second, images are streamed in batches to avoid system crash due to RAM overloading, and this is because of the memory limitations. Memory limitations come from training models using a RAM size of small bandwidth compared to the large dataset used in training (108,312 images) and the large size of our deployed CNN architecture. To solve this problem, our data is partitioned into batches of 32 images and the training data was transferred batch by batch from the hard disk to RAM during training. After the training of each batch ends, it is discarded, and a new batch of images is generated and stored in RAM. So, each training epoch will have 3385 iterations (equal to number of training dataset/batch size). This technique allows us to train our large training set without RAM overflowing.

Data Augmentation is applied for each batch. Rotation by range from 0 to 30 degrees, horizontal flipping and shifting by range between 0 and 30% from left, right, up and down is applied randomly on each batch of images in each epoch to allow deep learning architectures to train on different orientations of input image, thus improving the robustness of the system. After that, all the images are resized to 200 × 200 pixels.

Each pixel value is normalized to process all images in the same manner because some images may have high pixel range that could cause stronger loss and low pixel range that could cause weaker loss. Additionally, high pixel range will have a large number of votes to determine how to update weights. So, if the values of intensities are normalized to avoid discrepancy in image processing caused by different pixel ranges, this could decrease the gap and make a fair competition in votes between high pixel range and low pixel range.

4.2 RoI segmentation stage

The performance of deep learning is noteworthy. Nevertheless, a shortcoming of deep learning architectures is that they are considered as black boxes which do not provide any insight into the classification process. Therefore, in this phase the activation maps are generated by Norm-VGG16 for our classes to visualize and localize the areas used for classification. The activation maps highlight the focus areas, from which most of the CNN-based features were extracted.
In order to generate discriminative localization map, Gradient-weighted Class Activation Mapping (Grad-CAM) module [64] is used. The implemented Grad-CAM module structure is shown in Figure 4-4. Grad-CAM $L^c_{Grad\text{-}CAM} \in R^{u\times v}$ of width u and height v for any class c. The gradient of the score of the class c is computed before the softmax layer (y_c) with respect to feature maps A_k of a convolutional layer $\frac{\partial y^c}{\partial A^k}$ .These gradients flowing back are global average pooled to obtain the neuron importance weights α^c_k as shown in Equation (4-1).

$$\propto_k^c = \overbrace{\frac{1}{Z}\sum_i \sum_j}^{\substack{global \\ average\ pooling}} \underbrace{\frac{\partial y^c}{\partial A_{ij}^k}}_{\substack{gradiants\ via \\ backpropagation}} \qquad (4\text{-}1)$$

This weight α_c^k represents a partial linearization of the deep network downstream from A and captures the “importance” of feature map k for a target class c. After that a weighted combination of forward activation maps takes place. ReLU is applied on the combination of feature maps. ReLU is used to emphasize the positive influencer pixels whose intensity should be increased in order to increase y_c for a certain class and remove the negative influencer pixels that are likely to belong to other categories in the image as shown in Equation (4-2).

$$L^c_{Grad-CAM} = ReLU \underbrace{\left(\sum_k \propto_k^c A^k\right)}_{linear\ combination} \qquad (4\text{-}2)$$

The ReLU helps in highlighting the positive influencer pixels in the localization maps and avoid highlighting the negative influencer pixels, which achieve better localization . The generated heat maps (activation maps) seem to clearly portray the RoIs used in classification, which can be considered as implicit segmentation. Thus, they provide transparency of the model, which is required in medical diagnosis. The generated heat maps for all the four classes are shown in Figure 4-5, where the activation maps clearly mark the RoI.

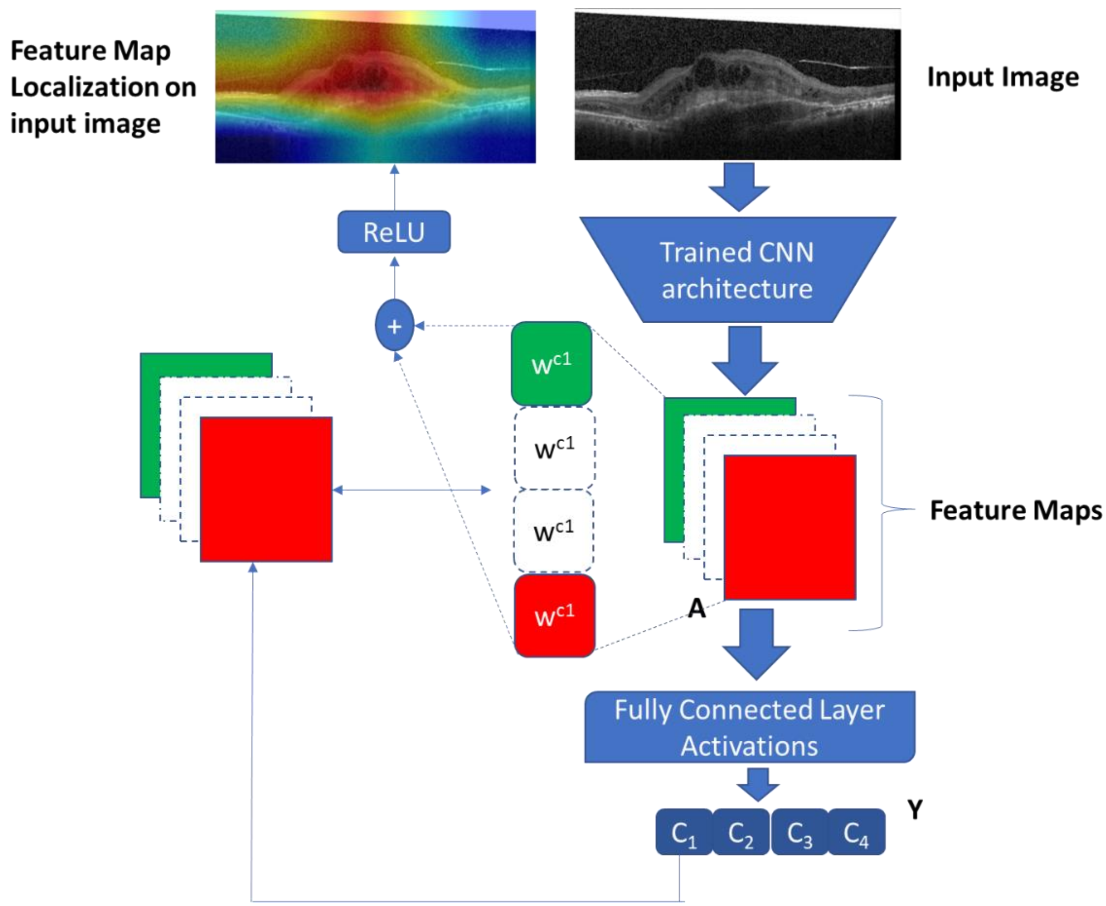


Figure 4-4 Grad-CAM overview

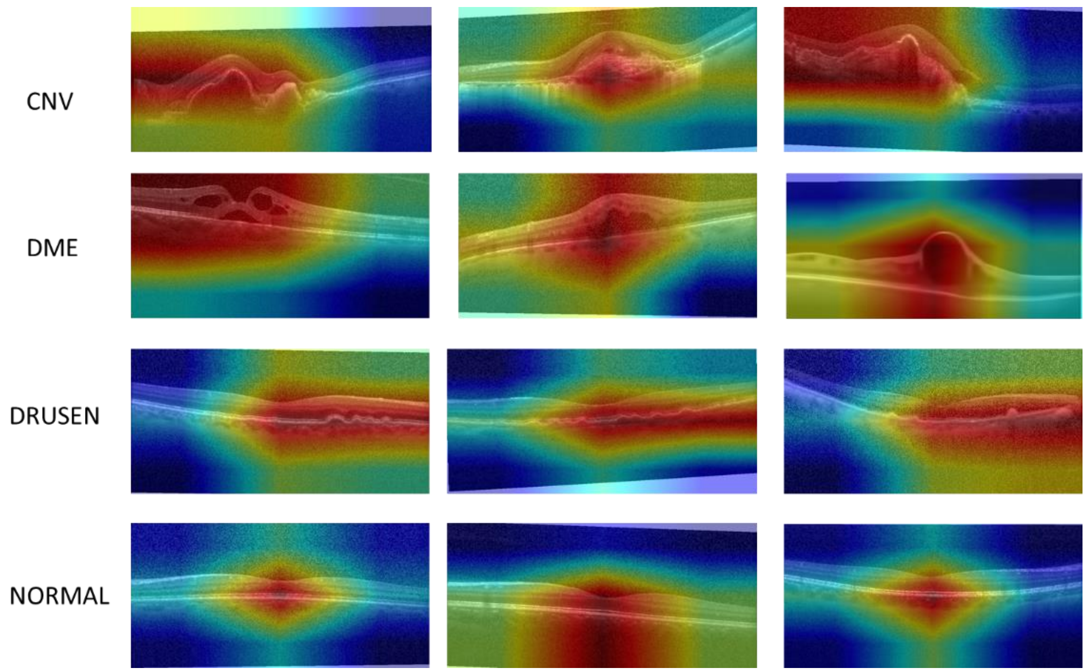


Figure 4-5. Examples of heat maps for each class.

The shown samples of the activation maps clearly mark the RoI that can be used in classification and feature generation. As shown in Figure 4-6, a binary mask is generated for each image from the RoI and the corresponding image is segmented using this binary mask.

The binary masks are generated based on the heat maps output by the Norm-VGG16 deep learning architecture. The heat maps highlight the RoI, which allow the generation of binary masks. Simple global thresholding [65, 66] is used, where a histogram is plotted for each heat map and the respective threshold value is determined. The threshold value is selected per individual heat map such that red and yellow areas in the heat map are used to create the mask (emphasizing the localized RoI region) as shown in Figure 4-7. The global threshold value is used to generate the binary mask.

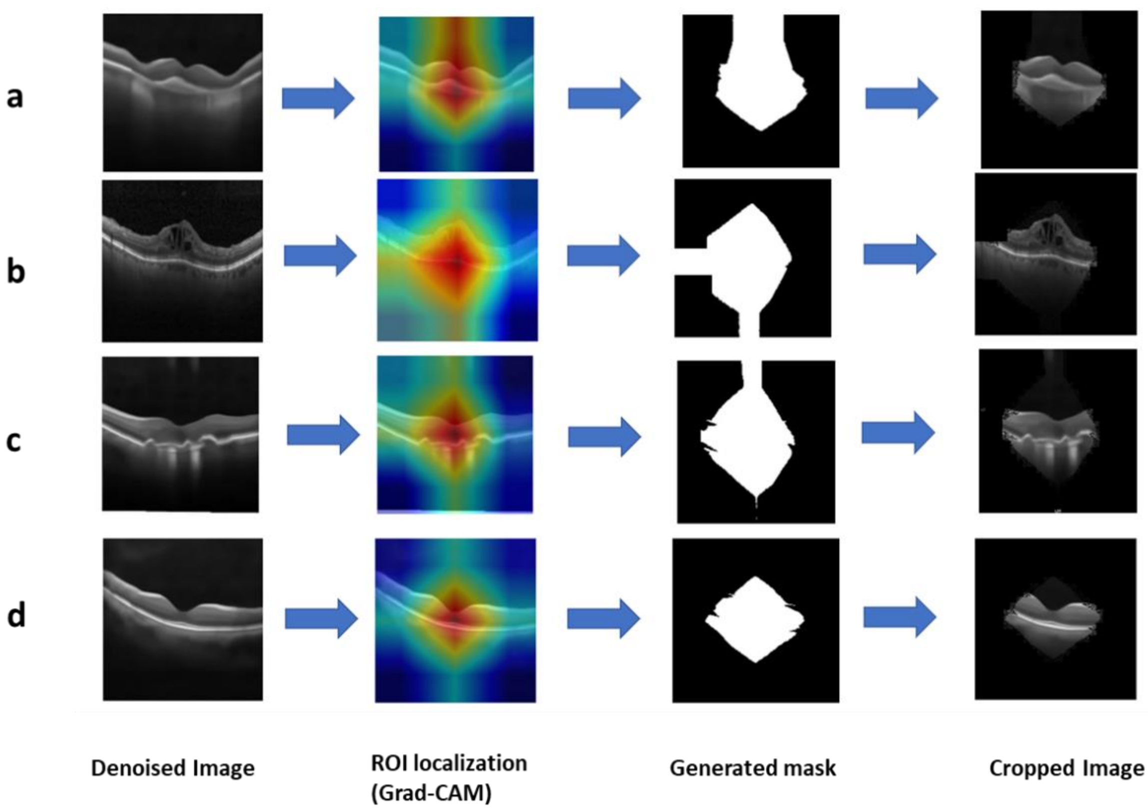


Figure 4-6. Pipeline of segmentation. (a) CNV class (b) DME class (c) DRUSEN class (d) NORMAL class.

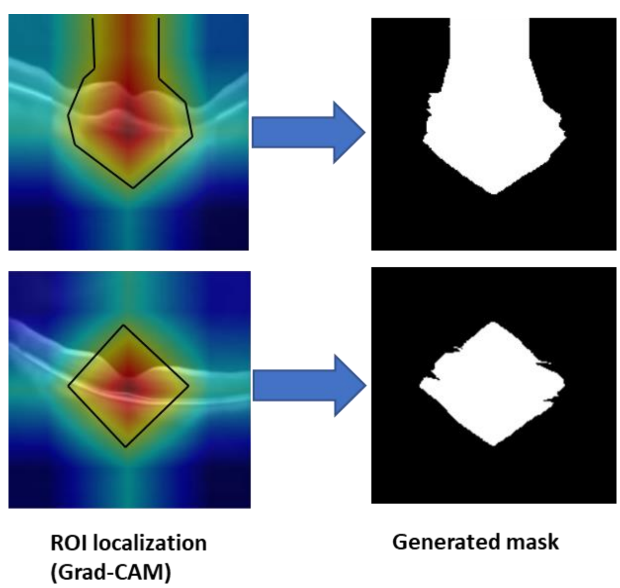


Figure 4-7. Mask Generation from heat map (generated from Norm-VGG16 (kernel regularized)).

4.3 Feature Extraction and Fusion

After RoI localization, features are extracted to be used in detection and classification of different diseases and disorders in medical images. In this phase, different type of features is extracted from the medical images. The features are divided to deep learning features and spatial hand-crafted features.

4.3.1 Deep Learning Features

One of the most prominent deep learning architectures is VGGs Network. VGG16 [56] architecture is utilized for its known superior performance, for example, compared to other models like AlexNet [55]. This may be attributed to its small kernel size (3×3) and more trainable parameters, which helps increase the depth of CNN without the problem of gradient loss in deep CNNs. A set of modifications to the standard VGG network architecture are proposed to create Norm-VGG16 Network variant architecture that enhances the performance of the basic architecture.

First, the number of convolution layers is increased from 13 layers in standard VGG to 16 convolution layers. Each of the added layers has a kernel of size 3×3 and stride = 1. The pipeline of the convolution layers starts with input layer which is convolved by 64 (3×3) kernels. The ending layer of the pipelined convolution layers has 512 (3×3) kernels with output of size 3×3. Each convolution layer is followed by a non-linear layer to trigger function to signal distinct identification of likely features on each hidden layer. Between each group of convolution layers there is a max pooling layer of size (2×2) to down sample the feature tensor to reduce overfitting. The second applied modification is adding a global average pooling layer followed by a softmax dense layer to reduce the overfitting of CNN. Figure 4-8 shows the Norm-VGG16 architecture.

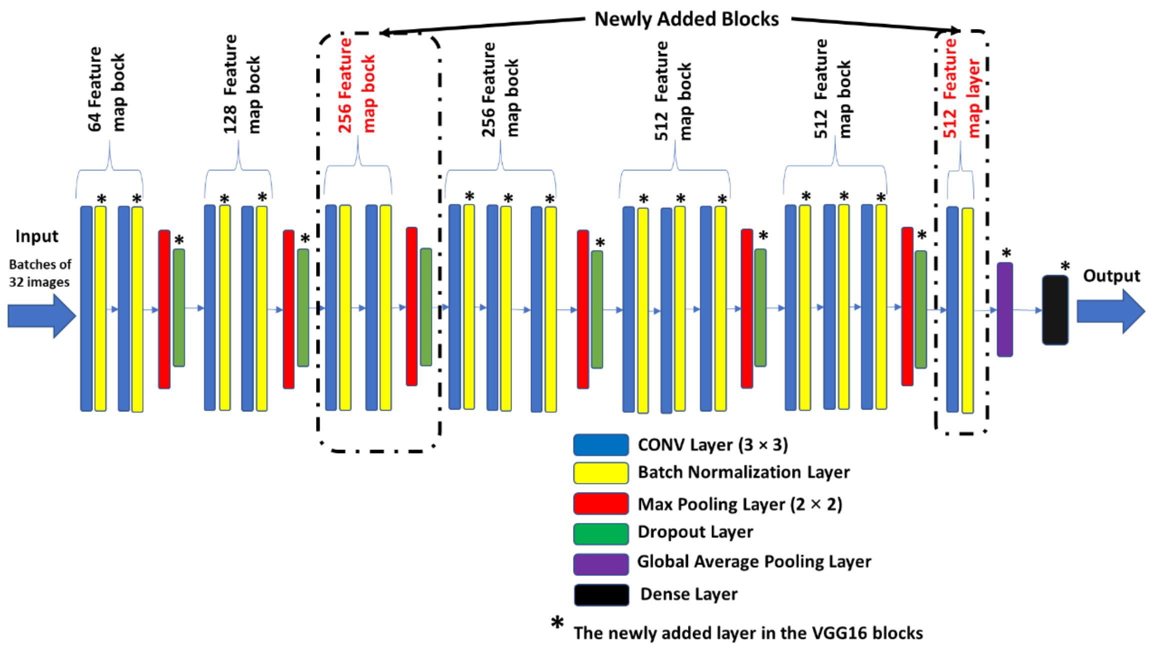


Figure 4-8. Norm-VGG16 proposed architecture

The Input layer of the proposed Norm-VGG16 is modified to have size of 200 × 200 This modification will affect the number of feature maps generated for each layer of the proposed Norm-VGG16 as shown in Figure 4-9.

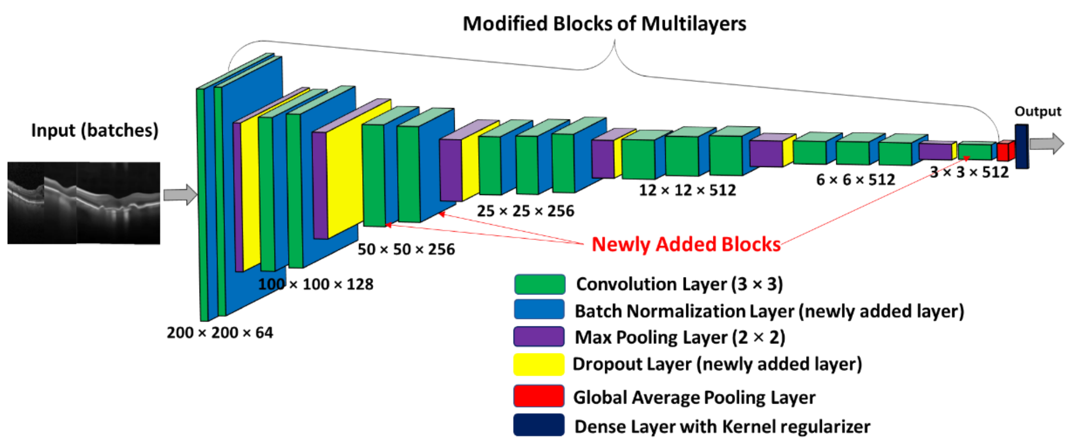


Figure 4-9. Modified Input layer size in Norm-VGG16

a. Convolution layers

Its main role is automatic feature extraction by passing different number of kernels (feature maps) on the input image [67]. The kernel weights change during the training stage and settle at the end of the training stage to be used in the testing stage [67]. Through the training process, each kernel will be convolved with each of the image's color (RGB layers) and compute the dot products between the entries of the kernel and the part of image at any position. The results 2-dimentional feature map gives the response of that filter at every spatial position. These feature maps will be calculated along the image color channels (depth dimension) The process of convolution layers is shown in Figure 4-10. The first convolution layers extract low-level features like edges, lines and corners while the last convolution layers extract higher level features.

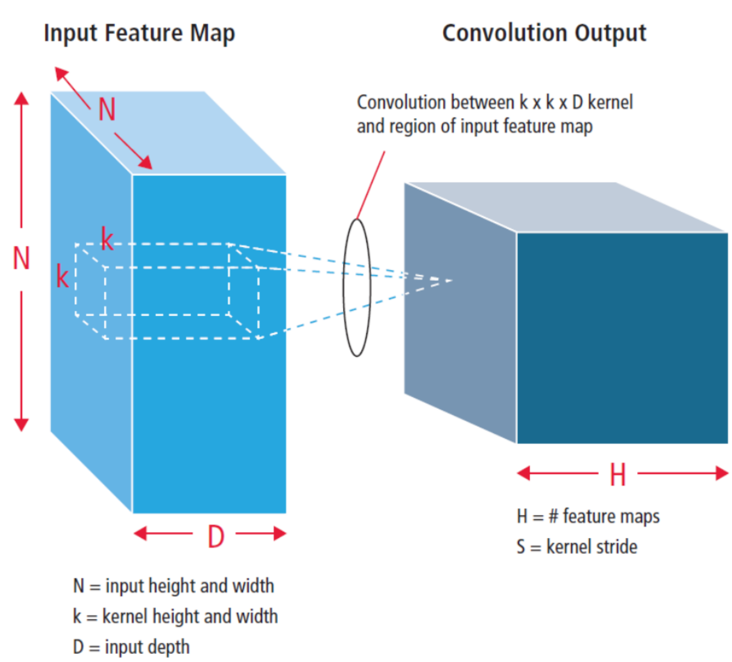


Figure 4-10. process of convolution layer

b. Non-Linear layer

Non-linear layers are used in CNNs to produce a non-linear trigger function to signal distinct identification of likely features on each hidden layer [68]. There are different non-linear layer activation functions for examples sigmoid/tanh and Rectified Linear Unit

(ReLUs). ReLU can be defined as h = max (0, a). In Norm-VGG16, the used activation function is Rectified Linear Unit (ReLUs) because it has advantages over others which are stated as follow [68]:

- *Reduced likelihood of the gradient to vanish:* This arises when a>0. In this case the gradient has a constant value. In contrast, the gradient of sigmoid becomes increasingly small as the absolute value of x increases. The constant gradient of ReLUs results in faster learning.
- *Sparsity:* it arises when a≤0. The more such units that exist in a layer the more sparse the resulting representation. Sigmoid on the other hand are always likely to generate some non-zero value resulting in dense representations. Sparse representations are more beneficial than dense representations.

c. Sub-sampling (Max Pooling / Global Average Pooling) layers

Sub-sampling layers produce a down-sampled version that is robust against noise and distortion [67]. Norm VGG16 uses different types of sub-sampling layers which are max pooling layers and average pooling layers. The max pooling layers consider the highest activation value of a window of size n × n of each feature map [67]. The max pooling layers is used to down sample the feature tensor to reduce overfitting. Figure 4-11 shows the operation of max pooling.

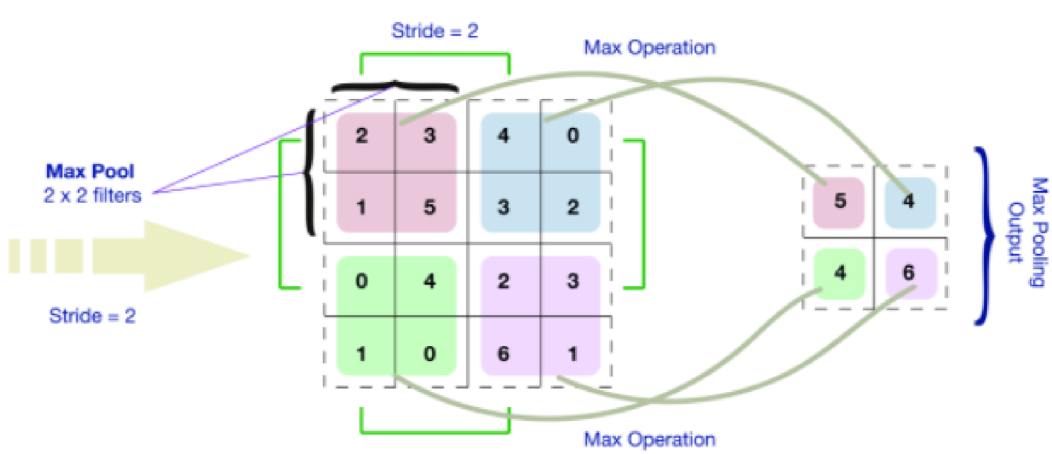


Figure 4-11. Operation of max pooling layer

The global average pooling layer computes the mean value of each feature map and forward it to the softmax in dense layer. The SoftMax in dense layer takes each value and converts it to a probability (with the probability of all digits summing to 1.0) [69]. Global Average Pooling is a pooling operation designed to replace fully connected layers in classical CNNs. The idea is to generate one feature map for each corresponding category of the classification task. Instead of adding fully connected layers on top of the feature maps, we take the average of each feature map, and the resulting vector is fed directly into the softmax layer that normalize the distribution probability of the output classes [70].

One advantage of global average pooling over the fully connected layers is that it is more native to the convolution structure by enforcing correspondences between feature maps and categories. Thus, the feature maps can be easily interpreted as categories confidence maps. Another advantage is that there is no parameter to optimize in the global average pooling thus overfitting is avoided at this layer. Furthermore, global average pooling sums out the spatial information, thus it is more robust to spatial translations of the input [70]. Figure 4-12 shows the operation of global average pooling.

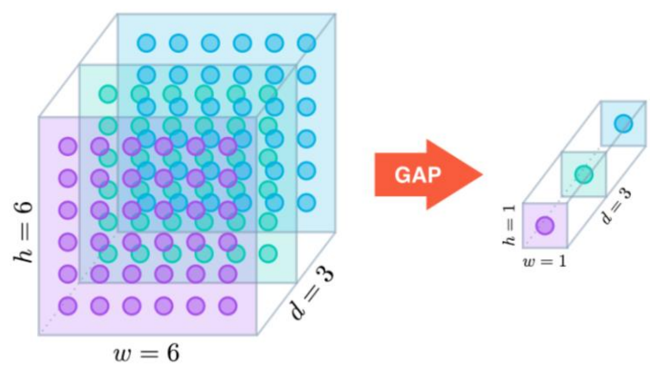


Figure 4-12. Operation of global average pooling layer

d. Batch Normalization and Dropout layers

NormVGG16 is a deep CNN which is prone to overfit training data. Batch Normalization layers and Dropout layers prevent overfitting of Deep CNNs. In Dropout layers, the term “dropout” refers to dropping out units (hidden visible) in a neural network. Dropping a unit out means it is temporarily removed from the network, along with all its incoming and outgoing connections [71, 72]. The choice of which units to drop is random. Each unit is retained with a fixed probability p independent of other units [71, 72]. Batch Normalization allows using higher learning rates and not to care about initialization. It also acts as a regularizer and helps dropout layers in avoiding overfitting [71, 72].

e. Kernel Regulaizers

Another enhancement to the standard architecture is adding L2 kernel regularization to limit overfitting. The added regularization is shown in Figure 4-8 at the last dense layer. Kernel regularization was first added in different layers, but the best performance is attained by the variant scenario implementing L2 kernel regularization at the dense layer (last layer). The value of the hyperparameter of L2 regularizer, which determines the relative importance of the regularization component compared to the loss component, is set to 0.01. L2 regularizer puts a constraint on the complexity of a network by giving a small value to its weights, which makes the distribution of weights regular. This is done by the increasing loss function of the network, which has large weights [73]. To explain how the regularizer works, a training function ŷ: f(x) should be first defined as a function that maps an input vector x to output ŷ where ŷ is predicted value for actual value y. The loss between ŷ (f(x)) and y can be defined for one sample xi with corresponding target y_i. Loss (L) can be computed as $L((y_i), \hat{y}_i) = L(f(x_i), y_i)$ [74]. For all input samples $x_i \ldots\ldots x_n$. The sum of all loss functions between each input xi and its corresponding output ŷ is minimized as given in Equation (4-3).

$$\min_{\mathrm{f}} \sum_{i=1}^{n} L(f(x_i), y_i) \tag{4-3}$$

The regularizer (R(f)) is added to the loss equation to increase the training loss to decrease overfitting as shown in Equation (4-4).

$$L(f(x_i), y_i) = \sum_{i=1}^{n} L_{losscomponent}(f(x_i), y_i) + \lambda R(f) \tag{4-4}$$

where λ is a hyperparameter that determines the relative importance of the regularization component compared to the loss component [74]. We choose our λ = 0.01. From Equation (1) and Equation (2), minimization after adding the regularization would occur to both the component of loss and the regularization component. Different values for λ are tested in the range between 0 and 1 (0.001, 0.01, 0.02, 0.1, etc.) according to the commonly used values in literature [73, 74]. The best results are obtained at λ = 0.01.

The L2 regularization is a type of regularization that is proportional to the square of the value of the weight coefficients (the L2 norm of the weights) [75] as shown in Equation (4-5).

$$R(f) = \sum\nolimits_{f\ i=1}^{n} w_i^2 \tag{4-5}$$

By substituting Equation (3) into Equation (2), L2 regularization Equation (4-6) is generated

$$L(f(x_i), y_i) = \sum_{i=1}^{n} L_{losscomponent}(f(x_i), y_i) + \lambda \sum_{i=1}^{n} w_i^2 \tag{4-6}$$

4.3.2 Spatial hand-crafted features

After RoI localization and cropping, a set of spatially articulated features are generated. HOG and DAISY feature descriptors were generated.

HOG descriptors are from global features family that generates a compact texture features, but they are most sensitive to clutter and occlusion [76]. On the other hand, DAISY descriptors are from local features family which generates key descriptors that are calculated in multiple interest points of local image and are not sensitive to clutter and occlusion [76]. The key point of extracting both, HOG and DAISY descriptors, is to combine different information of different families of features, which is expected to improve the results [76].

HOG descriptors are calculated by normalizing colors then dividing the image into blocks and each block is divided into smaller units called cells. Each cell contains number of pixels (pixel intensities). First, the gradient magnitude and direction of each cell's pixel intensities is calculated. If (x, y) is assumed as a pixel intensity, then gradient magnitude is calculated from Equation (4-7) and gradient angle is calculated using Equation (4-8) [76].

$$G(x, y) = \sqrt{\left(G_x(x, y)^2 + G_y(x, y)^2\right)} \tag{4-7}$$

$$\theta(x, y) = \arctan\left(\frac{G_{y(x,y)}}{G_{x(x,y)}}\right) \tag{4-8}$$

After calculating magnitude and angle, the HOG is calculated for each cell by calculating the histogram. Q bins for angles are chosen with unsigned orientation angles between 0 and 180. Afterwards, normalization is applied since different images may have different contrasts [40]. The pipeline of HOG can be shown in Figure 4-13. In our implementation, a [4 × 4] cell size, [2 × 2] cells per block and 9 orientation histogram bins creating 8100 features were used.

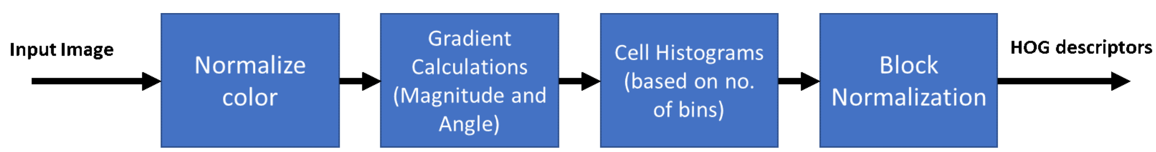


Figure 4-13. Histogram of Oriented Gradients (HOG) descriptor pipeline

DAISY feature descriptor is an algorithm that generates low dimensional invariant descriptors that convert local image regions into low dimensional invariant descriptors, which can be used for matching and classification. DAISY feature descriptor is faster than GLOH and SIFT feature descriptors [77]. It can be computed efficiently at every pixel unlike SURF. For DAISY descriptors generation, eight orientation maps, G, are computed for each image. one for each quantized direction, where G (u, v) equals the image gradient at location (u, v) [77].

First, orientation maps are computed from the original images, which are then convolved to obtain the convolved orientation maps $G_o^{\Sigma i}$. as show in Figure 4-14. Gaussian kernels of different $\sum$ values convolve each orientation map several times to obtain convolved orientation maps for different sized regions [77].

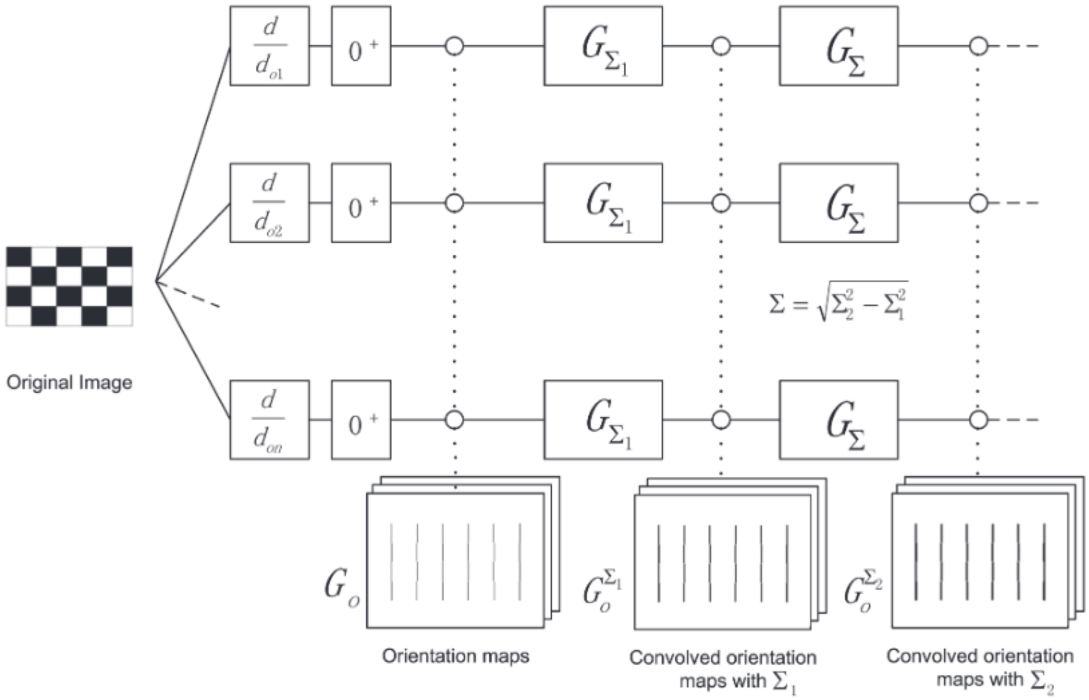


Figure 4-14. The calculation process of the convolved orientation maps in DAISY

If h∑ (u, v) is the vector made of the values at location (u, v) in the orientation maps after convolution by a Gaussian kernel of standard deviation Σ as shown in Equation (4-9), after that, the If h∑ (u, v) is normalized.

$$h_{\Sigma} \quad (u, v) = \left[G_1^{\Sigma} \quad (u, v), \dots \dots \dots, G_H^{\Sigma} \quad (u, v) \ \right]^T \qquad \text{(4-9)}$$

where H refers to number of orientation maps and G_1^{Σ}, G_2^{Σ}……G_H^{Σ} denote the *Σ-convolved orientation maps*. The vector is normalized and denoted by $\hat{h}_{\Sigma}(u, v)$. To correctly represent the pixels near occlusions, the normalization is performed in each histogram independently because normalizing the descriptor as a whole would lead to the same point, which is close to occlusions, appearing different in two images. The full DAISY descriptors D (u_0, v_0) for location (u_0, v_0) is then defined as a concatenation of $\hat{h}$ vectors as shown in Equation (4-10)

$$D(u_0, v_0) = \begin{bmatrix} \hat{h}^T_{\Sigma\,1}\,(u_0, v_0), \\ \hat{h}^T_{\Sigma\,1}\,\big(I_1(u_0, v_0, R_1)\big), \dots\dots\dots\, \hat{h}^T_{\Sigma\,1}\,\big(I_N(u_0, v_0, R_1)\big), \\ \hat{h}^T_{\Sigma\,2}\,\big(I_1(u_0, v_0, R_2)\big), \dots\dots\dots\, \hat{h}^T_{\Sigma\,2}\,\big(I_N(u_0, v_0, R_2)\big), \\ \hat{h}^T_{\Sigma\,Q}\,\big(I_1(u_0, v_0, R_3)\big), \dots\dots\dots\, \hat{h}^T_{\Sigma\,Q}\,\big(I_N(u_0, v_0, R_3)\big), \end{bmatrix}^T \tag{4-10}$$

Where Q is the number of convolved orientation layers with different ∑'s and I_j (u, v, R) is the location with distance R from (u,v) in the direction given by j when the directions are quantized into N values. Figure 4-15 shows the explanation of DAISY descriptor where each circle represents a region where the radius is proportional to the standard deviations of the Gaussian kernels and the '+' sign represents the locations where the convolved orientation maps center are sampled as a pixel location where we compute the descriptor.

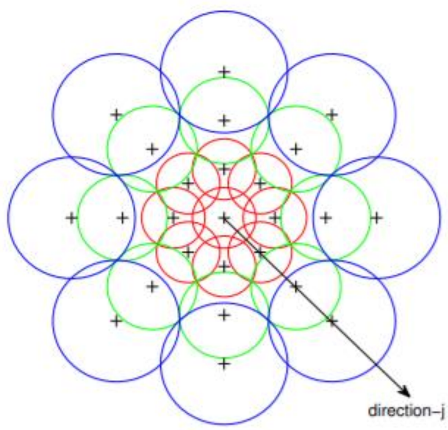


Figure 4-15. The explanation of DAISY descriptor.

In DAISY, a circular grid is used instead of SIFT's regular one since it has been shown to have better localization properties [77]. So, DAISY descriptor is closer to GLOH before PCA than to SIFT [77]. Also, the descriptor is naturally resistant to rotational perturbations as well by the use of Gaussian kernels with a circular grid. The overlapping regions ensure a smooth changing descriptor along the rotation axis and by increasing the overlap, we can further increase the robustness up to a certain point.

One advantage of the circular design and using symmetric kernels is that the descriptor can be computed orientation simply rotating the grid without recalculating the convolved orientation maps [77].

4.3.3 Feature Fusion

In this stage, spatial hand-crafted feature descriptors of global and local features are extracted from RoI medical images and fused with the automatic features generated from modified CNN (NormVGG16). The complex networks of CNN (Norm-VGG16) obtains generic features, to some extent. The idea behind CNN is to discover multiple levels of representation so that higher level features can represent the semantics of the data, which in turn can provide greater robustness to intra-class variability [78]. Moreover, HOG descriptors are from global features family that generates a compact texture features, but they are most sensitive to clutter and occlusion [76]. On the other hand, DAISY descriptors are from local features family which generates key descriptors that are calculated in multiple interest points of local image and are not sensitive to clutter and occlusion [76]. The key point of extraction of HOG descriptors, DAISY descriptors and automatically generated features from Norm-VGG16, is to combine different information of different families, which is expected to improve the results [76]. The benefit of feature fusion is the detection of correlated feature values generated by different algorithms [79]. The fusion of features of different properties and families creates a compact set of salient features that can improve robustness and accuracy of classification model [79]. In this stage, the global and local features handcrafted features which are 8100 descriptors of HOG and 400 DAISY descriptors of are extracted from RoI medical images and fused with the 512 automatic features generated from modified CNN (NormVGG16) and used as input for classification stage.

4.4 Classificaiotn

Classification is considered an instance of supervised learning where a training feature set are correctly identified samples is available and used for training the classifier. The new feature sample is then tested to identify to which set of categories it belongs.

During the learning stage, the classifier is trained to learn the differences between the known classes on the basis of input feature vectors of known labels. Based on the input feature vectors, the classifier creates a “unique‟ description of each predefined classification category. The input feature vectors from each class are referred to as “training sets‟. When learning is completed, the classifier is ready to conclude classification-labels on new data that haven't been used for its training.

A trained model is built to discriminate between different features of different diseases that were extracted from different medical images. Two types of classifiers, which are Artificial Neural Network and Support Vector Machine, are used to test the degree of correlation between the extracted features and discriminating different disorders and stages of diseases.

4.4.1 Artificial Neural Network (ANN)

ANN can be composed of different number of neurons. The classification of the extracted hybrid features from Norm-VGG16, HOG descriptors and DAISY descriptors is done in batches. A Neural Network consisting of three dense layers is used to train these features with first layer of input size of 9012 inputs followed by a hidden layer and the final layer contains four neurons with a softmax activation to normalize the distribution probability of the output classes. The model is trained using Adam optimizer with batch size 32 for 100 epochs. The starting learning rate used to train the model is 0.001.

ANN training is based on idea of training artificial neuron Training neural network is based on two passes which are feed forward learning and back propagation error learning. In feed forward learning, each neuron is trained on a linear combination of the input variables, $\propto_1$, $\propto_2$, ... $\propto_m$, multiplied with the coefficients, w_{ji}, called 'weights', then a second function serves as a 'transfer function' (ReLU) to produce a non-linear trigger function to signal distinct identification of likely features [80]. Moreover, a bias b_{ji} is added to try to approximate where the value of the new neuron starts to be meaningful. The output, y_j, of jth neuron is calculated according to Equation (4-11)

$$y_j = ReLU\left(\sum_{i=1}^{m} w_{j_i} \propto_{j_i} + bj_i\right) \tag{4-11}$$

In backpropagation, the gradient is calculated efficiently [80].It is used to adjust the weights and biases throughout the network, so that the desired output is generated from the output layer. The back-propagation method has obtained its name due to its learning procedure where the weights of neurons are first corrected in the output layer, then in the second hidden layer (if there is one), and at the end in the first hidden layer, i.e., in the first layer that obtains the signals directly from the input [80]. Let last layer as L, second last layer as L-1, output value to predict as y, cost function as C, weight (connection from one neuron to another) as w, Bias (added value to each activation) as b, z = (w × ∝ + b) and a = ReLU (z). first the cost function for last layer can be calculated as shown in Equation (4-12).

$$C = (a^L - y)^2 \tag{4-12}$$

Three Equations (4-13, 4-14, 4-15) are needed to calculate the relationships between components of neural network.

$$\frac{\partial C}{\partial w^{(L)}} = \frac{\partial C}{\partial a^{(L)}} \frac{\partial a^{(L)}}{\partial z^{(L)}} \frac{\partial z^{(L)}}{\partial w^{(L)}} \tag{4-13}$$

$$\frac{\partial C}{\partial b^{(L)}} = \frac{\partial C}{\partial a^{(L)}} \frac{\partial a^{(L)}}{\partial z^{(L)}} \frac{\partial z^{(L)}}{\partial b^{(L)}} \tag{4-14}$$

$$\frac{\partial C}{\partial a^{(L-1)}} = \frac{\partial C}{\partial a^{(L)}} \frac{\partial a^{(L)}}{\partial z^{(L)}} \frac{\partial z^{(L)}}{\partial a^{(L-1)}} \tag{4-15}$$

Finally, the weight and bias are updated for each layer l as in Equations (4-16, 4-17) respectively.

$$w^l = w^l - learning\,rate \times \frac{\partial C}{\partial w^{(l)}} \tag{4-16}$$

$$b^l = b^l - learning\,rate \times \frac{\partial C}{\partial b^{(l)}} \tag{4-17}$$

ANNs offer a number of advantages which are less formal statistical training, ability to implicitly detect complex nonlinear relationships between dependent and independent variables.

The main disadvantages of ANN include its "black box" nature and its greater computational burden.

4.4.2 Support Vector Machine (SVM)

Support Vector Machine is a family of nonlinear function-based classification algorithms that can be used to classify binary and multi class feature space. The SVM is a kernel-based approach with a strong theoretical background, which has become an increasingly popular tool for machine learning tasks involving classification and regression. The goal of a Support Vector Machine (SVM) is to find a hyper plane that corresponds to the largest possible margin between the data points of different classes as shown in Figure 4-16.

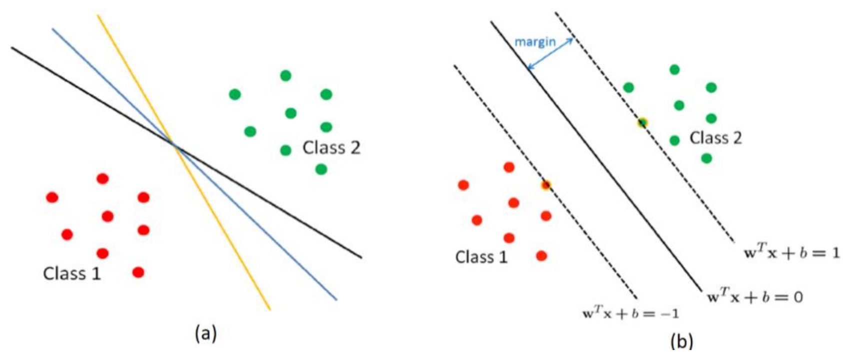


Figure 4-16. SVM hyperplane principles (a) Examples for possible solutions of the two-class classification (b) Illustration of the maximum margin.

First, SVM for a two-class uses a hyperplane for linearly separable classification problem. For two classes A_+ and A_-, SVM can separate them by a pair of parallel bounding planes. The first plane in Equation (4-18). bounds the class A+ and the second plane in Equation (4-19) bounds the class A-. w in Equations (4-18, 4-19) represents the normal vector to these planes and b determines their location relative to the origin. Figure 4-17 illustrates the linearly separable SVM [48].

$$w^T x + b = +1 \tag{4-18}$$

$$w^T x + b = -1 \tag{4-19}$$

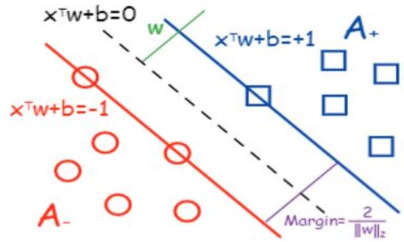


Figure 4-17. Illustration of linearly separable SVM

According to the statistical learning theory, SVM achieves a better prediction ability via maximizing the margin between two bounding planes. Hence, SVM searches for a separating hyperplane by maximizing $\frac{2}{||w||_2}$ [48] . It can be done by means of minimizing $\frac{1}{2}||w||_2^2$ and leads to a quadratic program, as shown in Equation (4-20)

$$\min_{(w,b)\,\varepsilon\,R^{n+1}} \frac{1}{2}||w||_2^2, \qquad s.t. \quad y_i(w^T x_i + b) \geq 1 \; for\; i = 1,2,\ldots.m \tag{4-20}$$

The linear separating hyperplane presented in Equation (4-21) is the plane midway between the bounding planes in Equations (4-18, 4-19).

$$w^T x + b = 0 \tag{4-21}$$

The data points on the bounding planes, in Equations (4-18, 4-19), are called support vectors. If any point which is not a support vector is removed, the training result will not be changed which is one of the SVM characteristics [48]. Moreover, a slack variable is introduced which is called penalty term [48]. It is used to increase performance on testing dataset. The parallel bounding planes is updated as shown in Equations (4-22, 4-23).

$$w^T x_i + b + \xi_i = +1 \quad for \quad x_i \in A_+ \tag{4-22}$$

$$w^T x_i + b + \xi_i = -1 \quad for \quad x_i \in A_- \tag{4-23}$$

Hence, the margin equation between two bounding planes is updated and presented in Equation (4-24)

$$\min_{(w,b)\, \epsilon\, R^{n+1+m}} C\sum_{i=1}^{m} \xi_i + \frac{1}{2}\left\|w\right\|_2^2, \tag{4-24}$$
$$s.t. \quad y_i(w^T x_i + b) + \xi_i \geq 1 \qquad \xi_i \geq 0, for\ i = 1,2, \dots .m$$

Here C >0 is a positive parameter which balances the weights of the penalty term $\sum_{i=1}^{m} \xi_i$ versus the margin maximization term $\frac{1}{2}\left\|w\right\|_2^2$. Higher complexity of the separating hyperplane may cause overfitting leading to poor generalization. The positive parameter C which can be determined by a tuning procedure (where a surrogate testing set is extracted from the training set), plays the role of balancing this trade off [48].

The SVM has high accuracy, robust performance, and low computational load but it is a binary classifier, while many problems we are interested in solving, are multiclass. There are several approaches to a multiclass SVM. One approach involves constructing and combining several binary classifiers. This method is called the “one-against-one” or OAO, which is used in the study. In this method, *k(k–1)* binary classifiers, where k is number of classes, are trained to separate a pair of two classes. To classify a new sample, a class that gains most votes of the binary classifiers is chosen as the final output [48].

4.5 Summary

In this chapter, the main phases of the proposed model had been described in detail. First, a preprocessing phase is explained followed by a detailed explanation of a newly proposed modified version of VGG16 network (Norm-VGG16). The preprocessing was applied to remove the noise from the OCT images and adjust it to be suitable for classification. The proposed Norm-VGG16 was used as a feature extractor and used in RoI localization. Second, different hand-crafted features were extracted from the RoI, these handcrafted features were used to increase the interpretability of the model. Third, the handcrafted features were fused with CNN features to have higher level features that can represent the semantics of the data and increase the robustness of model. Finally, the fused features are classified using two types of classifier that are detailed in fourth section of chapter. In the following chapter, different experiments will be conducted to evaluate the efficiency of the proposed model.

Chapter Five

OCT retinal disorders

Experimentation and Results

5 OCT RETINAL DISORDERS EXPERIMENTATION AND RESULTS

This chapter presents the dataset and experimental setup used in conduction of experiments. Moreover, it presents the results of the experiments that have been conducted to evaluate the accuracy and performance of the proposed HyCAD framework using Large Dataset of Labeled Optical Coherence Tomography (OCT) v3 benchmark dataset [8]. Different experiments have been carried out to validate the efficiency of the proposed HyCAD framework and compared with previously proposed methods in the literature.

5.1 Dataset description

The current study uses a Large Dataset of Labeled Optical Coherence Tomography (OCT) v3 benchmark dataset [8] which consists of 109,312 retinal OCT images. The dataset was selected from 207,130 retinal OCT images for 5319 adult patients taken between 2013 and 2017. The 207,130 retinal OCT images were collected from retrospective cohorts from the Shiley Eye Institute of the University of California San Diego, the California Retinal Research Foundation, Medical Center Ophthalmology Associates, the Shanghai First People's Hospital and the Beijing Tongren Eye Center [8].

Two retinal specialists provided the class labels for the images. Four class labels were available CNV, DME, DRUSEN, NORMAL) as shown in Figure 5-1. Furthermore, only 109,312 out of 207,130 images were clearly distributed between these classes by the agreement of the two retinal specialists and formed the dataset while the other images contained severe artifacts. The dataset is divided into 108,312 images as training set and 1000 images as testing set, according to Li et al. [37]. The distribution of the classes within the training set is illustrated in Table 5-1, where both classes: DME and DRUSEN, present a small percentage of the training set and can be considered to impose the challenges of minority classes on our classification problem [8]. The testing set is sampled to be of equal distribution for all classes as illustrated in
Table 5-2.

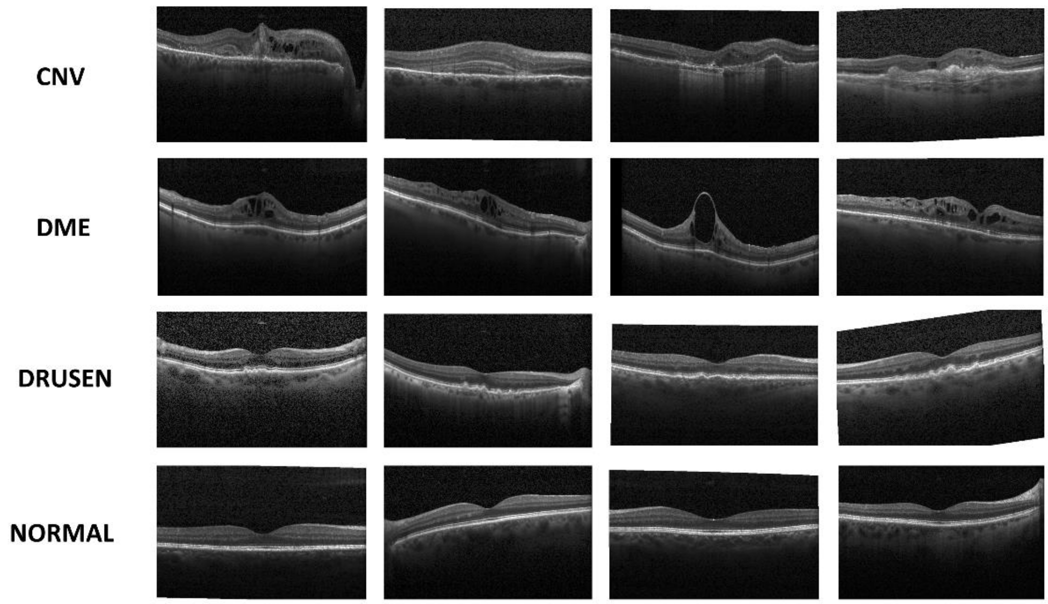


Figure 5-1. Dataset classes (CNV, DME, DRUSEN and NORMAL).

Table 5-1. Training dataset class distribution

Training Dataset				Total
Urgent Referrals Group		Non-urgent Referals Group		
CNV	DME	DRUSEN	NORMAL	
37,206	11,349	8617	51,140	108312
34.35%	10.48%	7.96%	47.22%	100%

Table 5-2. Testing dataset class distribution

Testing Dataset				Total
Urgent Referrals Group		Non-urgent Referrals Group		
CNV	DME	DRUSEN	NORMAL	
250	250	250	250	1000
25%	25%	25%	25%	100%

The classes are categorized as Urgent referrals and Nonurgent referrals groups. The CNV and DME classes represent the Urgent referrals group, where patients need to be transferred to an ophthalmologist in time for definitive anti-VEGF treatment [37, 42]. The referral is urgent in this case because if the treatment is delayed, it would almost cause irreversible vision impairment and even lead to blindness. The Nonurgent referrals group, represented by DRUSEN and NORMAL classes, does not need immediate Urgent referral from an ophthalmologist [37, 42].

5.2 Experimental Setup

Models implementation, training and results are done using Python language v3.7.6 (designed by Guido van Rossum and developed by Python Software Foundation, this version is released on December 2019) with Numpy v1.18.4 (Travis Oliphant), Tensorflow v2.1.0 (Google Brain team), Keras v2.3.1 (François Chollet) and OpenCV v4.2.0 (Intel Corporation, Willow Garage, Itseez) packages Experiments are conducted on Google Colab (free cloud service hosted by Google). Google Colab consists of 1xNvidia Tesla K80, 12.6 GB RAM and 320 GB Disk space.

All deep learning architectures are trained for 30 epochs from scratch using Adam optimizer with starting learning rate of 0.001 on the 108,312 training images. Inputs are divided into batches of size 32. Validation accuracy and cross-entropy loss are monitored for each epoch. In addition, the learning rate is reduced by factor of 0.2 for each three epochs without improvement in validation loss. For the deep learning architectures, the best model is defined as having minimum validation loss, then it is stored and applied on the testing set.

5.3 Results and Discussion

5.3.1 Performance Metrics

The following measures are used to evaluate the performance of trained models to decide which one has advantage over the others.

5.3.1.1 Confusion Matrix

It is a table that is used to visualize and describe the performance of a classification model on a testing set of data as shown in Figure 5-2. It is a summary of prediction results in a classification problem [81]. In confusion matrix, values of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) are represented by assuming C_i one of the four classes in our dataset.

- TP (C_i) = All the instance of C_i. that are classified as C_i.
- TN (C_i) = All the non-C_i. instances that are not classified as C_i.
- FP (C_i) = All the non-C_i. instances that are classified as C_i.
- FN (C_i) = All the C_i instances that are not classified as C_i.

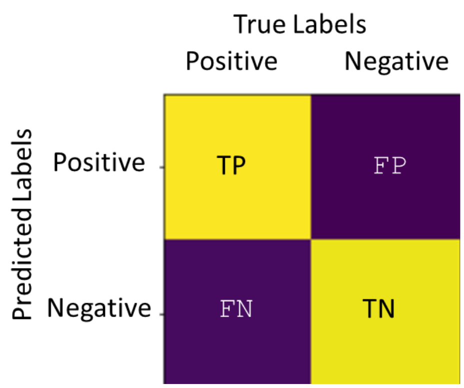


Figure 5-2. Confusion Matrix

5.3.1.2 Accuracy (Acc)

It is measured by dividing the number of correctly labeled images by the total number of test images [81]. Equation (6-1) depicts single class accuracy measurements

$$Accuracy = \frac{TruePositive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \qquad (6\text{-}1)$$

5.3.1.3 Sensitivity (Recall)

It is determined by dividing the total number of correctly classified Urgent referrals by total number of actual Urgent referrals [81]. Equation (6-2) depicts single class sensitivity measurements

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (6\text{-}2)$$

5.3.1.4 Specificity (Sp)

It is determined by dividing the total number of correctly classified non-referrals by total number of actual Nonurgent referrals [81]. Equation (6-3) depicts single class specificity measurements.

$$Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive} \quad (6\text{-}3)$$

5.3.2 Experimental Results

In this section, the implemented experimental scenarios used for performance comparison and the corresponding results using the described performance evaluation metrics are described.

5.3.2.1 Experimental Scenarios

1. ***HyCAD Architecture***

The proposed system architecture presented in chapter 6 is applied to the Large Dataset of Labeled Optical Coherence Tomography (OCT) v3 benchmark [8]. Norm-VGG16 integrating kernel regularizer is trained on the dataset from scratch aiming at higher performance metrics. It is used for RoI generation and feature extraction. A set of 512 features are extracted from the global average pooling layer of the CNN network. The hand-crafted feature extraction methods, namely HOG and DAISY, when applied on the segmented RoI generated 8500 features. At the fusion stage of our HyCAD system, all the extracted features are fused and fed into different classifiers which are SVM and a three sequential dense layer neural network (ANN) for a classification decision.

2. ***The proposed Norm-VGG16 Deep Learning Architecture***

In this scenario, the modified Norm-VGG16 Deep Learning architecture is applied on the OCT dataset. The performed classification is based solely on the CNN-based features.

This scenario is presented to elucidate the merit of incorporating human articulated features on the performance of the system.

3. ***ResNet-50 Net (kernel regularized) Deep Learning Architecture***

ResNets [58, 82] can be considered one of the most applied deep learning architectures for image recognition and classification. ResNets include residual blocks which allow deeper network architecture to avoid information loss during training. In addition, it was shown to achieve lower validation loss compared to VGG16 on the ImageNet dataset [44]. Therefore, ResNet-50 architecture (shown in section 3.2.2.1) is chosen as one of our experimental scenarios, to compare its performance to the proposed HyCAD system. The ResNet-50 best model is modified by adding kernel regularization in the final dense layer.

In addition to these implemented experimental scenarios, the results of HyCAD are compared with the work of Kermany et al. [42] and Li et al. [37]. These were chosen as their results were reported on the same dataset using deep learning approaches and attained high results.

Our experimental results are shown in two folds. First, bootstrapping is adopted, and a limited set of results are reported to show the performance of HyCAD across several testing data partitions and investigate its stability relative to Norm-VGG16 and RESNET-50 and to clarify the impact of hand-crafted features' fusion with deep learning architectures. In the second fold of experiments, elaborate results are reported on the same training and test percentages used by Kermany et al. [42] and Li et al. [37]. The results of the best run are displayed to be able to compare to Kermany et al. [42] and Li et al. [37].

5.3.2.2 Bootstrapping Experiments Fold

Ten bootstrapping experiments are conducted to evaluate the performance of the models at different boot strapped test partitions. In each experiment, the 109,312 images are shuffled and a random selection of 108,312 images is used as training dataset and the rest of the 1000 images are used as testing dataset. Each new partition tested on our implemented architectures (HyCAD, Norm-VGG16 integrating kernel regularizer and ResNet-50 integrating kernel regularizer) taking into consideration that the testing dataset is partitioned equally between four classes (CNV, DME, DRUSEN and NORMAL).

Table 6-3 presents the mean values and the standard deviation of overall accuracy and urgent group sensitivity for 10 different bootstrapped partitions. The depicted results reveal the improvement achieved through integrating the handcrafted features with Norm-VGG16 integrating kernel regularizer. The proposed HyCAD-ANN classifier model attains the highest mean accuracy and mean sensitivity for the critical Urgent referrals compared to Norm-VGG16 integrating kernel regularizer and ResNet-50. In addition, it attains the lowest standard deviation in comparison to the pure CNN architectures. Compared to Norm-VGG16 integrating kernel regularizer, HyCAD-ANN classifier achieves a substantial increase of 2.9% and 4.9% in terms of mean accuracy and mean sensitivity of Urgent referrals, respectively. In addition, the HyCAD-ANN classifier model has a lower standard deviation relative to the pure deep learning architecture Norm-VGG16 integrating kernel regularizer. Such findings emphasize the positive impact of fusing hand-crafted features with learned features from deep learning architecture.

Table 5-3. 10 different bootstrapping experiments with selection of 1000 test images

Model	Mean Accuracy ± Std	Mean UR Sensitivity ± Std
The proposed HyCAD-ANN classifier model	97.2 ± 1.2	98.1 ± 1.1
The proposed HyCAD-SVM classifier model	96.9 ± 1.7	98.0 ± 1.8
The proposed Norm-VGG16 architecture	94.3 ± 2.2	93.2 ± 4.6

The Training and Validation losses and accuracies for the two deep learning architecture (Norm-VGG16 kernel regularized and ResNet-50 kernel regularized) for each epoch are shown in Figure 5-3. The Norm-VGG16 architecture best model achieves training accuracy of 94.86% and training loss of 15.75%. In addition, it achieves a testing accuracy of 97.3% and testing loss of 8.34%. ResNet-50 achieves a training accuracy of 96.95% and a training loss of 9.02%. In addition, it achieves a testing accuracy of 97% and a testing loss of 8.76%. From the shown learning curves, it is manifest that the performance stabilizes around the 15th epoch, which can lead to reducing the training time considerably.

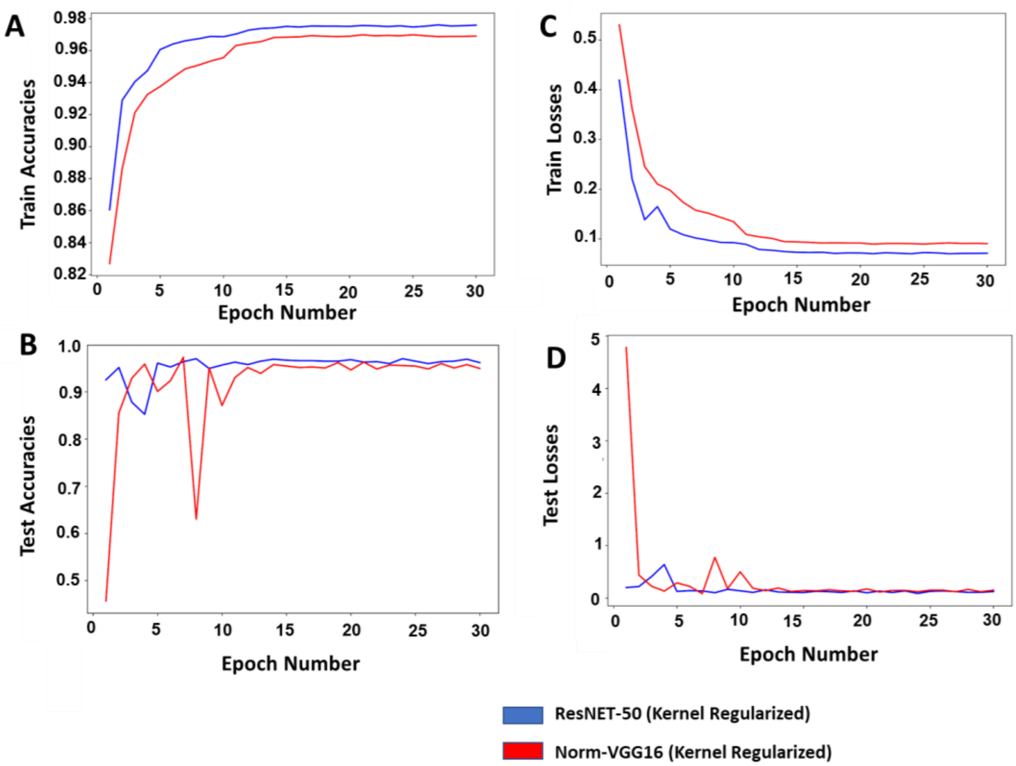


Figure 5-3. Training and Testing Curves for Norm-VGG16 (kernel regularized) and ResNet-50 (kernel regularized) (A) Training accuracy curves (B) Testing accuracy curve (C) Training loss curves (D) Testing loss curve.

The confusion matrices of the best models of the presented experimental scenarios on the test set are presented in Figure 5-4. ResNet-50 (Figure 5-4D) achieves the lowest testing accuracy of 97%. The model's sensitivity for Urgent referral group is 98.6%, while its specificity is 95.4%. The confusion matrix output by applying the best Norm-VGG16 model on the test set is presented in Figure 5-4C. The confusion matrix illustrates that 973 out of 1000 testing images were correctly classified. The model's sensitivity for the Urgent referral group is 98%, while its specificity is 96.6%.

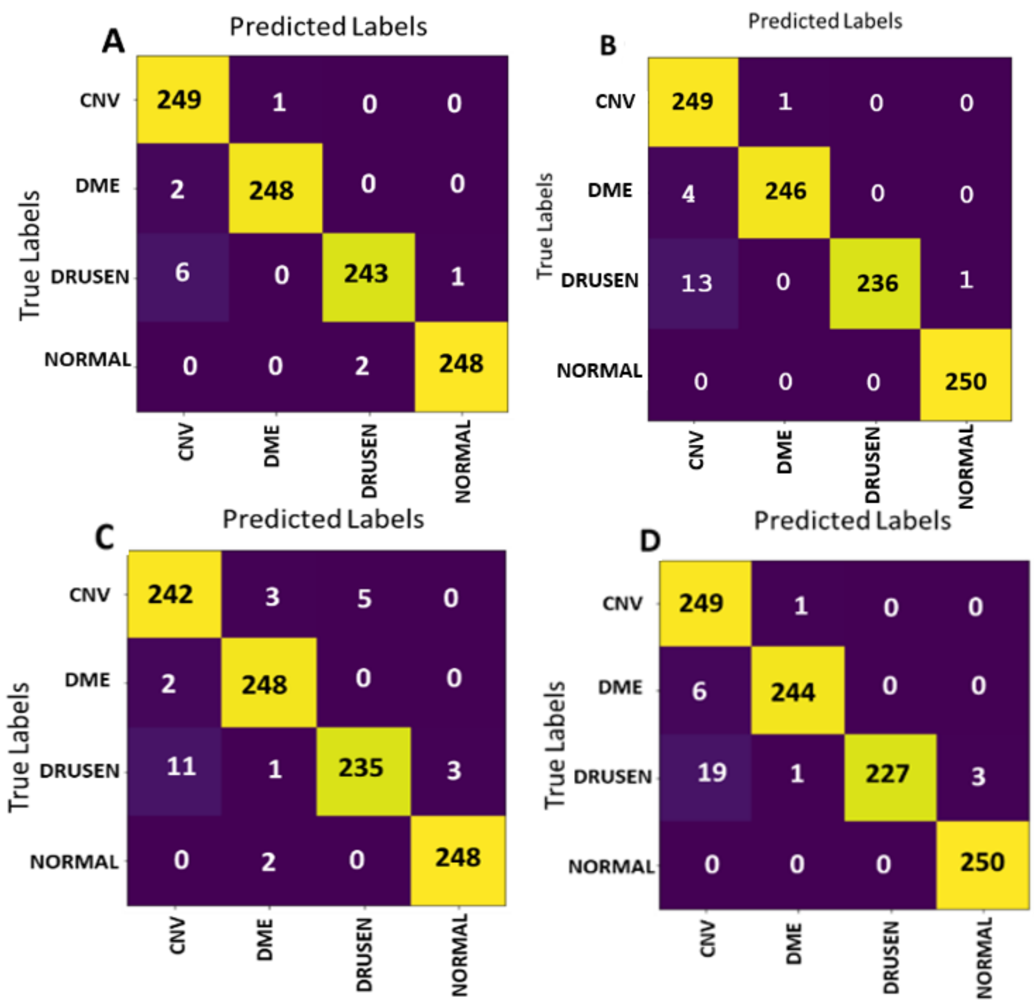


Figure 5-4 Confusion matrices of different proposed models (A) HyCAD-ANN classifier (B) HyCAD-SVM classifier (C) Norm-VGG16 (kernel regularized) (D) ResNet-50 (kernel regularized).

The performance of the models is summarized and compared in Table 5-4. From the shown results, it is determined that the best performing models are the HyCAD models (both version that uses SVM or ANN). The best metrics where achieved by HyCAD-ANN classifier model according to its confusion matrix presented in Figure 5-4A, scoring an accuracy of 98.8%. In Urgent referrals group, the HyCAD-ANN classifier model only misclassified 3 images out of 500 and achieves a sensitivity of 99.4% as shown in Table 5-4. In the Nonurgent referrals group, the same model misclassified only 9 images out of 500 images and achieves 98.2%. It is worth mentioning that the Norm-VGG16 architecture without Kernel regularization attained a testing accuracy of 96.7% when trained on the dataset from scratch.

Table 5-4. Comparison between different implemented architectures.

Model	Accuracy	Sensitivity (UR) (Recall)	Specificity (UR)
ResNet-50 (kernel regularized)	97%	98.6%	95.4%
Norm-VGG16	97.3%	98%	96.6%
HyCAD-SVM classifier model	98.1%	99%	97.2%
HyCAD-ANN classifier model	**98.8%**	**99.4%**	**98.2%**

5.3.2.3 HyCAD Performance Compared to Pure CNNs Performance and State of the Art

The accuracy per class is calculated and compared with previous models of Kermany et al. [42] and Li et al. [37] in Table 5-5. The per class accuracy is given by dividing the truly classified image for each class by their total number. It is noticeable from the accuracies in Table 5 that the lowest accuracy scored by HyCAD (both version that uses SVM or ANN) is of DRUSEN class, which is explainable due to the fact that it has the lowest percentage in the training set. Nevertheless, HyCAD-ANN classifier with the fused hand-crafted features managed to surpass its CNN counterpart Norm-VGG16 in separating the DRUSEN minority class with a difference of 3.2%.

As shown in Table 5-5, the proposed HyCAD-ANN classifier model achieves a significant increase in accuracy of CNV, DME, DRUSEN by 4.8%, 2.4%, 2.8%, while it attains similar NORMAL accuracy compared to Kermany et al. [42]. The proposed HyCAD models achieved an increase in accuracies of CNV class by 3.2% compared to Li et al. [37].

Table 5-5. Accuracy per class comparison between HyCAD proposed models, Kermany et al. and Li et al. models.

Accuracy / Class Model	Urgent Referrals Group		Nonurgent Referrals Group	
	CNV	DME	DRUSEN	NORMAL
Kermany et al. 2018 [42]	94.8%	96.8%	94.4%	98.4%
Li et al. 2019 [37]	96.4%	99.2%	99.2%	99.6%
HyCAD-ANN model	99.6%	99.2%	97.2%	99.2%
HyCAD-SVM model	99.6%	98.4%	94.4%	100%
Norm-VGG16	96.8%	99.2%	94%	99.2%
ResNet-50(kernel regularized)	99.6%	97.6%	90.8%	100%

In order to further examine the performance of our best performing HyCAD (HyCAD-ANN classifier) relative to the state of the art, its overall performance is compared to the work of Kermany et al. [42] and Li et al. [37]. The comparison is depicted in Table 5-6 and Figure 5-5. The values for Kermany et al. [42] and Li et al. [37] are calculated from the best reported confusion matrices as shown in Figure 5-6. The HyCAD-ANN classifier model outperforms the work of Kermany et al. [42] in terms of accuracy, sensitivity and specificity scoring an increase of 2.5%, 1.6% and 0.8% respectively. In addition, compared to Li et al. [37], the proposed HyCAD-ANN classifier model achieves a noticeable increase in sensitivity by 1.8%, a comparable specificity and a slight increase in accuracy. HyCAD-ANN classifer model attains the highest sensitivity, scoring an increase of 1.6% compared to the best model of the state of art models [37, 42] to reach 99.4% for the Urgent referral group. Such a high sensitivity is required for this group as they represent the critical group, which needs immediate attention. This improvement is achieved while maintaining a competitive specificity with the state of the art.

Table 5-6. Performance metrics comparison between Kermany et al., Li et al. models and HyCAD-ANN classifier model performance metrics.

Model	Accuracy	Sensitivity (UR)	Specificity (UR)
Kermany et al.2018 [42]	96.1%	97.8%	97.4%
Li et al. 2019 [37]	98.6%	97.8%	**99.4%**
HyCAD-ANN classifier model (Proposed model)	**98.8%**	**99.4%**	98.2%

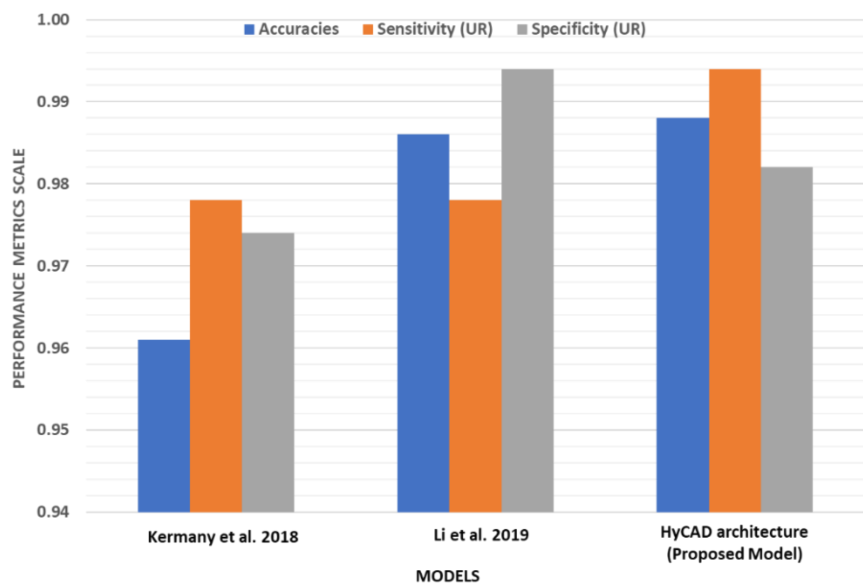


Figure 5-5. Performance metrics comparison between Kermany et al., Li et al. models and HyCAD-ANN classifier.

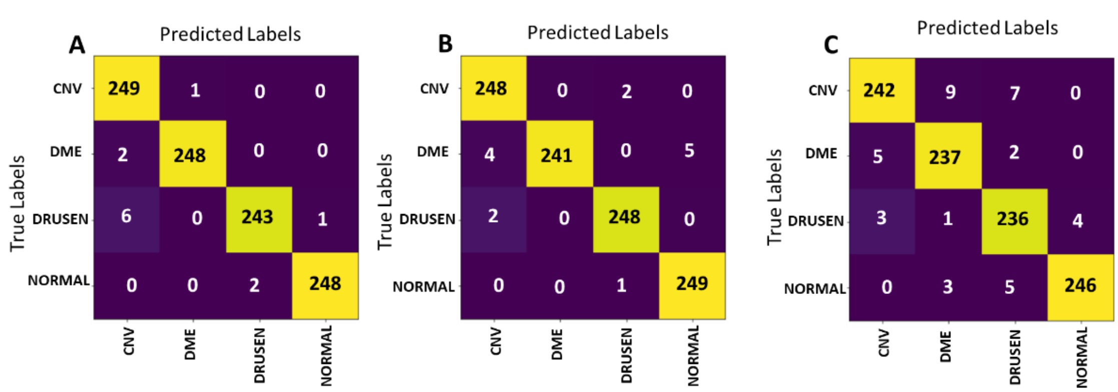


Figure 5-6. Confusion Matrices of proposed model and state-of-the-art models (A) HyCAD-ANN classifier model Confusion Matrix (B) Li et al. model Confusion Matrix (C) Kermany et al. Confusion Matrix.

Another important aspect to study is the performance of our HyCAD-ANN classifier model using hybrid integration of Norm-VGG16 CNN features and hand-crafted features in case of binary classifications. Hence, multiple binary models of CNV vs. NORMAL, DME vs. NORMAL and DRUSEN vs. NORMAL classes are trained and achieve significant results. The respective confusion matrices are shown in Figure 5-7 and the results are compared to Kermany et al. [42] and Li et al. [37] in Table 5-7. The two binary models CNV vs. NORMAL vs. DME and NORMAL achieve an accuracy of 100%, sensitivity of 100% and specificity of 100% without any wrong classification of 500 testing set as can be seen from the confusion matrix in Figure 5-7A,B. The binary model between DRUSEN vs. NORMAL achieves an accuracy of 99.79%, sensitivity of 99.6% and specificity of 100% by classifying only one wrong image as shown in Figure 5-7C. The CNN features are extracted by training Norm-VGG16 for only 15 epochs with learning rate starting with 0.001.

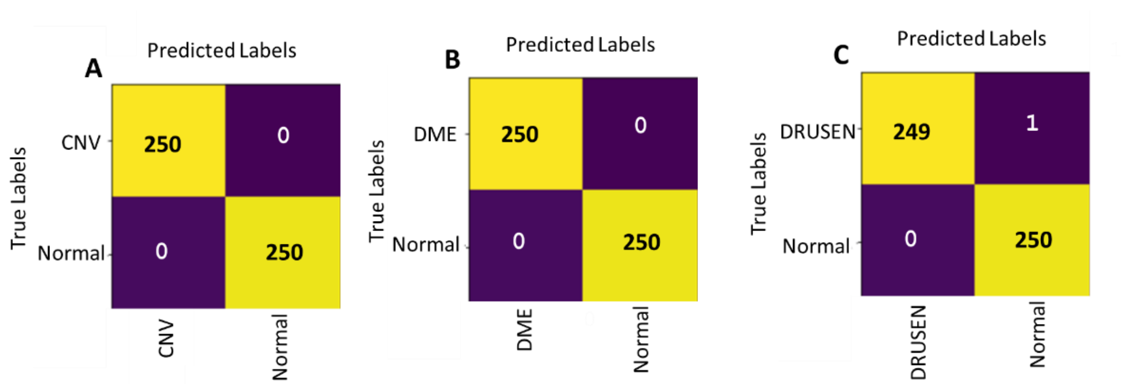


Figure 5-7. Confusion Matrices of HyCAD-ANN classifier binary models (A) CNV vs. NORMAL (B) DME vs. NORMAL (C) DRUSEN vs. NORMAL.

Table 5-7. Performance metrics comparison between binary models of HyCAD-ANN classifier, Kermany et al. and Li et al. models.

Models	**CNV vs. NORMAL**			**DME vs. NORMAL**			**DRUSEN vs. NORMAL**		
	Acc	**Sn (CNV)**	**Sp (CNV)**	**Acc**	**Sn (DME)**	**Sp (DME)**	**Acc**	**Sn (DRUSEN)**	**Sp (DRUSEN)**
Kermany et al. 2018 [42]	100%	100%	100%	98.2%	96.8%	99.6%	99%	98%	99.2%
Li et al. 2019 [37]	100%	100%	100%	98.8%	98.8%	98.8%	99.2%	98.4%	100%
HyCAD-ANN classifier (Proposed Model)	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**99.79%**	**99.6%**	**100%**

5.4 Summary

This chapter started by dataset description followed by identification of the performance metrics used in evaluation of our proposed models and comparing it with the state-of-the-art models. Finally, the chapter described the implemented scenarios and its results. First, bootstrapping experiments were applied on proposed Norm-VGG16, HyCAD with ANN and SVM classifier (kernel regularizer) to test the robustness of the HyCAD model. Second, the proposed Norm-VGG16 and HyCAD models are compared to state of the art models. The winning model of the implemented models was the HyCAD model with ANN classifier (HyCAD-ANN classifier) which achieves an accuracy, sensitivity and specificity of 98.8%, 99.4% and 98.2% respectively. The results of the HyCAD-ANN model had an improvement over Kermany et al. 2018 model [42] in accuracy, sensitivity with 2.7%, and 1.6% respectively. Moreover, the HyCAD-ANN model had an improved sensitivity to urgent referrals compared to Li et al. 2019 model [37] by 1.6%. Finally, binary models are implemented to detect a single retinal disorder and it achieved an accuracy of 100%, 100% and 99.7% for CNV vs Normal, DME vs Normal and DRUSEN vs Normal respectively.

Chapter Six

Conclusion and Future work

6 CONCLUSION AND FUTURE WORK

This chapter briefly summarizes and discusses the work completed, followed by conclusions. The chapter ends up with a discussion of some directions for the future work.

6.1 Conclusion

Medical imaging has a key role in early detection and diagnosis of numerous diseases. It provides direct visualization and investigation to human tissues, organs and bones. It can provide information about any anatomical changes that helps in early detection of diseases. The detection of abnormalities especially in early stages of any disease is subjective to physicians' opinions. Recently, different CAD systems are proposed and used in hospitals to assist and provide a supportive decision for detecting cancers and other different diseases.

In this thesis, a novel hybrid computer aided diagnostic system (HyCAD framework) is introduced that effectively distinguishes a range of retinal disorders from OCT images scans. The proposed HyCAD framework is subjected to different experiments and the results presented in chapter 7 of this thesis, highlight the significance and key contribution of this framework.

The proposed HyCAD architecture integrates deep and traditional learning paradigms to present an accurate timely diagnosis. In addition, HyCAD framework provides an explainable diagnostic decision through automatic RoI localization and the extraction of human articulated features. The automatic RoI localization provides a relative advantage compared to traditional segmentation methods as it does not require expert involvement. Our HyCAD architecture proves the importance of combining automatic feature extraction and hand-crafted features in achieving higher performance metrics. Moreover, the hand-crafted features are more interpretable than the learned features by deep learning architectures models. For the employed deep learning architecture, a set of modifications are applied on the standard VGG16 architecture creating the Norm-VGG16 integrating kernel regularizer variant. Our HyCAD framework achieves significant results compared to state-of-art methods in literature.

Our experiments are conducted on Large Dataset of Labeled Optical Coherence Tomography (OCT) v3 benchmark. The proposed HyCAD model applies noise filtering on the acquired OCT medical images. After that, the RoI is localized using Norm-VGG16 activation maps. Then, the Norm-VGG16 are fused with human articulated features generated from RoI and different classifiers are used to in classification process. Our modified Norm-VGG16 achieved better results than ResNet-50 and Kermany et al. models. HyCAD outperforms the Kermany et al. pure deep learning model in terms of accuracy and Urgent Referrals sensitivity. Similarly, our HyCAD architecture surpassed the Urgent Referrals sensitivity of Li et al. In case of binary classifications of retinal diseases vs. normal, HyCAD had a superior performance compared to both Kermany et al. and Li et al. achieving 100% accuracy, 100% sensitivity and 100% specificity for CNV and DME. For the bootstrapping experiment, HyCAD showed higher mean accuracy and Urgent referrals sensitivity when compared to Norm-VGG16 and ResNet-50. In addition, the results had lower standard deviation.

Overall, a new robust diagnostic system is proposed which offers automatic RoI segmentation that can be used in assisting physicians in their decision with a range of different diseases. Moreover, HyCAD is a general architecture that can be trained and applied on similar problems, since no underlying specific assumptions were made that would hinder its generalization.

6.2 Future Work

An extension to this study would be to implement and examine the performance of the proposed HyCAD system to classify different use cases as classifying different anatomical abnormalities in different imaging techniques as MRI, ultrasounds and mammograms. Another improvement to system can be implemented by using Natural Language Processing (NLP) to combine the reports provided after medical imaging with results of medical image classification to generate more robust decision. Moreover, utilize an adaptive thresholding technique to segment the RoI after Norm-VGG16 localization with Grad-CAM.

Another extension could be modifying the Norm-VGG16 to accept 3D and 4D inputs to try to increase correlation between slices of medical images and to compare its results with current developed architecture.

Also, applying different experiments using different global hand articulated features as Local Binary pattern (LBP) and Gabor Features and show the effect of fusion of different global hand articulated features with each other compared to fusion of automatic CNN features with hand articulated features would be worth further investigation.

Also, applying different feature reduction techniques as Principle Component Analysis (PCA) to reduce the number of features and to compare its results with the current developed architecture.

Finally, a stacking technique could be tried to combine a dynamic local hand articulated features as SIFT to provide more key points than that of static local hand-crafted features as DAISY.

References

REFERENCES

[1] M. T. Cicerone, and C. H. Camp, "21 - Potential Roles for Spectroscopic Coherent Raman Imaging for Histopathology and Biomedicine," *Neurophotonics and Biomedical Spectroscopy*, R. R. Alfano and L. Shi, eds., pp. 547-570: Elsevier, 2019.

[2] C.-M. Chen ,Y.-H. Chou, N. Tagawa, and Y. Do, “Computer-Aided Detection and Diagnosis in Medical Imaging,” *Computational and Mathematical Methods in Medicine,* vol. 2013, pp. 790608, 2013/09/01, 2013.

[3] H. Molly Subhash, and R. Wang, “Optical Coherence Tomography :Technical Aspects,” *Biomedical Optical Imaging Technologies, Biological and Medical Physics, Biomedical Engineering. ISBN 978-3-642-28390-1. Springer-Verlag Berlin Heidelberg, 2013, p. 163*, pp. 163, 08/01, 2013.

[4] K. Doi, “Computer-aided diagnosis in medical imaging: historical review, current status and future potential,” *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society,* vol. 31, no. 4-5, pp. 198-211, Jun-Jul, 2007.

[5] W. H. O. (WHO). Accessed: 17 September 2020 ;https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment.

[6] S. Naz, T. Hassan, M. U. Akram, and S. A. Khan, "A practical approach to OCT based classification of Diabetic Macular Edema." pp. 217-220.

[7] T. Hassan, M. U. Akram ,B. Hassan, A. Nasim, and S. A. Bazaz, "Review of OCT and fundus images for detection of Macular Edema." pp. 1-4.

[8] D. S. Kermany, K. Zhang, and M. Goldbaum, "Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images”, Mendeley Data, v3," 2018.

[9] R. European Society of, “The future role of radiology in healthcare,” *Insights into Imaging,* vol. 1, no. 1, pp. 2-11, 2010/01/01, 2010.

[10] J. G. Fujimoto, C. Pitris, S. A. Boppart, and M. E. Brezinski, “Optical coherence tomography : an emerging technology for biomedical imaging and optical biopsy,” *Neoplasia (New York, N.Y.),* vol. 2, no. 1-2, pp. 9-25, Jan-Apr, 2000.

[11] D. S. Aumann S., Fischer J., Müller F., "High Resolution Imaging in Microscopy and Ophthalmology," Springer International Publishing, 2019.

[12] A. F. Fercher, C. K. Hitzenberger, G. Kamp, and S. Y. El-Zaiat, “Measurement of intraocular distances by backscattering spectral interferometry,” *Optics Communications,* vol. 117, no. 1, pp. 43-48, 1995/05/15/, 1995.

[13] N. Eladawi, M. Elmogy, M. Ghazal, A. H. Mahmoud, H. Mahmoud, M. T. Alhalabi, A. Aboelfetouh, A. Riad, R. Keynton, S. Schaal, and A. El-Baz, "7 - Optical coherence tomography: A review," *Diabetes and Fundus OCT*, A. S. El-Baz and J. S. Suri, eds., pp. 191-2 :21Elsevier, 2020.

[14] M. Gabriele, G. Wollstein, H. Ishikawa, L. Kagemann, J. Xu, L. Folio, and J. Schuman, "Optical Coherence Tomography: History, Current Status, and Laboratory Work," *Investigative ophthalmology & visual science*, vol. 52, pp. 2425-36.2011 ,01/04 ,

[15] K. Alsaih, G. Lemaître, M. Rastgoo, J. Massich, D. Sidibé, and F. Meriaudeau, "Machine learning techniques for diabetic macular edema (DME) classification on SD-OCT images," *BioMedical Engineering OnLine*, vol. 16, 06/07, 2017.

[16] "Know your retinal layers," Accessed: 29 November 2020; https://business-lounge.heidelbergengineering.com/ca/en/news/news/know-your-retinal-layers-33401465.

[17] S. Roychowdhury, *3-D localization of Diabetic Macular Edema using OCT thickness maps*, 2015.

[18] T. Alasil ,K. Wang, F. Yu, M. G. Field, H. Lee, N. Baniasadi, J. F. de Boer, A. L. Coleman, and T. C. Chen, "Correlation of retinal nerve fiber layer thickness and visual fields in glaucoma: a broken stick model," *American journal of ophthalmology*, vol. 157, no. 5 ,pp. 953-959, 2014.

[19] G. Soubrane, and G. Coscas, "Choroidal Neovascular Membrane in Degenerative Myopia," *Retina*, vol. 2, pp. 1136-1152, 12/31, 2006.

[20] J. Cafasso. "What You Need to Know About Drusen," Accessed: 15 Febrauary 2020; https://www.healthline.com/health/drusen.

[21] V. E. "Macular Edema - What Is It, What Causes it and How Does it Affect Vision?," Accessed: 12 February 2020; https://www.enhancedvision.com/low-vision-info/macular-edema-what-is-it-what-causes-it-and-how-does-it-affect-vision.html.

[22] M. Clinic. "Diabetic macular edema," Accessed: 12 February 2020; www.mayoclinic.org/diseases-conditions/diabetic-retinopathy/multimedia/diabetic-macular-edema/img-20124558.

[23] S. Sanjay, T. K. Yeo, and K.-G. Au Eong, "Spontaneous closure of traumatic macular hole," *Saudi journal of ophthalmology : official journal of the Saudi Ophthalmological Society*, vol. 26, no. 3, pp. 343-345, 2012.

[24] A. M. Bagci, A. Rashid, and S. Mahnaz, "A method for detection of retinal layers by optical coherence tomography image segmentation." pp. 144-147.

[25] S. Lu, C. Y. Cheung, J. Liu, J. H. Lim, C. K. Leung, and T. Y. Wong, “Automated Layer Segmentation of Optical Coherence Tomography Images,” *IEEE Transactions on Biomedical Engineering,* vol. 57, no. 10, pp. 2605-2608, 2010.

[26] M. Garvin, M. Abramoff, R .Kardon, S. Russell, X. Wu, and M. Sonka, “Intraretinal Layer Segmentation of Macular Optical Coherence Tomography Images Using Optimal 3-D Graph Search,” *Medical Imaging, IEEE Transactions on*, vol. 27, pp. 1495-1505, 11/01, 2008.

[27] I. Ghorbel, F. Rossant, I. Bloch, S. Tick, and M. Paques, “Automated segmentation of macular layers in OCT images and quantitative evaluation of performances,” *Pattern Recognition,* vol. 44, no. 8, pp. 1590-1603, 2011/08/01/, 2011.

[28] F. Shi, X. Chen, H. Zhao, W. Zhu, D .Xiang, E. Gao, M. Sonka, and H. Chen, “Automated 3-D Retinal Layer Segmentation of Macular Optical Coherence Tomography Images With Serous Pigment Epithelial Detachments,” *IEEE transactions on medical imaging*, vol. 34, 09/24, 2014.

[29] J. Sugmk, S. Kiattisin, and A. Leelasantitham, “Automated classification between age-related macular degeneration and Diabetic macular edema in OCT image using image segmentation,” *BMEiCON 2014 - 7th Biomedical Engineering International Conference*, 01/20, 2015.

[30] M. Salarian, R. Ansari, J. Wanek, and M. Shahidi, "Acuurate segmentation of retina nerve fiber layer in OCT images." pp. 653-656.

[31] S. Luo, J. Yang, Q. Gao, S. Zhou, and C. a. Zhan, “The Edge Detectors Suitable for Retinal OCT Image Segmentation,” *Journal of Healthcare Engineering,* vol. 2017, pp. 1-13, 08/17, 2017.

[32] S. Niu, Q. Chen, L. de Sisternes, D. L. Rubin, W. Zhang, and Q. Liu, “Automated retinal layers segmentation in SD-OCT images using dual-gradient and spatial correlation smoothness constraint ”,*Computers in Biology and Medicine*, vol. 54, pp. 116-128, 2014/11/01/, 2014.

[33] A. ElTanboly, M. Ismail, A. Shalaby, A. Switala, A. El-Baz, S. Schaal, G. Gimel’farb, and M. El-Azab, “A computer-aided diagnostic system for detecting diabetic retinopathy in optical coherence tomography images,” *Medical Physics*, vol. 44, no. 3, pp. 914-923, 2017/03/01, 2017.

[34] P. Dash, and A. N. Sigappi, “Denoising and segmentation of retinal layers in optical coherence tomography images,” *AIP Conference Proceedings,* vol. 1952, no. 1, pp. 020072, 2018/04/24, 2018.

[35] B. I. Dodo, Y. Li, K. Eltayef, and X. Liu, “Automatic Annotation of Retinal Layers in Optical Coherence Tomography Images,” *Journal of Medical Systems,* vol. 43, no. 12, pp. 336, 2019/11/13, 2019.

[36] P .P. Srinivasan, L. A. Kim, P. S. Mettu, S. W. Cousins, G. M. Comer, J. A. Izatt, and S. Farsiu, "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images," *Biomedical optics express*, 5, http://europepmc.org/abstract/MED/25360373
https://doi.org/10.1364/BOE.5.003568

[37] F. Li, H. Chen, Z. Liu, X. Zhang, and Z. Wu, "Fully automated detection of retinal disorders by image-based deep learning," *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 257, no. 3, pp. 495-505, 2019/03/01, 2019.

[38] Z. Tianyu, M. Zhenjiang, and Z. Jianhu, *Combining CNN with Hand-Crafted Features for Image Classification*, 2018.

[39] M. Awais, H. Müller, T. B. Tang, and F. Meriaudeau, "Classification of SD-OCT images using a Deep learning approach." pp. 489-492.

[40] O. Perdomo, S. Otálora, F. A. González, F. Meriaudeau, and H. Müller, "OCT-NET: A convolutional network for automatic classification of normal and diabetic macular edema using sd-oct volumes." pp. 1423-1426.

[41] N. Motozawa, G. An, S. Takagi, S. Kitahata, M. Mandai, Y. Hirami, H. Yokota, M. Akiba, A. Tsujikawa, M. Takahashi, and Y. Kurimoto, "Optical Coherence Tomography-Based Deep-Learning Models for Classifying Normal and Age-Related Macular Degeneration and Exudative and Non-Exudative Age-Related Macular Degeneration Changes," *Ophthalmology and Therapy*, vol. 8, no ,4 .pp. 527-539, 2019/12/01, 2019.

[42] D. S. Kermany, M. Goldbaum, W. Cai, C. C. S. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, J. Dong, M. K. Prasadha, J. Pei, M. Y. L. Ting, J. Zhu, C. Li, S. Hewett, J. Dong, I. Ziyar, A. Shi ,R. Zhang, L. Zheng, R. Hou, W. Shi, X. Fu, Y. Duan, V. A. N. Huu, C. Wen, E. D. Zhang, C. L. Zhang, O. Li, X. Wang, M. A. Singer, X. Sun, J. Xu, A. Tafreshi, M. A. Lewis, H. Xia, and K. Zhang, "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning," *Cell*, vol. 172, no. 5, pp. 1122-1131.e9, 2018/02/22/, 2018.

[43] E. Abdellatef, E. M. Omran, R. F. Soliman, N. A. Ismail, S. E. S. E. Abd Elrahman, K. N. Ismail, M. Rihan, F. E. Abd El-Samie, and A. A. Eisa, "Fusion of deep-learned and hand-crafted features for cancelable recognition systems," *Soft Computing*, 2020/04/13, 2020.

[44] F. Wang, C. Zhang, S. Chen, G. Ying, and J. Lv, "Engineering Hand-designed and Deeply-learned features for person Re-identification," *Pattern Recognition Letters*, vol. 130, pp. 293-298, 2020/02/01/, 2020.

[45] T. Georgiou, Y. Liu, W. Chen, and M. Lew, "A survey of traditional and deep learning-based feature descriptors for high dimensional data in computer vision," *International Journal of Multimedia Information Retrieval*, 2019/11/22, 2019.

[46] H. Zhang, H. Han, J. Cui, S. Shan, and X. Chen, "RGB-D Face Recognition via Deep Complementary and Common Feature Learning." pp. 8-15.

[47] Y. Xie, Y. Xia, J. Zhang, D. D. F. Feng, M. Fulham, and W. Cai, "Transferable Multi-model Ensemble for Benign-Malignant Lung Nodule Classification on Chest CT," *Lecture Notes in Computer Science*, vol. 10435, pp. 656-664, 09/04, 2017.

[48] Y.-J. Lee, Y.-R. Yeh, and H.-K. Pao, "Introduction to Support Vector Machines and Their Applications in Bankruptcy Prognosis," *Handbook of Computational Finance*, 07/01, 2012.

[49] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: a review of classification and combining techniques," *Artificial Intelligence Review*, vol. 26, no. 3, pp. 159-190, 2006/11/01, 2006.

[50] S. Dargan, M. Kumar, M. R. Ayyagari, and G. Kumar, "A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning," *Archives of Computational Methods in Engineering*, vol. 27, no. 4, pp.2020 ,01/09/2020 ,1092-1071 .

[51] D. Li, and Y. Dong, *Deep Learning: Methods and Applications*: now, 2014.

[52] O. Faust, Y. Hagiwara, J. H. Tan, S. L. Oh, and U. R. Acharya, "Deep learning for healthcare applications based on physiological signals: A review," *Computer Methods and Programs in Biomedicine*, vol. 161, 04/01, 2018.

[53] N. Lopes, and B. Ribeiro, *Machine Learning for Adaptive Many-Core Machines - A Practical Approach*, 2015.

[54] J. Heaton, *Artificial Intelligence for Humans, Volume 3: Deep Learning and Neural Networks*: Createspace Independent Publishing Platform, 2015.

[55] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[56] K. Simonyan, and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv 1409.1556*, 09/04, 2014.

[57] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. J. I. C. o. C. V. Wojna, and P. Recognition, "Rethinking the Inception Architecture for Computer Vision," pp. 2818-2826, 2016.

[58] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition." pp. 770-778.

[59] E. Rezende, G. Ruppert, A. Theophilo, and T. Carvalho, “Exposing Computer Generated Images by Using Deep Convolutional Neural Networks,” *Signal Processing: Image Communication*, 11/28, 2017.

[60] A. A. Ardakani, A. R. Kanafi, U. R. Acharya, N. Khadem, and A. Mohammadi, “Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks,” *Computers in Biology and Medicine*, vol. 121, pp. 103795, 2020/06/01/, 2020.

[61] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. J. I. C. C. o. C. V. Chen, and P. Recognition, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” pp. 4510-4520, 2018.

[62] A. Buades, B. Coll, and J. Morel, "A non-local algorithm for image denoising." pp. 60-65 vol. 2.

[63] M. Lebrun, “An Analysis and Implementation of the BM3D Image Denoising Method,” *Image Processing On Line*, vol. 2, pp. 175-213, 08/08, 2012.

[64] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336-359, 2020/02/01, 2020.

[65] E. R. Davies, *Computer and Machine Vision, Fourth Edition: Theory, Algorithms, Practicalities*, 4th ed.: Academic Press Inc., 2012.

[66] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*: Thomson-Engineering, 2007.

[67] M. Srinivas, D. Roy, and C. K. Mohan, "Discriminative feature extraction from X-ray images using deep convolutional neural networks." pp. 917-921.

[68] S. Hijazi, R. Kumar, and C. Rowen, "Using Convolutional Neural Networks for Image Recognition By".

[69] Z. Mohsin, and L. Alzubaidi, *Convolutional Neural Network with Global Average Pooling for Image Classification*, 2020.

[70] M. Lin, Q. Chen, and S. Yan, “Network In Network,” 12/16, 2013.

[71] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” vol. 15, no. 1 %J J. Mach. Learn. Res., pp. 1929–1958, 2014.

[72] S. Ioffe, and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” 2015, pp. 448-456.

[73] J. Patterson, and A. Gibson, *Deep Learning: A Practitioner's Approach*: O'Reilly Media, Inc., 2017.

[74] Chris. "What are L1, L2 and Elastic Net Regularization in neural networks?," Accessed: 4 August 2020; https://www.machinecurve.com/index.php/2020/01/21/what-are-l1-l2-and-elastic-net-regularization-in-neural-networks./

[75] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*: The MIT Press, 2016.

[76] D. A. Lisin, M. A. Mattar, M. B. Blaschko, E. G. Learned-Miller, and M. C. Benfield, "Combining Local and Global Image Features for Object Class Recognition." pp. 47-47.

[77] E. Tola, V. Lepetit, and P. Fua, "A fast local descriptor for dense matching ",*Proc. CVPR*, 06/01, 2008.

[78] L. Nanni, S. Ghidoni, and S. Brahnam, "Handcrafted vs. non-handcrafted features for computer vision classification," *Pattern Recognition*, vol. 71, pp. 158-172, 2017/11/01/, 2017.

[79] A. Ross, "Fusion, Feature-Level," *Encyclopedia of Biometrics*, S. Z. Li and A. Jain, eds., pp. 597-602, Boston, MA: Springer US, 2009.

[80] J. Zupan, "Introduction to Artificial Neural Network (ANN) Methods: What They Are and How to Use Them," *Acta Chimica Slovenica*, vol. 41, 01/01, 1994.

[81] M .Hasan, M. M. Islam, M. I. I. Zarif, and M. M. A. Hashem, "Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches," *Internet of Things*, vol. 7, pp. 100059, 2019/09/01/, 2019.

[82] K. Simonyan, and A. J. C. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," vol. abs/1409.1556, 2015.

Appendix

APPENDIX: VALIDATION REPORT FROM MEDICAL EXPERT

Dr. Eman Elbanna

MSC of Ophthalmology
FRCS of Ophthalmology
Glasgow University
Consultant of Ophthalmology

دكتورة
إيمان سعيد البنا

زميل كلية الجراحين الملكية بإنجلترا

إستشارى طب و جراحة العيون

Validation Report

Master thesis titled: A Hybrid Computer Aided Medical Diagnosis System Integrating Machine Learning and Automatic Region of Interest Segmentation: A Case-Study on Diagnosis of Optical Coherence Tomography Retinal Disorders, By Mohamed Ramzy Ibrahim for fulfilling the requirements of master degree of science in computer engineering.

The thesis consists of 86 pages and 7 pages of references.

It is divided into:

- Abstract
- List of Publication
- Chapter 1: Introduction
- Chapter 2: Medical Background
- Chapter 3: Literature Survey
- Chapter 4: HyCAD: Hybrid Computer Aided Diagnosis Proposed Framework
- Chapter 5: OCT Retinal Disorders Experimentation and Results
- Chapter 6: Conclusion and Future Work

The thesis presents a new contribution in computer aided diagnosis systems that aims to localize the abnormalities in retinal layers of OCT images to detect the early stages of different retinal disorders. The proposed model idea can be used to assist the ophthalmologists to confirm the early stages diagnosis and prevent further complications.

The thesis clearly explained the medical point of view of different retinal disorders and focuses on the critical cases of retinal disorders. The proposed Hybrid Computer Aided Diagnosis Framework (HyCAD) provides reasonable results and future work was also explained.

The thesis is accepted regarding the contents and framework and valid for international publication.

د . إيمان سعيد البنا
إستشارى طب وجراحة العيون
FRCS.GLasgow

يتم إجراء تجارب مختلفة للتحقق لاثبات صحة الأداء الفعال للنظام المقترح على مجموعة البيانات الكبيرة لمجموعة البيانات المرجعية للتصوير التصوير للترابط البصري المسمى OCT) v3). مقارنة بالأداء للنماذج الاخري الموجودة فى الابحاِث. أظهرت النتائج التجريبية أن النموذج المقترح يحقق أداءً عاليًا نسبيًا من حيث الدقة والحساسية والنوعية. تم تحقيق متوسط دقة وحساسية ونوعية 98.8٪ و 99.4٪ و 98.2٪ على التوالي. أظهر التقييم التجريبي أداءً متميزًا مقارنة بالأداء للنماذج الاخري الموجودة فى الابحاِث ، فإن نموذج (HyCAD) المقترح يحقق حساسية تصنيف عالية للحالات العاجلة. نتائج نماذج HyCAD مقارنة بنتائج العمود الفقري لـ CNN (Norm-VGG16) وغيرها من بنى التعلم العميق الشهيرة. لقد حققت زيادة كبيرة في الدقة والحساسية والنوعية.

في هذه الأطروحة ، تم تقديم HyCAD بنية عامة لنظام التشخيص بمساعدة الكمبيوتر. يمكن تدريب الهيكل المقترح وتطبيقه على مشاكل مماثلة ، حيث لم يتم وضع افتراضات أساسية محددة من شأنها أن تعيق تعميمها. تعكس النتائج التجريبية أن مرحلة الاندماج (automatic CNN features and hand-crafted features) يمكن أن تحسن بشكل فعال نسبة تحديد الصور التشخيصية للمرضى العاجلين. بالإضافة إلى ذلك ، يتم تحقيق أداء متميز مقارنة بالأداء للنماذج الاخري الموجودة فى الابحاِث.

الفصل الأول :مقدمة وتشمل نظرة عامة للموضوع وتطبيقاته، مع عرض لأهداف الرسالة والطرق المستخدمة وكذلك الهيكلة العامة للرسالة.

الفصل الثاني :يعرض فيه الباحث تغطية للموضوع و انواع امراض الشبكيه المختلفة.

الفصل الثالث: تغطية للملامح المميزة أنظمة CAD لاستئصال طبقات الشبكية و تشخيص امراض الشبكية و طرق التصنيف.

الفصل الرابع :يعرض النموذج المقترح مع شرح للمراحل المختلفة.

الفصل الخامس :يعرض نتائج تطبيق النظام المقترح على مجموعة كبيرة من انواع امراض الشبكيه المختلفة ، لإثبات حساسية ودقة النموذج المقترح، مع مقارنة بالنماذج الأخرى.

الفصل السادس : الاستنتاجات والتوقعات المستقبلية.

ملخص الرسالة

فحص التصوير للترابط البصري (OCT) هو تقنية تصوير غير جراحية تُستخدم بشكل متزايد لتشخيص وإدارة مجموعة متنوعة من أمراض الشبكية والزرق. يتمتع تصوير OCT بمزايا رئيسية في تحديد وجود العديد من أمراض العين بشكل فعال واكتشاف مجموعة واسعة من الأمراض البقعية. يمكن أن تساعد فحوصات OCT في الكشف عن العديد من اضطرابات الشبكية في المراحل المبكرة والتي لا يمكن اكتشافها في صور الشبكية التقليدية. تم استخدام OCT لتحديد الثقوب البقعية ، والخراجات البقعية ، والانفصال الظهاري الصبغي ، وتكوين الأوعية الدموية المشيمية. يمكن استخدامه لتحديد وقياس الوذمة البقعية ، وقياس تغيرات سمك الشبكية استجابة للعلاج.

غالبًا ما يتم تشخيص وعلاج اضطرابات الشبكية من قبل طبيب عيون. ومع ذلك ، لتقييم مرض الشبكية بدقة ، سيحتاج طبيب العيون إلى تحليل نوعي وكمي للمرض. علاوة على ذلك ، تؤدي الزيادة غير المتناسبة في عبء العمل بالنسبة للقوى العاملة إلى التناقض والخطأ في تشخيص الأمراض والاضطرابات المختلفة. أيضًا ، هناك ضرورة للكشف المبكر عن اضطرابات الشبكية لتحسين التشخيص وتجنب حدوث المضاعفات التي قد تؤدي إلى فقدان البصر. للأسباب السابقة ، يجب أن يكون هناك نظام تشخيص بمساعدة الكمبيوتر (CAD) لمساعدة أطباء العيون في التشخيص المبكر الدقيق لاضطرابات الشبكية المختلفة.

في هذه الأطروحة ، تم اقتراح نظام تشخيص هجين جديد بمساعدة الكمبيوتر (HyCAD) لتصنيف اضطرابات الشبكية: الوذمة البقعية السكري (DME) ، واضطراب الأوعية الدموية المشيمية (CNV) واضطرابات البراريق ، مع فصلها عن بنية الشبكية الطبيعية باستخدام صور OCT . يستوعب نظام التعلم الهجين HyCAD المقترح مجموعة من التقنيات بما في ذلك توطين ROI على أساس التعلم العميق ، واستخراج الميزات الهجينة ، متبوعًا بالتصنيف والتشخيص. تم تقديم مرحلة دمج ميزة فعالة للجمع بين ميزات صورة OCT ، المستخرجة بواسطة الشبكة العصبية التلافيفية العميقة (CNN) ، مع الميزات المستخرجة من مرحلة تجزئة RoI. تُستخدم مجموعة الميزات المدمجة هذه للتنبؤ باضطرابات شبكية العين OCT متعددة الطبقات. تضيف مرحلة التجزئة المقترحة لمناطق العائد على الاستثمار في شبكية العين مساهمة كبيرة لأنها تلفت الانتباه إلى أهم المناطق المرشحة للتشخيص. تم تقديم بنية التعلم العميق المعدلة الجديدة (Norm-VGG16) بدمج منظم النواة (kernel regularizer) . يتم تدريب Norm-VGG16 من نقطة الصفر على مجموعة بيانات معيارية كبيرة وتستخدم في توطين وتجزئة المناطق المرشحة للتشخيص (RoI).

إقرار لجنة التحكيم والمناقشة

محمد رمزى ابراهيم

تم مناقشة هذه الرسالة و اجازتها بتاريخ: 13 / 12 / 2020

أعضـاء اللجنـة

الأساتذة المشرفون على الرسالة:

الإسـم: **ا.د. شيرين مصطفى يوسف** **التوقيـع:....................................**
الوظيفة: أستاذ هندسة الحاسب – كلية الهندسة و التكنولوجيا – الأكاديمية العربية للعلوم و التكنولوجيا و النقل البحرى – الأسكندرية

الإسـم: **د. كارمة محمد فتح الله** **التوقيـع:....................................**
الوظيفة: مدرس هندسة الحاسب – كلية الهندسة و التكنولوجيا – الأكاديمية العربية للعلوم و التكنولوجيا و النقل البحرى - الأسكندرية

الأساتذة المحكمون:

الإسـم: **ا.د. صالح عبدالشكور الشهابى** **التوقيـع:....................................**
الوظيفة: أستاذ هندسة الحاسب – معهد الدراسات العليا والبحوث – جامعة الأسكندرية

الإسـم: **ا.د. علياء عبدالحليم عبدالرازق يوسف** **التوقيـع:....................................**
الوظيفة: أستاذ علوم الحاسب و تكنولوجيا المعلومات – كلية الحاسبات و تكنولوجيا المعلومات – الأكاديمية العربية للعلوم و التكنولوجيا و النقل البحرى – القرية الذكية

الإسـم: **ا.د. شيرين مصطفى يوسف** **التوقيـع:....................................**
الوظيفة: أستاذ هندسة الحاسب – كلية الهندسة و التكنولوجيا – الأكاديمية العربية للعلوم و التكنولوجيا و النقل البحرى - الأسكندرية

أودعت هذه الرسالة بالمكتبة بتاريخ / /

إقرار الباحث

أقر بأن المادة العلمية الواردة في هذه الرسالة قد تم تحديد مصدرها العلمي وأن محتوى الرسالة غير مقدم للحصول علي أي درجة علمية أخرى، وأن مضمون هذه الرسالة يعكس أراء الباحث الخاصة وهى ليست بالضرورة الآراء التي تتبناها الجهـة المانحة.

الاسم: محمد رمزى ابراهيم

التوقيع: ....

التاريخ: ....

الأكاديمية العربية للعلوم والتكنولوجيا والنقل البحري
كليـــــة الهندســـــة والتكنولوجيـــــــــــا
قسم هندسة الحاسب

نظام هجين للتشخيص الطبى بدمج التعلم الآّلى وفصل مناطق الاهتمام تلقائيا: دراسة في مجال فحص التصوير المقطعي للترابط البصري في الكشف عن اضطرابات الشبكية

إعداد
محمد رمزي ابراهيم
بكالوريوس هندسة الحاسب 2018
كلية الهندسة والتكنولوجيا – الأكاديمية العربية للعلوم والتكنولوجيا والنقل البحري

رسالة مقدمة للأكاديمية العربية للعلوم والتكنولوجيا والنقل البحرى لاستكمال متطلبات نيل درجة
ماجستير العلوم
فى
هندسة الحاسب الآلي

إشـــــــــراف

2020

أ.د. شيرين مصطفى يوسف
قسم هندسة الحاسب
كلية الهندسة والتكنولوجيا

د. كارمة محمد فتح الله
قسم هندسة الحاسب
كلية الهندسة والتكنولوجيا

202
الأكاديمية العربية للعلوم والتكنولوجيا و النقل البحري
الأسكندرية