Gabor Wavelet Based Diabetic Retinopathy Detection Using Deep Learning

Ghulam Mustuffa Khan

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Electrical and Electronic Engineering

Eastern Mediterranean University
August 2021
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Ali Hakan Ulusoy
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science in Electrical and Electronic Engineering.

Assoc. Prof. Dr. Rasime Uyguroğlu
Chair, Department of Electrical and Electronic Engineering

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Master of Science in Electrical and Electronic Engineering.

Prof. Dr. Hasan Demirel
Supervisor

Examining Committee

1. Prof. Dr. Hasan Demirel

2. Prof. Dr. Önsen Toygar

3. Asst. Prof. Dr. Kamil Yurtkan

ABSTRACT

Diabetic Retinopathy (DR) is the disease that causes blindness when it reaches to the proliferative stage. There are four stages in DR namely: NO DR, Mild DR, Severe DR and Proliferative DR. The detection of diabetic retinopathy in the early stages can prevent sight loss in a notable amount of the population worldwide. The earliest signs of diabetic retinopathy are hemorrhages and hard exudates which are red and yellow lesions on the retina of the eye. Diagnosis of the DR is performed using retinal image analysis. Manual Analysis of these images to decide the presence of the DR is somewhat slow and costly. To ease the job of a medical consultant, we can take the images of the retina and feed them to the trained machine learning model and get the results whether the person has diabetic retinopathy or not.

In this thesis, we propose a system to detect and classify diabetic retinopathy using the Gabor filter-based CNN with fusion. Proposed model fuses the decisions generated by Gabor-based CNN and traditional CNN pipelines. A preprocessing stage is employed to normalize images before feeding them into the CNN architecture. Preprocessing includes cropping and image resizing. A pre-trained CNN model VGG16, is selected as a deep machine learning architecture throughout the thesis. Decision fusion approaches including Sum Rule and Product Rule are employed throughout the fusion process. The recall, specificity, f1score, precision, and accuracy of the models have been studied and generated results are compared with each other and methods in the literature. The receiver characteristics curve (ROC) and Area under the curve (AUC) metrics are also measured to evaluate the classification performance. Messidor retina image database have been used to measure the performance of the

proposed system. The results show that the proposed system to detect and classify diabetic retinopathy using the Gabor filter-based CNN with fusion generates higher performance over the state-of-the-art alternative methods in the literature.

Keywords: Convolutional Neural Networks, Deep Learning, Machine Learning, Gabor Filters, Data Fusion.

ÖZ

Diyabetik Retinopati (DR), proliferatif aşamaya ulaştığında körlüğe neden olan bir hastalıktır. Diyabetik retinopatinin erken evrelerde saptanması, dünya nüfusunun önemli bir kısmında görme kaybını önleyebilir. Diyabetik retinopatinin en erken belirtileri, gözün retinasında kırmızı ve sarı lezyonlar olan kanamalar ve sert salgılardır. DR'nin teşhisi, retina görüntü analizi kullanılarak gerçekleştirilebilmektedir. DR'nin varlığına karar vermek için bu görüntülerin Manuel Analizi biraz yavaş ve maliyetlidir. Bir tıbbi danışmanın işini kolaylaştırmak için retinanın görüntülerini alıp eğitimli makine öğrenme modeline besleyebilir ve kişide diyabetik retinopati olup olmadığına bakabiliriz.

Bu tezde, füzyonlu Gabor filtre tabanlı CNN kullanarak diyabetik retinopatiyi tespit etmek ve sınıflandırmak için bir sistem öneriyoruz. Önerilen model, Gabor tabanlı CNN ve geleneksel CNN kanalları tarafından üretilen kararları birleştirir. Görüntüleri CNN mimarisine beslemeden önce normalleştirmek için bir ön işleme aşaması kullanılmaktadır. Ön işleme süreci, kırpma ve görüntünün yeniden boyutlandırılmasını içerir. Tez kapsamında önceden eğitilmiş bir CNN modeli olan VGG16 derin makine öğrenimi mimarisi olarak seçilmiştir. Birleştirme süreci boyunca toplam kuralı ve çarpım kuralı içeren karar birleştirme yaklaşımları kullanılmaktadır. Modellerin hatırlama, özgüllük, f1skor, kesinlik ve doğrulukları çalışılmış ve elde edilen sonuçlar literatürdeki yöntemler ve birbirleriyle karşılaştırılmıştır. Alıcı özellikleri eğrisi (ROC) ve Eğrinin altındaki alan (AUC) metrikleri de sınıflandırma performansını değerlendirmek için ölçülür. Önerilen sistemin performansını ölçmek için Messidor retina görüntü veri tabanı kullanılmıştır. Sonuçlar, füzyonlu Gabor filtresi tabanlı CNN

kullanılarak diyabetik retinopatiyi tespit etmek ve sınıflandırmak için önerilen sistemin, literatürdeki son teknoloji alternatif yöntemlere göre daha yüksek performans ürettiğini göstermektedir.

Anahtar Kelimeler: Evrişimli Sinir Ağları, Derin Öğrenme, Makine Öğrenimi, Gabor Filtreleri, Veri Füzyonu.

DEDICATION

To my family and Loved ones

ACKNOWLEDGEMENTS

It gives me great pleasure to praise and thank Allah SWT for His greatness and that He gave me this opportunity, strength, and courage to complete my master's thesis.

I would like to thank my supervisor, Prof. Dr. Hasan Demirel, for guiding me and mentoring me throughout my master study, for being patient and sharing kindly his knowledge with me.

My sincere thanks go out to the jury members who agreed to review my work and attend my thesis defense.

Also, I would like to thank all the staff of the Electrical and Electronic department, Eastern Mediterranean University.

I also wish to thank my friends Isa Ahmed and Shimraaz Rabbani for always supporting me in any way possible.

In particular, I wish to thank my parents for supporting me throughout my life endeavors. As a result of their efforts, I strive to be the best I can be, and I will always be grateful for them.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVATIONS

1-D	One Dimensional
2-D	Two Dimensional
AI	Artificial Network
ANN	Artificial Neural network
ASOLO	Adaptive Optics Scanning Laser Ophthalmoscopy
CCD	Charged Coupled Device
CNN	Convolutional Neural Network
CONV	Convolutional
DF	Decision Fusion
DL	Deep Learning
DM	Diabetes Mellitus
DM2	Type II Diabetes Mellitus
DNN	Deep Neural network
EX	Exudates
FC	Fully Connected
FFA	Fundus Fluorescein Angiography
FN	False Negative
FOV	Field Of View
FP	False Positive
FPR	False Positive Ratio
HE	Hard Exudates
HM	Haemorrhages

MA	Microaneurysms
ML	Machine Learning
MSL	Mean square Loss
OCT	Optical Coherence Tomography
POOL	Pooling Layer
ReLU	Rectified Linear Unit
RGB	Red Green Blue
ROC	Receiver Operating Characteristics
SE	Soft Exudates
SE	Sensitivity
SP	Specificity
TN	True Negative
TP	True Positive
TPR	True Positive Ratio
WHO	World Health Organization

Chapter 1

INTRODUCTION

1.1 Introduction

Diabetes mellitus (DM) as a chronic condition is characterized by hyperglycemia [1]. Type 1, Type II, and gestational diabetes are the three major types of diabetes recognized by the World Health Organization (WHO). All these three types of diabetes have comparable signs, symptoms, and consequences, unlike causes and spread.

DM currently represents a serious public health issue due to the high prevalence in the world, especially in developing countries, due to fatality, and because it is one of the primary cardiovascular risk factors [2]. In the natural course of diabetes, chronic complications emerge, which are categorized as microvascular: retinopathy, neuropathy, nephropathy, cardiomyopathy, and Macrovascular: coronary artery disease, cerebrovascular disease, and peripheral arterial disease. The following complications are associated with: significant morbidity and mortality, including cardiovascular and renal mortality, blindness, limb amputation, as well as significant loss of function and quality of life compared to individuals without diabetes. In the aging population and an increasingly inactive lifestyle, there is an increase in patients with type II diabetes mellitus (DM2) due to their lifestyle and eating habits, both of which contribute to body fat accumulation.

Diagnostic and therapeutic procedures, hospitalizations, absenteeism, disability, and premature death are only a few of the complications that occur as a result of diabetes that contribute to substantial direct and indirect health care expenses. The disease and chronic complications that accompany it must both be prevented.

Diabetic Retinopathy (DR) is a complication of diabetes characterized by swelling of the retinal blood vessels and leaking of fluid. There are 2.6% of blind people in the world due to DR [3]. DR may present without symptoms in the early stages, and when the disease becomes more advanced, treating it may be more challenging. It is therefore imperative that DR be detected as soon as possible [4]. It is therefore mandatory for diabetic patients to undergo systematic eye examinations involving a checkup of their retina by an ophthalmologist. To detect impaired eyes, a variety of imaging techniques can be used, including slit lamp biomicroscopy [5], optical coherence tomography (OCT), fundus fluorescein angiography (FFA) [6], and fundus images.

Performing a manual diagnosis requires a lot of effort. The use of automatic systems has the potential to reduce time and effort, as well as costs. The increase in diabetes cases has prompted an increase in automatic analysis tools in recent years. Moreover, image processing, analysis, computer vision, and increasing computer speed are becoming increasingly important in the field of medical science and ophthalmology. Our motivation to use Gabor filter comes from the fact that the Gabor functional scheme is seen as a good representation for visual information in the combined spatial-frequency domain; simple cells in the visual cortex can also be modeled by Gabor function [7].

1.2 Thesis Objectives

The objective of this work is to offer an automatic detection system for Diabetic Retinopathy. We propose to use the Gabor filter-based CNN models and Image-based CNN model in this thesis in order to detect diabetic retinopathy in retinal images. In the first step, seven Gabor filters are used to represent retinal images. After that, the seven Gabor-based CNN models are learned to recognize diabetic retinopathy. Image-based CNN model is also generated using only retinal images. In the end we combine the CNN model outputs to estimate the diabetic retinopathy. Specifically, the following objectives are being pursued:

- Detecting diabetic retinopathy using Gabor filters. Investigate multiple directions and scales for robust detection.
- Decision Fusion using sum rule and product rule.
- The analysis of Receiver Operating Characteristics (ROC) for high performance measurements such as True Positive Ratio (TPR), False Positive Ratio (FPR), sensitivity, specificity, accuracy, precision, and f1score.
- Compare the results of the generated method with those obtained through alternative methods in the literature.

1.3 Thesis Contributions

In this thesis, we put forward the use of ensemble of CNN based models for the classification of DR. The first step is to apply bank of Gabor filters to the original images with certain degree of orientation to each other and collect the output. Then we feed these images to the CNN model and take output probabilities. The Original Image is also fed to the separate model. The proposed method fuses the output probabilities of these ensemble of CNN models and gives the classification for DR. The

probabilistic techniques used in this thesis are add rule and product rule. This thesis also proves that performance of ensemble of models is better than single model for DR classification. The results generated by this study are comparable with other state-of-the-art results in the relevant literature, using retina images found in a standard image database. In this study, we used the publicly available dataset Messidor, which contains 1200 color fundus images.

1.4 Thesis Organization

Chapter 2 begins with a description of the eye's structure and DR as it relates to the eye. Different stages of DR are also discussed. The discussion is briefly followed by a review of some previous work done on detecting and classifying DR. In Chapter 3, we talk about Deep Learning, Deep Neural Networks and how they can be used. Also, Convolutional Neural Networks are described and some pre-trained architectures are explained. In Chapter 4, we described Gabor filters and its application with respect to image processing. Chapter 5 gives insights about metrics evaluation methods and database used in this study. Proposed method is described in Chapter 6. Also, results and relevant graphs are shown as well. Conclusion and future work are briefly explained in Chapter 7.

Chapter 2

RETINAL IMAGE PROCESSING FOR DIABETIC RETINOPATHY

2.1 Introduction

A fundus image represents a two-dimensional representation of the three-dimensional structure of the retina [8]. It is the amount of reflected light that determines the image intensity. There is a microscope and camera attached to the eye fund camera, so you can look at it through the microscope. As with indirect ophthalmoscopes, the optic system can magnify internal observations of the eye. At an amplification of 2.5 times, the camera observes the retina at a 30- to 50-degree angle. An auxiliary lens increases this amplification fivefold. A variety of dyes are used to perform the test, including colored filters, fluorescein, and other types. To detect DR, the retina is examined using the following methods;

- Fundus Photography (red-free) - Image-capturing is done by measuring how much light is reflected by a specific wavelength band [9].
- Color Fundus Photography - Using reflected light, the image is captured as RGB (Red, Green, and Blue) [10].
- Indocyanine and Fluorescein Angiography - Images are formed by counting the number of photons emitted by dyes injected into the bloodstream, such as fluorescein or indocyanine [11]. As a result of a few advantages over the other methods, we choose only Color Fundus

Photography for this research. RGC fundus cameras only use the green channel for their photography, which can compromise useful information. Fluorescein angiography, however, requires injections of contrast material, which makes it more expensive and invasive.

2.2 Related Work

To classify retinal fundus images of DR type, several algorithms have been proposed. For instance, using a CNN, study [17] classified the images of Kaggle database as no diabetic or diabetic. A total of thousand images were used from the database. Prior to inputting the color fundus images to the CNN, the images were augmented and resized to 224*224*3. CNN consists of 8 CONV layers, 4 maximum-pooling layer, and 2 FC layer. For classification, the SoftMax was used at the last layer of CNN. An accuracy of 94.5% was obtained using this method.

In a study by M. T. Esfahan et al. [18], which classified Diabetic retinopathy of the Kaggle database into no diabetic or diabetic retinopathy, they used a well-known CNN called ResNet34. Among the pre-trained algorithms in ImageNet is ResNet34. To enhance the quality of images, they used a set of image processing techniques. The techniques are Gaussian filtering, and weighted addition. The images were 512 * 512 pixels in size and 35000 images were included in this study. A sensitivity of 86% and an accuracy of 85% were reported.

A study by [19] combined 3 pre-trained Convolutional Neural Networks. Adam optimizer was applied to update CNN parameters during training. The Adaboost algorithm was used to integrate these models. In this work, correctness was 88.21%, and the AUC was 0.94%.

In study [20] perceived red lesions and diabetic retinopathy images using enlarged 65*65 patches and CNNs. There were two networks used: a pre-trained VGG16 CNN and a custom CNN with 5 CONV layers, 5 maximum-pooling layers, and Fully connected layer. For the Messidor dataset, this research attained a recall of 0.94 and an AUC of 91.2%.

According to [21], all images were distinguished as referential diabetic retinopathy (refer to medium or higher stage) or non-referential diabetic retinopathy (no diabetic or slightly) by using 3 CNNs. Three datasets were used to create the images: Kaggle (88,702 images), DiaretDB1 (89 images), and E-ophtha (107,799 images). Prior to inputting them into network, the pictures were rescaled, cropped to 448 * 448 pixels, interpolated, and the FOV was eroded by 5%. In Kaggle database, they acheived AUC of 0.954. In E-ophtha, AUC of 0.949.

2.3 Anatomy of Eye

In the human body, vision is the most used sense. In fact, a significant portion of the brain is dedicated to visual processing, since most of the data that we receive is from the eyes. Often, the eye is compared to a camera as it changes light into data that our brains can process. They both use lenses to focus incoming light. While the eye uses a layer of cells called the retina to create a picture, the camera uses film. The eye, however, has a much superior ability to focus on objects of different sizes, contrast at a high speed, and luminosity than current cameras. On the back of the eye lies a multilayered sensory tissue that receives images formed by the lens and passes them on to the brain via the optic nerve.

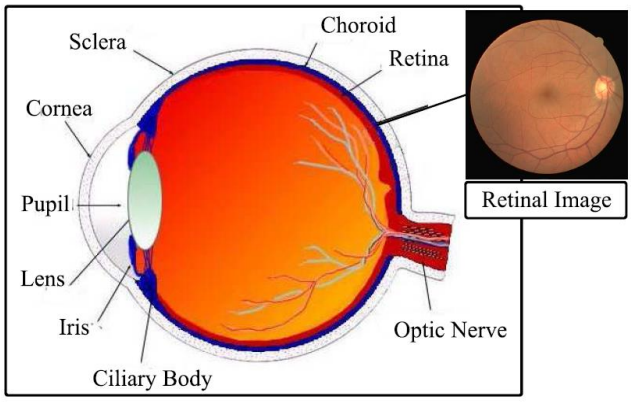


Figure 2. 1: Anatomy of Eye [46].

There are millions of photoreceptors in the retina that convert light into electrical impulses. Upon reaching the brain, the impulses are transformed into images by the optic nerve. Fundus images are often called retinal photography and are taken with the purpose of showing the eye's internal structure, including the retina, optic disk, retinal vessel tree, and macula (the small point in the retina where vision is most sensitive). During an eye examination, the fundus is the part of the eye that can be seen through the pupil. Images of the fundus retina are commonly used to study vessel features and vascular structure in many applications.

The retina has two types of photoreceptors, called rods and cones, respectively, based on their shapes. Rod cells are sensitive to fluctuations in contrast even at low light levels, which allows them to detect movement, however, they are inexact and insensitive to color. They are located in the peripheral part of the retina, which is responsible for night vision (scotoma). Conversely, cones are highly sensitive cells that can detect the presence of color very precisely. The macula is where they are located, and they are responsible for Photopic vision (day vision). This center portion of the macula is called the fovea, where the human eye can distinguish the finest details easily. Figure 2.2 illustrates the serious consequences of damage to the macula, which

can go unnoticed for some time while peripheral vision may be lost for a period of time without being noticed [12].

(1) (2)

Figure 2. 2: (1) Healthy Retina Vision. (2) Defected Retina Vision [46].

Approximately 1.2 million nerves connect all the photoreceptors with the brain [13]. The optic nerve is the only nerve bundle leaving the eye.

The blind spot in the retina results from the lack of photoreceptors on the retina where the optic nerve leaves the eye.

It is essential to prevent preventable vision loss by detecting and treating retinopathy as soon as possible. Human retinal photography was the first form of imaging the retina. An imaging method, fluorescein angiography is often used for diagnosing and treating retinal diseases [14]. Fluorescein angiography involves injecting fluorescein into the bloodstream. The recent development of digital photography of the retina has been made possible by the recent progress in communications technology and information. A digital fundus image of the retina currently plays a critical role in diagnosing various eye conditions in clinics, as well as detecting systemic diseases, such as diabetes, arteriosclerosis, and hypertension [15].

2.4 Clinical Characteristics of Diabetic Retinopathy

DR is detected in retinal images by the emergence of various types of lesions. Lesions causing these symptoms are microaneurysms (MA), haemorrhages (HM), and soft and hard exudates (EX) [16];

- MA- are the earliest signs of diabetic retinopathy, which appear as tiny, red, round dots on the retina. This is caused by the weakening of the vessels' walls. The size of the red dot is less than 125μm and the margins are sharp. The MA are classified into six types and using AOSLO reflectance and conventional fluorescein imaging, the types of MA can be identified.
- HM - emerge as bigger patches on the retina, which have an irregular margin and are wider than 125 μm in size. In HM, there are two types: flame (superficial HM) and blot (deeper HM).
- HE - These are leaks of protein and lipoproteins from abnormal vessels of the retina. Their appearance is that of small white or yellow-white deposits with sharp margins. On the outer part of the retina, they tend to appear as clusters or rings.
- SE - These exudates (also known as cotton wool) emerge as white patches on the retina as a result of swelling of the nerve fibers. Usually, it is round or oval in shape.

2.5 Stages of the Diabetic Retinopathy

DR can be classified into five stages when these lesions are present: no diabetic retinopathy, mild DR, moderate DR, severe DR and proliferative DR, which are described briefly below in Table 2.1. In Figure2.1, we see a sample of images of the DR stage.

Table 2. 1: DR Levels and Associated Lesions [3]

DR Severity Level	Lesions
No DR	Absent of lesions
Mild non-proliferative DR	MA only
Moderate non-proliferative DR	More than just MA but less than severe DR
Severe non-proliferative DR	Any of the following: • more than 20 intraretinal HM in each of 4 quadrants • definite venous beading in 2+quadrants • Prominent intraretinal microvascular abnormalities in 1+ quadrant • no signs of proliferative DR
Proliferative DR	One or more of the following: vitreous/pre-retinal HM, neovascularization

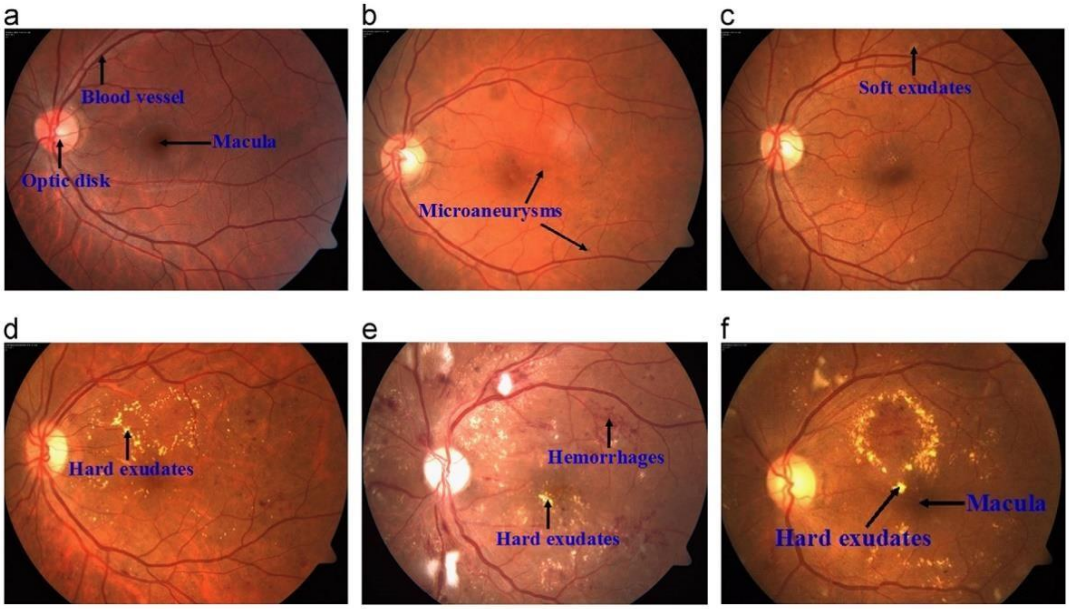


Figure 2. 3: Different Stages; (a) Normal (b) Mild DR (c) Severe DR (d) Severe Non-Proliferative DR (e) Proliferative DR (f) Macular edema [3]

Chapter 3

DEEP LEARNING AND CONVOLUTIONAL NEURAL NETWORKS

3.1 Introduction

Deep learning is a form of machine learning (ML) that is inspired by the human nervous system and similar to the structure of the brain [22]. The field of Machine Learning (ML) is part of Artificial Intelligence. Machine learning is used to teach machines how to handle data more efficiently. In concrete terms, machine learning is applied to interpret patterns in the data and extract relevant information. As defined in Nils J. Nilsson's book Introduction to Machine Learning: As a part of artificial intelligence (AI), machine learning refers to changes in systems that perform specific tasks. Recognizing, diagnosing, planning, controlling, and predicting are tasks that are involved. Changes may involve enhancing already-optimized systems or creating new ones from scratch [23]. An overview of the stages in an ML workflow is shown in the Figure 3.1. Typically, in a machine learning application, the processing takes place in two phases:

1. The learning phase: This phase occurs during the development/construction of the application. This process consists of the following steps:

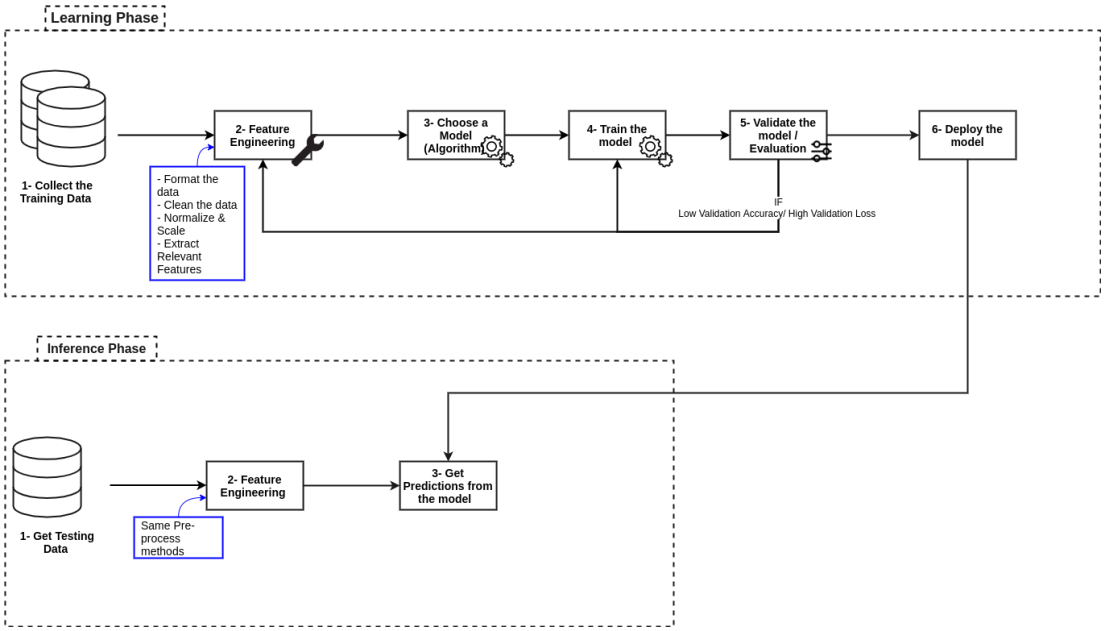


Figure 3. 1: Diagram of Typical Machine Learning(ML) Workflow [43].

- Data gathering: It is the first step in which the data scientist gathers all the data possible.
- Data preparation or feature engineering: It consists of cleaning and preparing the data for learning algorithms.
- Select a model: Different algorithms are appropriate for different tasks, so it is up to the developer to choose the best one.
- Training Model: During training the model, its parameters are updated to minimize an error function. The error function measures how frequently the algorithm gives incorrect predictions.
- Testing the model: The developer runs the model using new data and calculates the accuracy of the model. The model can be deployed if it performed well. Otherwise, the developer will have to prepare the data better or tweak the algorithm parameters.
- Deployment: Basically, the model is installed on a platform that can be accessed by clients.

2. The Inference Phase: In the Inference phase, we apply our model created in the previous phase to predict the results of the data we are given by the user.

The evolution of machine learning has progressed from purely statistical approaches to bio-inspired methods such as neural networks. As a result, we have classified them as Traditional Machine Learning models and Deep Learning models. In traditional machine learning techniques, only natural raw data can be analyzed. It takes significant domain knowledge to design the ideal feature extractor that transforms raw data into a useful format (mainly a feature vector) from which a classifier can detect patterns or classify the input. This problem can be solved by Deep Learning.

Assuming that we think about how to classify different dog's images, a classic machine-learning technique will use either the pixel values as features or a user-defined function that extracts edges, shapes, and colors to feed into the classifier. However, if we are using deep learning, the convolutional neural network itself will extract the features and learn which ones are useful.

Deep learning, by definition, is a machine learning technique suited to teaching computers what toddlers learn by example. A deep learning algorithm uses a bio-inspired architecture known as Deep Neural Networks. With this architecture, raw data can be fed into the machine, which can automatically extract the desired representation of the data (Classification, Detection, etc.). It powers many amazing applications including driverless cars, language translation, even some hospitals use predictive applications to manage patients in the emergency rooms.

3.2 Neural Network

In order to understand what deep learning is, we must describe the basic components of it: Neural Network. A deep neural network (DNN) is an artificial neural network (ANN) with multiple layers between the input and output layers [24]. Different types of neural networks have different components, but they all have neurons, synapses, weights, biases, and functions. Figure 3.2 shows the diagram of Deep Neural Network.

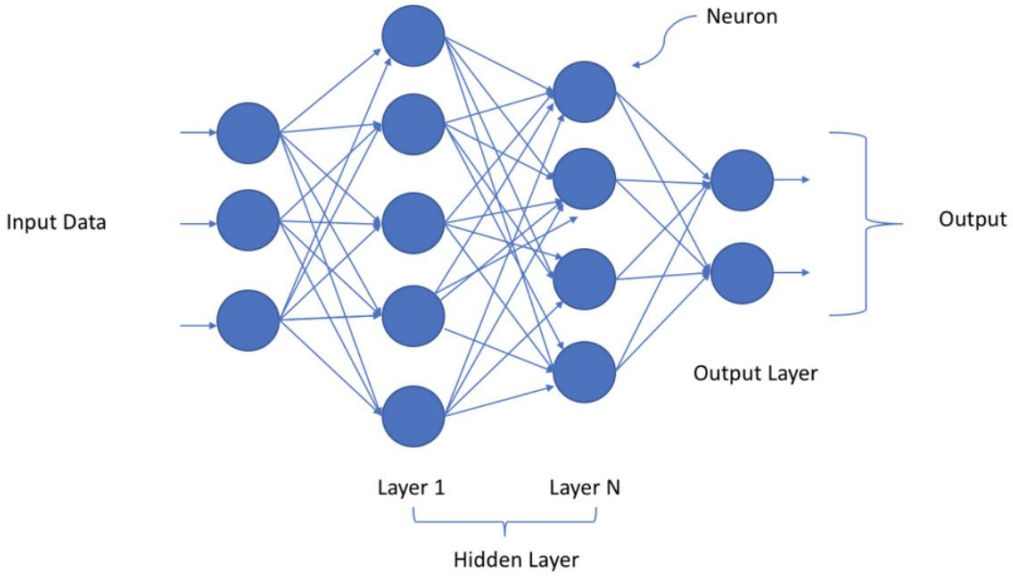


Figure 3. 2: Diagram of Deep Neural Network [44]

Deep neural networks are bio-inspired algorithms. A set of layers represents a vertical combination of nodes which is illustrated graphically in Figure 3.2. As each node is fed raw input data, it applies a linear function followed by a non-linear activation function, and each node represents processing unit. It transforms the data representations at each level and builds a statistical model that helps us later perform classifications and detections. Output consists of a series of probabilistic nodes, each of which indicates the chance of categorizing the input. When choosing the right non-linear functions (Activation functions) and composing their representations on

multiple levels, one can learn very complex functions according to the universal approximation theorem [25]. This number of processing layers and transformations is what inspired the label "deep".

Using neural networks, we can approximate any function, provided there is enough data and computation time. Initially, the network is naive, it does not know how to map inputs to outputs. In order to train the network, we update its parameters according to the losses the network makes each time. It is important to know how poorly the system predicts the real outputs in order to find such parameters. Thus, we calculate a loss function (also called a cost), which is a measure of the error in prediction. For instance, regression and binary classification problems often require the use of mean squared losses.

$$\text{Mean Square loss (MSL)} = \frac{1}{n}\sum_{i=1}^{n}(U_i - \hat{U}_i)^2 \quad (3.1)$$

U_i represent the true labels and $\hat{U}_i$ represent the predicted labels, where n is the number of training examples. Our goal here is to minimize this loss in relation to the network parameters and to find configurations where the loss can be minimized while the network can predict the correct labels with high accuracy. Gradient descent is used to find this minimum. Gradients are the slopes of loss functions and indicate the direction of the fastest change. We therefore wish to follow the gradient down (downwards) in order to reach the minimum in the shortest amount of time.

The training process consists of two principal passes, which we'll explain below:

1-FeedForward: In neural networks, Feedforward is the process of turning input into output. In other words, it is the outcome of feeding a neural network with data. Here is a closer look at one processing unit (Figure 3.3): Each processing unit receives some

data, multiplies each input by its weights, and adds these products using an activation function.

$$W_1' = W_1 - \alpha \frac{\partial \ell}{\partial W_1}$$

In summary, in a feedforward pass, we pass the input x through a linear transformation L1 with weights W1 and biases b1. In the next step, a sigmoid operation S is applied, followed by a linear transformation L2. We then calculate the loss. This loss is used as an indicator of how well the network performs.

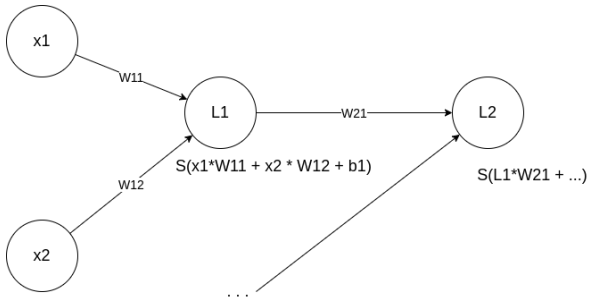


Figure 3. 3: Zoom into One Processing Unit [43].

2 – Backpropagation: By using backpropagation, one can train multilayer networks, and this is just extending the chain rule from calculus. Through the network, we propagate the gradient of loss backwards. There is a gradient between the inputs and outputs of every operation. By sending gradients backward, we multiply the incoming gradients by the operation's gradient. This simply involves the chain rule to determine the gradient of the loss w.r.t weights.

$$\frac{\partial \ell}{\partial W_1} = \frac{\partial L_1}{\partial W_1} \frac{\partial S}{\partial L_1} \frac{\partial L_2}{\partial S} \frac{\partial \ell}{\partial L_2} \tag{3.2}$$

The weights are updated using this gradient, along with a learning rate α.

$$W_1' = W_1 - \alpha \frac{\partial \ell}{\partial W_1} \tag{3.3}$$

We set the learning rate α efficiently. Not too small, which would slow down the convergence, nor too big, which would prevent the iterative method from reaching a minimum. The re-iteration of these two phases (Feedforward and Backpropagation) is what makes deep neural networks learn, and it's what constitutes the fundamental algorithms of deep learning.

3.3 Convolutional Neural Network

A convolutional neural network is presently regarded as one of the most prominent algorithms for doing deep learning with image data. Essentially, convolutional neural networks use 2D images as input and convolve them with filters to produce the desired outputs, using deep learning algorithms to train huge datasets with millions of parameters [26]. A CNN architecture varies according to the number and type of layers it implements. In a continuous response system, the regression layer should be placed on the end of the network. However, in categorical response systems, the classification layer should also be included. Every layer of the CNN has neurons arranged in a three-dimensional arrangement, and each layer outputs from a three-dimensional input. As we use 3D inputs in our application, the input layer holds the height, width, and RGB values of the images as dimensions. After that, neurons are added to the areas of the image in order to produce a three-dimensional output as shown in Figure 3.4.

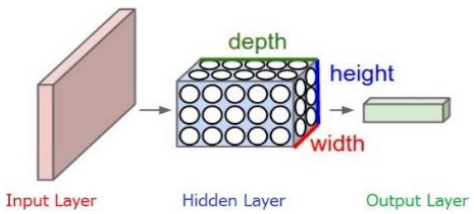


Figure 3. 4: Convolutional Neural Network [45].

CNN configurations contain a myriad of hidden layers. In each layer, activation volumes can be altered using differentiable functions. To build a CNN configuration, four different types of layers are used, as illustrated in Figure 3.5.

1. **Convolutional Layer (CONV):** A convolution is a mathematical operation that involves applying a filter to an n-dimensional array (image). A convolutional filter is used to obtain an activation map from input data.
2. **Rectified Linear Unit Layer (ReLU):** It filters out negative values to only provide positive ones, thus allowing for a much faster training speed.
3. **Pooling Layer (POOL):** Reduces the number of parameters for a simpler output by performing nonlinear down sampling.
4. **Fully Connected Layer (FC):** Produces a vector of N dimensions with N being the number of classes, giving the probability scores for each class.

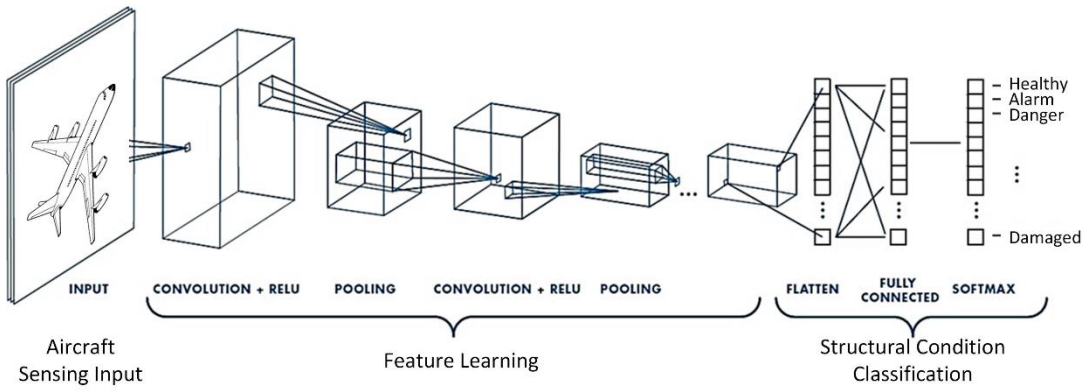


Figure 3. 5: The diagram of a CNN Used to Monitor Aircraft Systematic Health [27].

3.4 Standard CNNs

In this section, we list the key CNN Architectures that lie at the foundation of today's computer vision technologies.

AlexNet: Alextnet is based on the LeNet architecture [28], which is regarded as the first attempt to build a convolutional neural network. There are five convolutional layers and three fully interconnected layers in the architecture. In 2012, these 8 layers combined with two new concepts - MaxPooling and ReLU activation - gave their model an edge. In order to achieve such extraordinary efficiency at the time, the author used many techniques, including Dropout [29], Augmentation, and Stochastic Gradient Descent with momentum [30].

VGG16: In 2014, VGGNet took the second place for ImageNet challenge. VGGNet has two rules of thumb that should be followed. First, Only the number of kernels differs between the two convolution layers. Second, at every layer of pooling, we shrink the image by half. Figure 3.6 below shows the various layers and their configurations.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 3. 6: Different Configurations of VGG Network [31].

ResNet: The first ResNet (Residual Network) was proposed in 2015 at the Imagenet competition. This reduced the top-5 error rate to 3.57%, which is less than human error. An innovative approach was introduced called "skip connections" [32]. Skip connections are direct links between two layers that do not follow one another. With this method, the vanishing gradient problem is prevented by adding the input x to the output after a few convolution layers. Here, we try to figure out the residual (difference) between a layer's output and its input. As soon as we add the actual input, we get the same result as before shown in Figure 3.7.

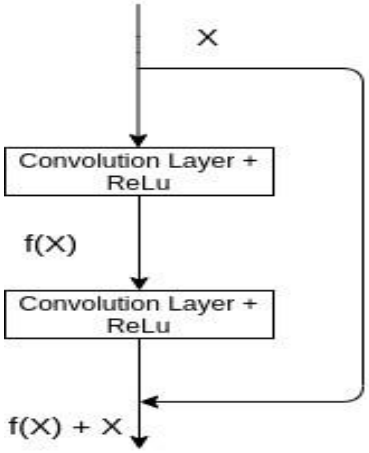


Figure 3. 7: Basic Building Block of Resnet Architecture [43]

Chapter 4

GABOR FILTERS

4.1 Introduction

The Gabor function introduced by Dennis Gabor in 1946 uses Gaussian functions modulated by sine waves to process signals and images [33]. Nevertheless, these functions had played a significant role in harmonic analysis prior to 1946 and in quantum mechanics. The Gaussian representation of signals through the use of complex exponentials modulated by Gaussian functions has been shown by Gabor to be optimal in terms of minimizing the joint uncertainty in the combined time-frequency space. In this way, both time resolution and frequency resolution are best suited by these functions. Daugman developed Gabor into a two-dimensional system [34]. In several years of research on the Human Visual System, Marcelja [35] and Daugman [34] developed a model of how the visual cortex is organized which led to the discovery that simple cells in the cortical structure can be modeled by Gabor functions. Based on an analytical method developed in one-dimensional case by Bastiaans [36], Porat and Zeevi [37] have shown that the Gabor scheme provides a suitable representation for visual information in a two-dimensional space of combined frequency and position. The findings of these studies have generated much interest in the computer vision and image processing communities, as in these communities there has been a constant endeavor to develop a mathematical model for human vision and apply it to applications. Since it is often assumed that our visual system evolved in response to the structure of our environment, it should be able to process visual data

optimally. Throughout the years, Gabor's original work has been extended and used to represent, enhance, restore, segment, and compress images, all the way up to identification and interpretation. The Gabor wavelet framework has enjoyed great popularity among wavelets enthusiasts due to the Gabor functions' popularity. For many image processing applications, Gabor wavelets provide optimal (minimal), inversely-related effective widths in position and frequency.

Firstly, we will introduce Gaussian functions as a part of Gabor function in this chapter. Next, we examine Gabor filters in more detail. It also includes a description of the various types of Gabor filters. At the end, this function's application has been explained.

4.2 Window Function

Filters can be represented mathematically as window functions that can be applied to signals for processing. Window functions are typically constants within and zero outside of a specified region. Using it, you can analyze certain regions and remove undesirable parts. With regard to the practical cases, a window function is appropriate inside the desired region but degrades rapidly outside, with an emphasis on trying to achieve compatibility with the ideal case.

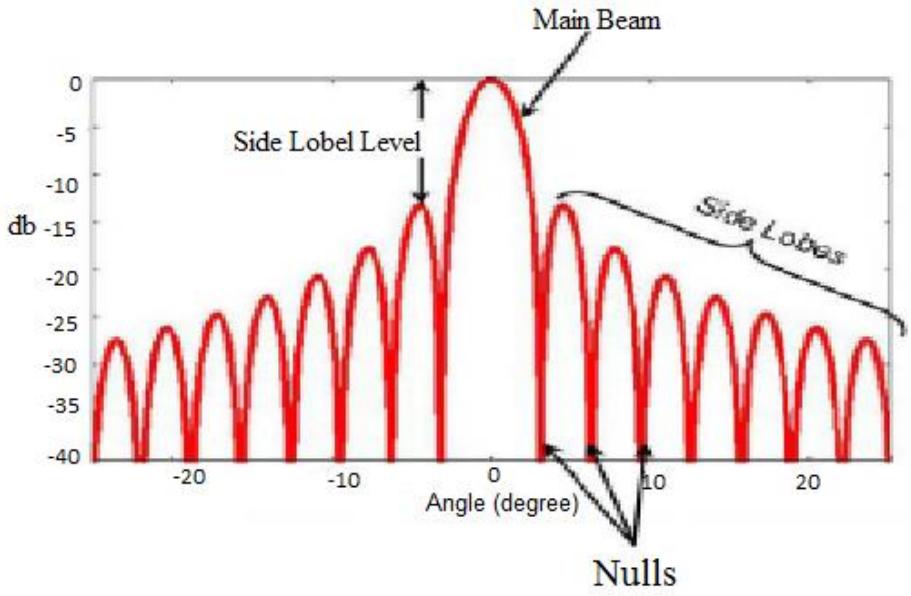


Figure 4. 1: Frequency Response of Window Functions [46]

Using the main lobe as the basis determines the ability to resolve comparable strength signals, whereas using the side lobes as the basis lets you resolve disparate strength signals. Figure 4.1 shows the frequency response of the window function.

4.3 Gaussian Window

Due to some unique characteristics of the Gaussian function, along with the Fourier transform and its derivative, it is selected. In time-frequency analysis, this property is extremely useful. Moreover, Gaussian functions are able to resolve disparate strength signals due to their window-like properties. Figure 4.2 shows a Gaussian window and its frequency response.

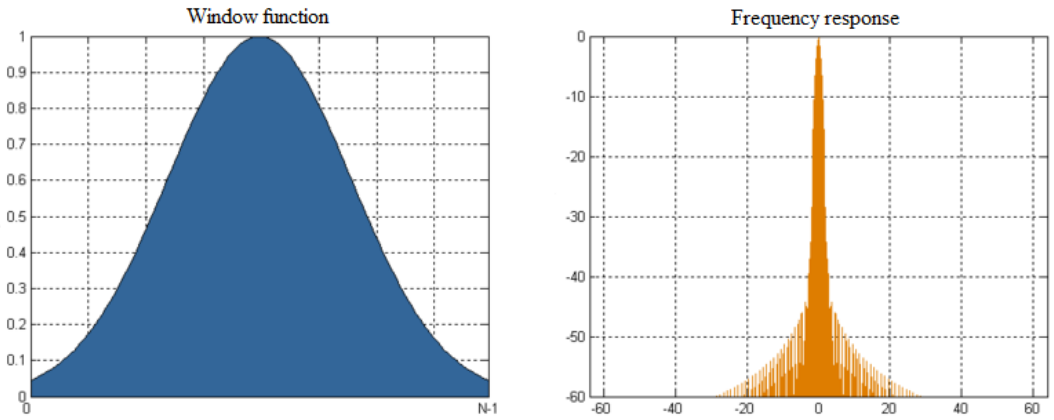


Figure 4. 2: Gaussian Window and its Response [46].

4.4 Gabor Filters

Gabor filters are linear filters whose impulse responses are defined by harmonic functions multiplied by Gaussian functions. The Fourier transform of a Gabor filter's impulse response is the convolution of the Fourier transforms of the harmonic function and the Fourier transform of the Gaussian function, due to the convolution property of multiplication (Convolution theorem).

The Gabor filter can be designed to dilate and rotate as needed, so that they are closely related to Gabor wavelets. However, in general, Gabor wavelets are not expanded since this would require computation of orthogonal wavelets, which can take a lot of time. As a result, a filter bank consisting of Gabor filters with different scales and rotations is created. The signal is convolved with the filters, resulting in a Gabor space. There is a close connection between this process and processes occurring in the primary visual cortex. As a result, the Gabor space is widely used in applications like image processing, edge detection, and texture segmentation. In an image, activations related to a certain spatial location are very distinct from one another. In addition, it is possible to extract important activations from the Gabor space to create a sparse object representation.

4.5 One-Dimensional Gabor Filter

A 1-D Gabor filter was first invented by D Gabor in 1946. Gabor filters can be described as sinusoidal planes modulated by Gaussian envelopes that have a set frequency and orientation.

$$h(x,y) = s(x,y)\, g(x,y) \tag{4.1}$$

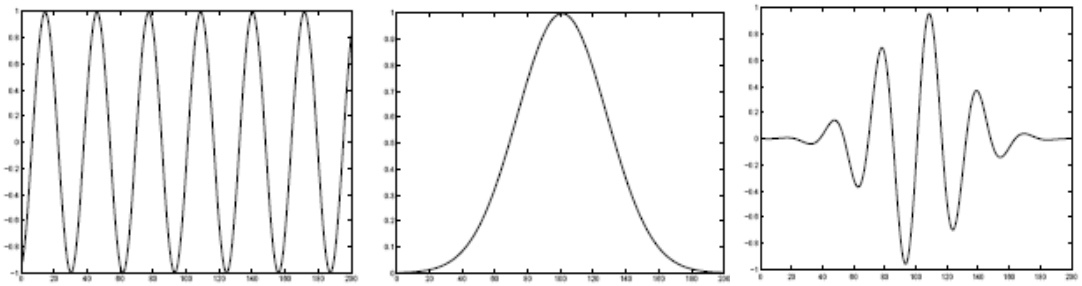
s(x, y) is a complex Sinusoidal function and g(x, y) is a Gaussian shaped function denoted by its envelope.

$$s(x,y) = e^{-2\pi j(u_0 x + v_0 y)} \tag{4.2}$$

$$g(x,y) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x{}^2} + \frac{y^2}{\sigma_y{}^2}\right)} \tag{4.3}$$

Here (u0, v0) is the spatial frequency, and $\sigma_x$and $\sigma_y$are the standard deviation on the x and y axes, respectively.

An easy way of showing Gabor filter responses in one dimension has been devised. The filter responses presented below are summarized in a one-dimensional graph and provide an introduction to our two-dimensional filter approach. Graph 4.3 shows a sinusoid, Gaussian Kernel for a Gabor filter in one dimension.



(a) (b) (c)

Figure 4. 3: (a) Sinusoidal Waveform (b) Gaussian Kernel (c) Gabor Filter Response [46].

4.6 Two-Dimensional Gabor Filter

Gabor's original proposal for filter collection was extended by Daugman to two dimensions. In Figure 4.4, a Gabor filter is shown centered at 2D frequency coordinates.

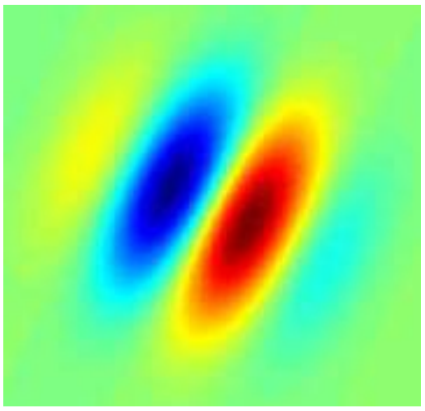


Figure 4. 4: 2-D Gabor Filter Response [46]

$$g(x,y) = s(x,y)\, w_r(x,y) \tag{4.4}$$

In this case, S(x, y) is a complex sinusoidal, or carrier, in which the carrier can be introduced as follows:

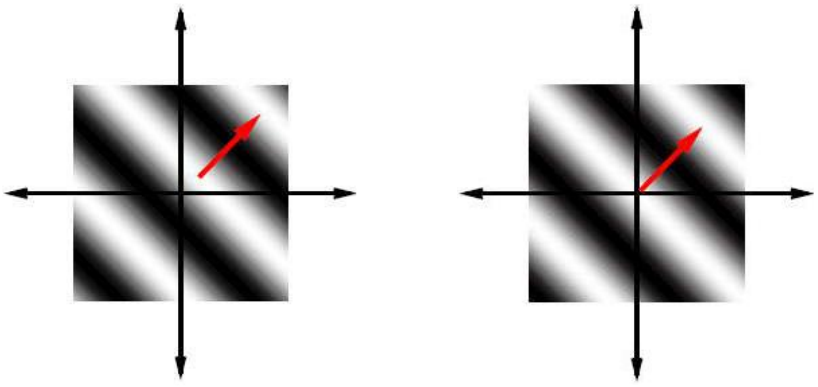
$$S(x,y) = e^{(j(2\pi(u_0 x + v_0 y) + p))} \tag{4.5}$$

$$\mathrm{Re}\big(s(x,y)\big) = \cos(2\pi(u_0 x + v_0 y) + p)) \tag{4.6}$$

$$\mathrm{Im}\big(s(x,y)\big) = \sin(2\pi(u_0 x + v_0 y) + p)) \tag{4.7}$$

As illustrated in Figure 3.5, the sinusoid has an imaginary component as well as a real component. Image dimensions are 128x128 pixels and the parameters are:

$$u_0 = v_0 = \frac{1}{80}\ ; p = 0.$$



(a) (b)

Figure 4. 5: (a) Real Part of Complex Sinusoidal (b) Imaginary Part of Complex Sinusoidal [46].

According to this representation, the complex sinusoid is expressed as follows:

$$S(x,y) = e^{(j(2\pi F_0(x\cos w_0 + y\sin w_0) + p))} \tag{4.8}$$

$w_r(x,y)$ is a 2-D Gaussian-shaped function, also known as the envelope.

$$w_r(x,y) = K\, e^{(-\pi((a^2(x-x_0)_r^2 + b^2(y-y_0)_r^2))} \tag{4.9}$$

K stands for Scales the magnitude of the Gaussian envelop, (x0, y0) for the peak of the function, a and b for the scale parameters of the Gaussian, and the r subscript stands for rotation. The Gabor filters will then be able to have different orientations.

Figure 4.6 shows a Gaussian envelope in the frequency domain. Clearly, the complexity of convolution varies with the size of the convolution mask.

$$(x - x_0)_r = (x - x_0)\cos\theta + (y - y_0)\sin\theta \quad (4.10)$$

$$(y - y_0)_r = -(x - x_0)\sin\theta + (y - y_0)\cos\theta \quad (4.11)$$

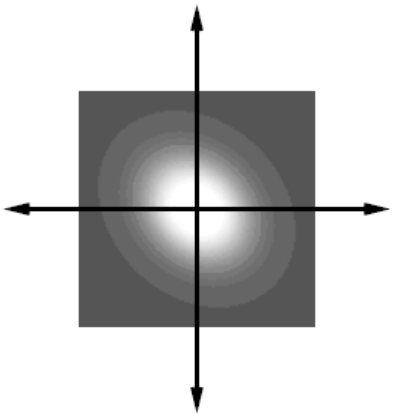


Figure 4. 6: The Gaussian Envelope [46]

4.7 Complex Gabor Function in Space Domain

In space domain, we can write the Gabor filter as follows:

$$g(x,y) = ke^{(-\pi(a^2(x-x_0)_r^2+b^2(y-y_0)_r^2))}\exp\left(j(2\pi(u_0x+v_0\,y)\,p)\right) \quad (4.12)$$

Where:

K : Scales the Gaussian envelope magnitude.

(a, b) : Scale the Gaussian envelop axes up and down.

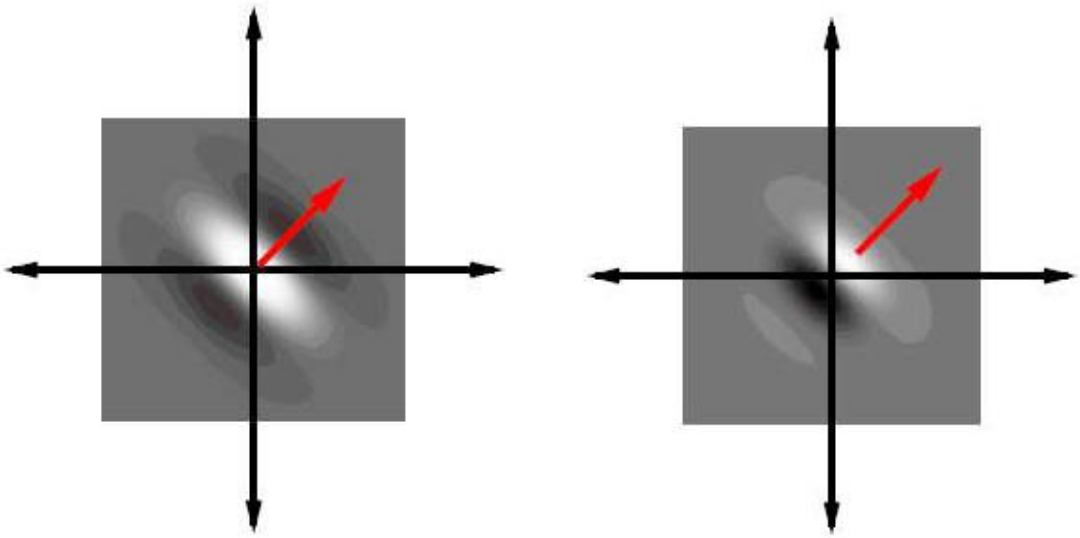
θ :Angle of rotation of the Gaussian envelope.

(x_0, y_0) : Gaussian envelope peak.

(u_0, v_0) : Cartesian frequency space of the sinusoidal carrier.

P : Sinusoidal carrier phase.

In Figure 4.7, the real and imaginary components of the complex Gabor function are displayed.



(a) (b)

Figure 4. 7: (a) Real part (b) Imaginary part [46]

(u_0 , v_0) The spatial frequencies of a sinusoidal carrier in Cartesian coordinate can be defined in Polar coordinates as (F_0, φ_0).

$$F_0 = \sqrt{u_0{}^2 + v_0{}^2} \tag{4.13}$$

$$\varphi_0 = \tan^{-1} \frac{v_0}{u_0} \tag{4.14}$$

We can also write:

$$u_0 = F_0 \cos \varphi_0 \tag{4.15}$$

$$v_0 = F_0 \sin \varphi_0 \tag{4.16}$$

Therefore, the Gabor function in polar coordinates is:

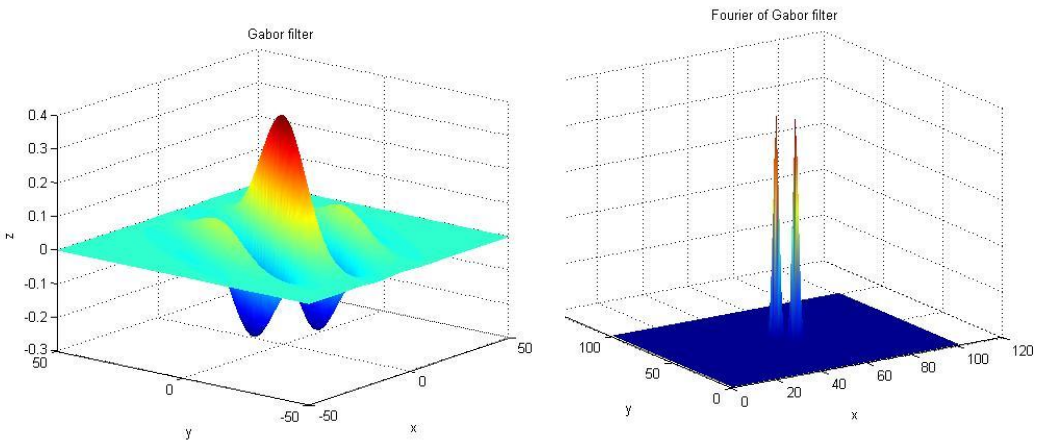
$$g(x,y) = k\, e^{(a^2(x-x_0)_r^2 + b^2(y-y_0)_r^2))} * e^{(j(2\pi F_0(x\cos\varphi_0 + y\sin\varphi_0) + p))} \tag{4.17}$$

As a result of applying the filter, edges along the rotation of the Gaussian envelope are highlighted.

Fourier transform of Gabor function defined as follows:

$$G(u,v) = \frac{k}{ab}\ \mathrm{e}^{(j(-2\pi(x_0(\mathrm{u}-u_0)+y_0(v-v_0))+\mathrm{p}))} * \mathrm{e}^{(-\pi\left(\frac{(\mathrm{u}-u_0)_\mathrm{r}^2}{a^2}+\frac{(v-v_0)_\mathrm{r}^2}{b^2}\right))} \qquad (4.18)$$

Figure 4.8 shows the Gabor filter with its Fourier transform.



(a) (b)

Figure 4. 8: (a) Gabor Filter (b) Fourier response of Gabor Filter [46]

4.8 Gabor Filters in Application

Gabor filters are commonly used band pass filters that can be applied successfully to a variety of applications including texture segmentation/classification, target detection, character recognition, fingerprint recognition, face recognition, image analysis, compression, and edge detection. As part of this thesis we are introducing edges in oriented structures, and using the Gabor filter we detect them in different orientations. Below are given few application of Gabor filter presented in the literature.

Descriptive texture description: Text descriptions have improved tremendously through Gabor filters, and it contributes to any kind of research [34].

Enhancing the image information: A Gabor filter can boost the information in an image for surveillance purposes in dark-lit situations or when noise is present [35].

Fingerprint: The use of Gabor filters operating at different frequencies and with different orientations has been used to identify and extract text-only regions from complicated image documents (both grayscale and color). This can be accomplished because text is marked by high frequency components, while pictures are relatively smooth in appearance [39].

Script Identification: In the processing of document images, Gabor features may be used to identify and position the script of each word on the page [40].

Chapter 5

METHODOLOGY AND DATABASE

5.1 Introduction

The ability to measure the performance of a system subjectively or objectively can be determined by developing index measures that enable an exact measurement of its performance [41]. It is necessary to use many different indices in order to achieve a good estimation of classification performance, even for simple tasks like binary classification. Binary classification considers instances of a set to fall into one of two categories: belong to a class or not. The classification scheme will yield four outcomes: positive classes classified correctly (true positive, TP), positive classes classified incorrectly (false positive, FP), negative classes classified correctly (true negative, TN), and negative classes classified incorrectly (false negative, FN) [42]. It is also possible to have a balance in the prediction set, meaning that the number of positives and negatives are equal, or an unbalanced prediction set, meaning there are more instances of one class than another, resulting in a difference between absolute and relative predictions. It is difficult to produce a unique metric for evaluating prediction ability because of all these variables. It is because of this reason that binary classification uses many different indexes. An overview of some of the most commonly used indexes for binary classification follows below.

5.2 Confusion Matrix

Four possible outcomes can be represented as a confusion matrix, or contingency table. Figure 5.1 illustrates how information is represented in the table. Confusion matrixes

summarize all the information about a problem. The other indexes are calculated using the information in the table. This confusion matrix can readily be applied to multi-class classification problems simply by adding rows and columns for as many classes as predicted.

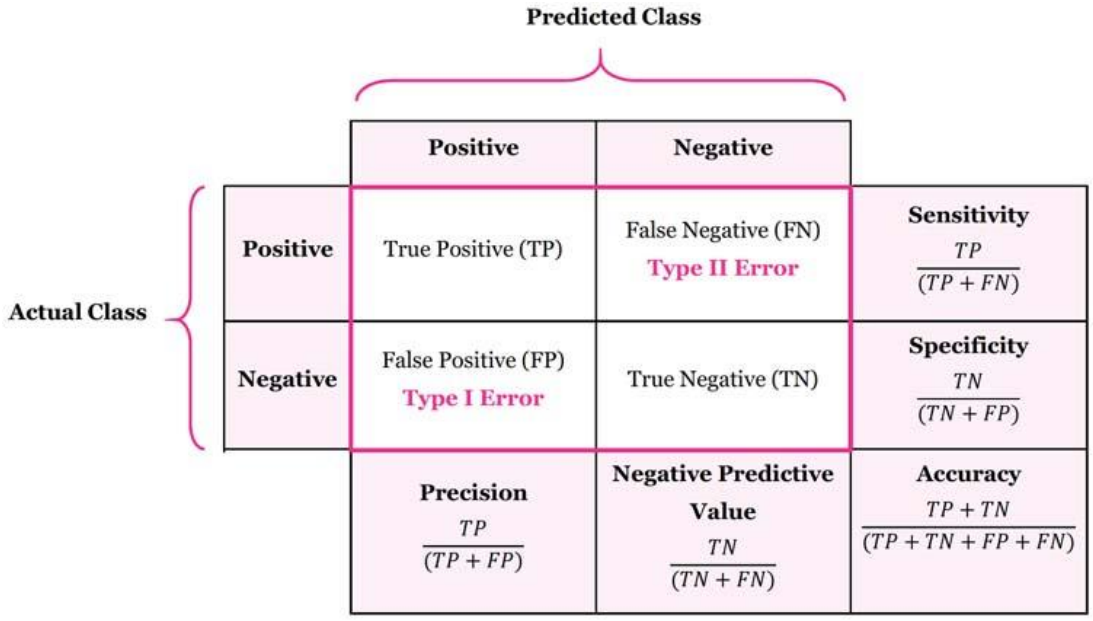


Figure 5. 1: Confusion Matrix for Binary Classification [38].

5.3 Sensitivity and Specificity

Confusion matrix indexes like Sensitivity and Specificity are the most important measures. Recall (eq. 5.1), also termed as sensitivity, an indication of how often actual positive results are identified correctly. The specificity (eq. 5.2) measures whether or not true negatives are correctly identified as such as a result of actual testing. This can be expressed mathematically as:

$$SE = \frac{TP}{(TP+FN)} \times 100 \tag{5.1}$$

$$SP = \frac{TN}{(TN+FP)} \times 100 \tag{5.2}$$

In medical diagnosis, sensitivity is the model's capability of correctly identifying those

who have the disease, and specificity is its capability of correctly identifying those who don't.

6.4 False Positive and Negative Rates

The False Negative Rate (eq. 5.3) and False Positive Rate (eq. 5.4) index measures the likelihood of the model reporting false values. These indexes are known as type I and II errors in statistical hypothesis testing. Among medical diagnosis, they are of great importance for determining screening and testing strategies.

$$FPR = \frac{FP}{(FP+TN)} \times 100 \tag{5.3}$$

$$FNR = \frac{FN}{(FN+TP)} \times 100 \tag{5.4}$$

5.5 Accuracy

It is a measure of how successful a diagnostic test is based on factors such as its "sensitivity" and "specificity.". Sensitivity and specificity are two types of accuracy. These measures, as well as the related indices, TPR and FPR, are more beneficial than "accuracy" but do not describe diagnostic performance uniquely because they depend on an arbitrary choice of a decision threshold. The accuracy is given by equation 5.4.

$$\text{Accuracy} = \frac{\text{TP+TN}}{(\text{TP+TN+FP+FN})} \tag{5.4}$$

5.6 Precision

This is the ratio of true positives to the total number of true positives and false positives. Precision looks at how many false positives were thrown into the mix. A model with no bad positives (such FPs) has a 100% level of accuracy. As more FPs are added to the mix, the worse the precision will look. The precision is given by equation 6.5 below.

$$\text{Precision} = \frac{\text{TP}}{(\text{TP+FP})} \tag{5.5}$$

5.7 F1-Score

F-scores, also called F1-scores, are measures of a model's accuracy on a dataset. It is used to evaluate binary classification systems that categorize examples into 'positive' or 'negative' categories. The F-score is a way to combine the precision and recall of a model, and it is defined as its harmonic mean. It is given by the equation 5.6.

$$\text{fscore/f1score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5.6}$$

5.8 ROC Analysis

An analysis of the receiver operating characteristic (ROC) curve offers a visual representation of the trade-offs between sensitivity and specificity for different values of the test result. Moreover, it may be used to choose the most appropriate threshold for an application. The ROC curve can be plotted on a 2-dimensional unit square plot where the y-axis and x-axis represent the TPR (or sensitivity) and FPR (1-specificity), respectively, and are close to (0, 1). Figure 5.2 shows the ROC analysis for binary classification done in this thesis. Elaboration is given in the next chapter.

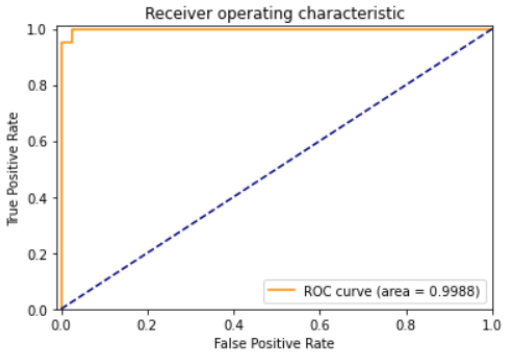


Figure 5. 2: ROC Curve of Binary Classification for Diabetic Retinopathy.

5.9 Database of Retina Images

Numerous databases with manual annotations of quality are publicly available. For this study, we chose the Messidor [47] database using the following criteria: (i) the

database annotation was handled by experienced physicians, (ii) the images were classified as suitable for medical analysis (gradable) or not (ungradable), and (iii) the database was public. Messidor is a database designed to facilitate studies on computer-aided diagnostics of diabetic retinopathy. The Messidor database contains 1200 color numerical images of the posterior pole acquired by three ophthalmology departments using a color video 3CCD camera mounted on a 45-degree nonmydriatic retinograph, the Topcon TRC NW6. The images were captured with 8 bits per color plane at 1440 by 960 pixels, 2240 by 1488 pixels, or 2304 by 1536 pixels. According to the annotations given in the dataset, the images are divided into train set (948 images) and test set (238 images). Some of the images were duplicated and mislabeled in the dataset, they are taken care of according to instruction given in the Messidor website [47]. This dataset has 4 classes as:

Class0 (R0): No diabetic retinopathy

Classs1 (R1): Mild case of DR

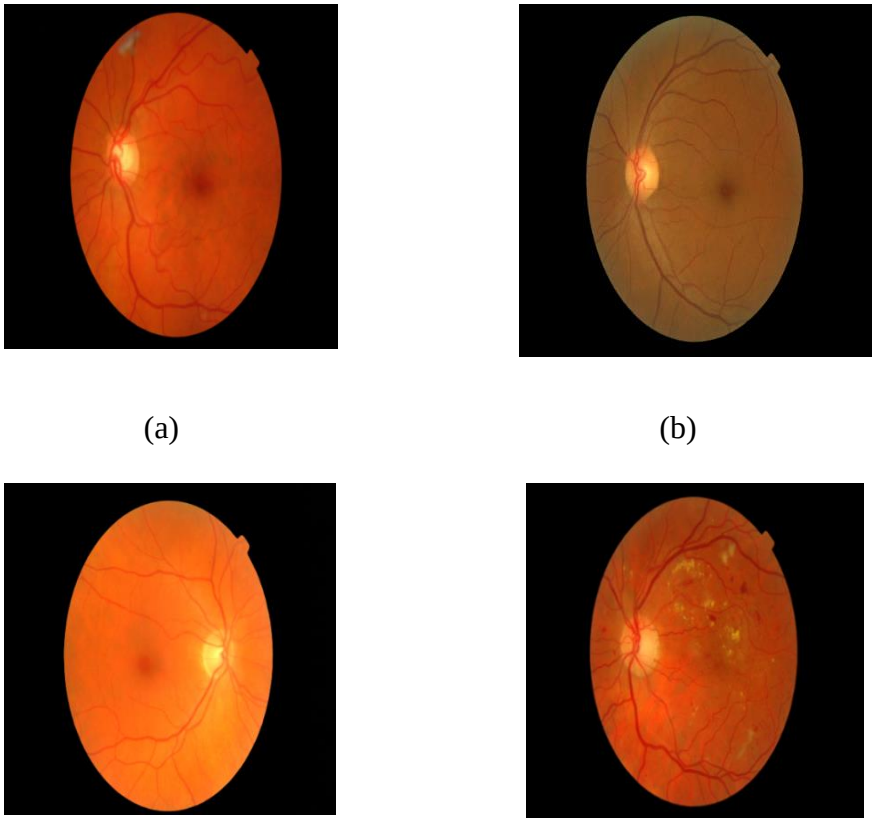
Class2 (R2): Severe case of DR

Class3 (R3): Proliferative case of DR.

Table 5.1 shows the different classes present in our dataset. Figure 5.3 shows the images from different classes.

Table 5. 1: Messidor Dataset Classes [47].

Messidor Dataset Division		
Class	**Number of Train Images**	**Number of Test Images**
Class0 (R0)	438	109
Class1 (R1)	119	30
Class2 (R2)	190	49
Class3 (R3)	202	50



(a) (b)

(c) (d)

Figure 5. 3: Sample Images: (a) NO DR (healthy retina) (b) Mild Case of DR (c) severe ase of DR (d)Proliferative Case of DR [47]

5.10 Decision Fusion

In decision fusion, output probabilities from different CNN orientations are given to the decision fusion block, and the final outcome regarding detection of the DR is obtained at the end of decision fusion. For our thesis, Sum Rule and Product Rule are applied in decision fusion to give the final decision as to whether it is an obvious DR or not.

The Sum Rule is given by equation 5.6 and Product Rule is given by equation 5.7 [43]. The probabilities can be expressed by p_i and pr_i respectively, where pr_i represents the probability given by Product Rule and p_i represents probability given by Sum

Rule, i = 1, 2..., n, and n represents the number of output probabilities. The n value in our case is 2.

$$p_i = \frac{\sum_{j=1}^{m} p_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}} \qquad i = 1, \ldots\ldots, n \qquad (5.6)$$

where p_{ij} denotes the i class probability of the CNN_j model.

$$pr_i = \frac{\prod_{j=1}^{m} p_{ij}}{\sum_{i=1}^{n}\prod_{j=1}^{m} p_{ij}} \qquad i = 1, \ldots\ldots\ldots, n \qquad (5.7)$$

Chapter 6

DIABETIC RETINOPATHY DETECTION USING GABOR FILTER BASED CNN

6.1 Introduction

In this Chapter we will discuss about our proposed method for detection of diabetic retinopathy. We are using ensemble of Gabor based CNN and image based CNN for detecting the DR. As we will see in this chapter, ensemble of Gabor based CNN and Image based CNN will outperform the single CNN with respect to accuracy, recall, f1score, precision, and Area under ROC curve. The following sections will take you through the steps done in this method.

6.2 Image Preprocessing

An image preprocessing step is necessary for the removal of noise, enhanced image features, and consistency of the images. The images are resized to 224×224 due to the architecture configuration of CNN (VGG16) we are using in our work. A cropping technique was also used to remove the unnecessary regions of the images. The result is shown in Figure 6.1.

(a) (b)

Figure 6. 1: (a) Original Image (b) Preprocessed Image

6.3 Proposed Method

In this thesis, we propose to use Ensemble of Gabor filter-based CNN models and image-based CNN model for detection of diabetic retinopathy. First, images are preprocessed, then the Gabor filters are applied to images at different orientations and fed to the CNN, and the output probabilities are then collected from the model. Additionally, images without Gabor filter are fed to a separate CNN model, and the output probabilities are calculated from the model. As a final step, these previously taken probabilities are fed into decision fusion, and a decision is made as to whether or not the patient has diabetic retinopathy. Figure 6.2 shows the whole process.

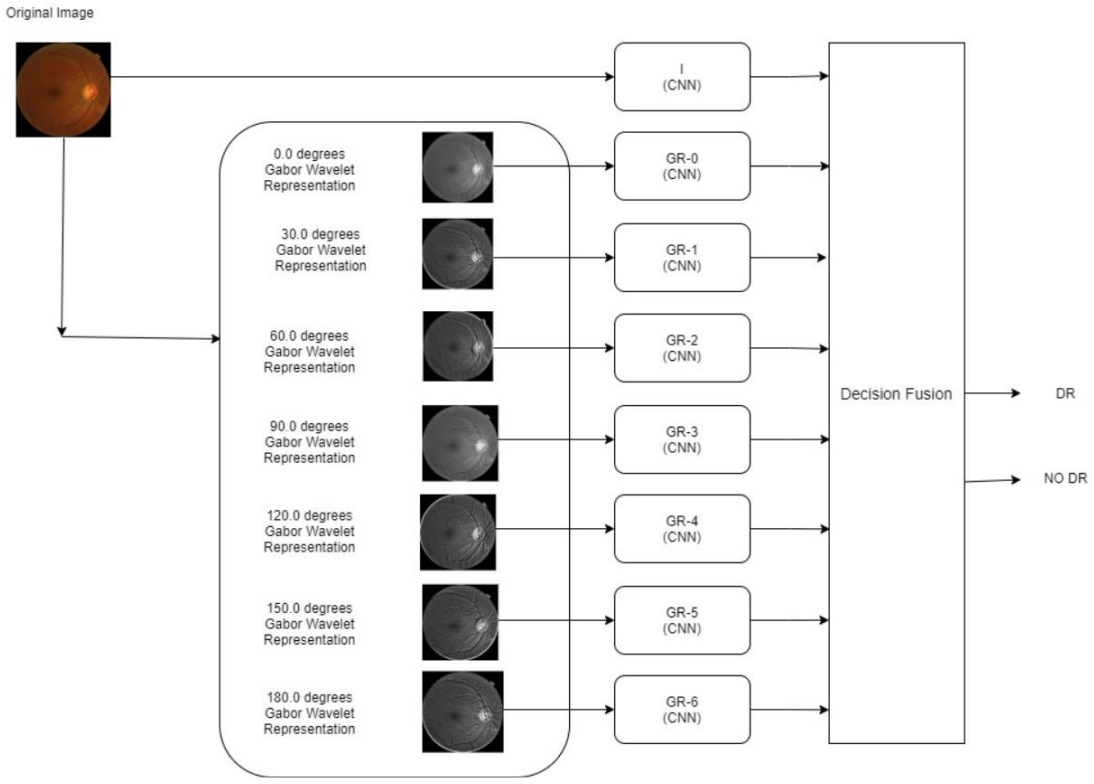


Figure 6. 2: Detection Of Diabetic Retinopathy According To The Proposed Approach.

In decision fusion, the output probabilities of CNNs are combined. The CNN provides probability values for DR, as well as for retinal images with no DR. The probabilities

can be expressed by p_i, where i ranges from 1 to n, and n represents the number of output probabilities. The n value in our case is 2.

A fusion of CNNs can also provide output probabilities. The probabilities are denoted by p_{ij}, where i ranges from 1 to n, and j ranges from 1 to m. In this case, m denotes the number of CNN models used. By combining probabilities from CNNs, the sum of probabilities (SP) is performed [43]. The SP is given by equation 6.1.

$$p_i = \frac{\sum_{j=1}^{m} p_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}} \qquad i = 1, \ldots\ldots n \tag{6.1}$$

where p_{ij} denotes the i class probability of the CNN_j model.

6.4 Gabor Filter Representation

Gabor filter is applied to the original images at 0^0, 30^0, 60^0, 90^0, 120^0, 150^0, 180^0 and corresponding output is collected as shown in Figure 6.3.

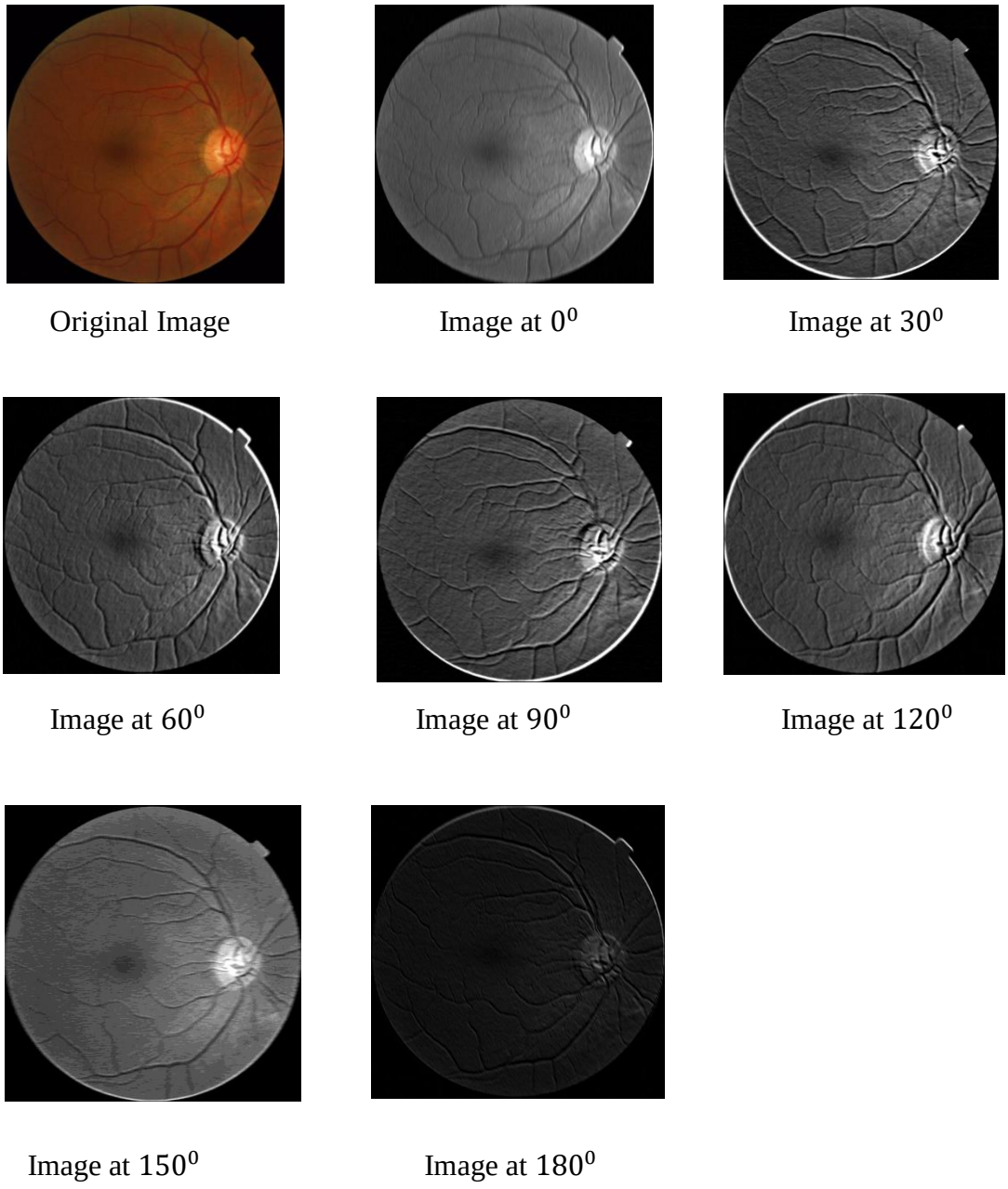


Figure 6. 3: Sample Images From Different Orientations.

Gabor filters are used for a variety of image processing and computer vision applications, including texture segmentation and face recognition. The vessels in a retinal image are connected and linear, so Gabor filters are ideally suited, as they are able to detect oriented features as well as being tuned to specific frequencies. There has been considerable evidence that the Gabor functions offer a good model of the receptive fields of simple cells in the striate cortex. The Gabor functions offer optimal joint resolutions both in Fourier and time space, and they provide a complete basis set when they phase shift, scale or dilate the original mother wavelet function.

Gabor filtering techniques have been used to detect blood vessels in retinal images based on Gabor wavelets. Real Gabor filter kernels (or mother wavelets) orientated at angle = π can be expressed as follows [46]:

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \mathrm{e}^{\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)\right)} \cos(2\pi f_0 \mathrm{x}) \tag{6.2}$$

Equation 6.2 gives us the standard deviation values $\sigma_x$and σ_y for x and y direction, as well as the modulating sinusoidal frequency fo. The mother wavelet is rotated to produce kernels at other angles. In this thesis, we use seven Kernels separated by an angle of 30^0 with a coordinate transform over the range [0, π].

$$\mathrm{x}' = \mathrm{x}\cos\theta + \mathrm{y}\sin\theta$$

$$\mathrm{y}' = -\mathrm{x}\sin\theta + \mathrm{y}\cos\theta \tag{6.3}$$

where $(x', y')$is a set of coordinates rotated by theta. In order to determine the parameters in equation above, namely σ_x and σ_y , it is necessary to take the size of the lines or curvilinear structures into account.

6.5 Model Generation

A list of the generated models can be found in Table 6.1 for the proposed Gabor based CNN. I represents the CNN model generated for original images. The remaining seven Gabor-based CNN models are the G0, G1, G2, G3, G4, G5, G6, G6, and G7 models. CNN model G0 is trained using retinal images taken from the Gabor filter at 0^0. Similarly G1, G2, G3, G4, G5, G6, G7 CNN models are trained using retinal images from $30^0, 60^0, 90^0, 120^0, 150^0, 180^0$ respectively.

GR model describes the fusion of Gabor-based CNN models. To develop this model, seven Gabor-based CNN models were trained on seven on images from seven directions as described above and then the output probabilities were merged together. Additionally, the I-GR model describes CNN model trained on original images and seven other CNN models trained Gabor filtered images from multiple directions (seven). The output probabilities are then combined for DR classification.

A Gabor-based CNN model and original retinal image-based CNN models are built using the VGG16 model. The VGG16 model is pre-trained. The proposed method uses this pre-trained VGG16 model to fine-tune it for DR detection. With this pre-trained model, you can better estimate network parameters and Messidor images.

Table 6. 1: Generated Models For Each CNN Pipeline.

Model	Explanation
I	CNN for original retinal images.
G0	CNN for 0^0 Gabor filter images.
G30	CNN for 30^0 Gabor filter images.
G60	CNN for 60^0 Gabor filter images.
G90	CNN for 90^0 Gabor filter images.
G120	CNN for 120^0 Gabor filter images.
G150	CNN for 150^0 Gabor filter images.
G180	CNN for 180^0 Gabor filter images.
GR	CNN for fusion of above seven Gabor filter images .
IGR	CNN for original images and GR .

6.6 Performance Evaluation (R0R1 vs R2R3)

The performance of the proposed model is evaluated for model accuracy, Area under ROC (Receiver Operating Characteristics) curve, sensitivity, specificity, f1score, precision, and false positive rate. Evaluation is performed on 238 test images of Messidor dataset. Since we are doing binary classification, we combined class0 (R0) and class1 (R1) as healthy retinal images. Also, we combined class2 (R2) and class3 (R3) as images with diabetic retinopathy. Therefore, we will evaluate our method using R0R1 (healthy retina) vs R1R2 (diabetic retinopathy). Table 6.2 shows evaluation with the performance metrics Accuracy. IGR (Sum Rule) is the proposed method throughout the evaluation section.

Table 6. 2: Evaluation With Performance Metrics Accuracy.

Class	**CNN for I**	**CNN for G**	**Method**	**Accuracy (%)**	**Decision Fusion**
R0R1 vs R2R3	VGG16	-	I	87.39	-
R0R1 vs R2R3	-	VGG16	G0	84.45	-
R0R1 vs R2R3	-	VGG16	G30	86.55	-
R0R1 vs R2R3	-	VGG16	G60	86.97	-
R0R1 vs R2R3	-	VGG16	G90	86.97	-
R0R1 vs R2R3	-	VGG16	G120	85.71	-
R0R1 vs R2R3	-	VGG16	G150	86.55	-
R0R1 vs R2R3	-	VGG16	G180	85.71	-
R0R1 vs R2R3	VGG16	VGG16	IGR	93.27	Product rule
R0R1 vs R2R3	VGG16	VGG16	IGR	**93.7**	Sum rule
R0R1 vs R2R3	-	VGG16	GR	92.85	Product rule
R0R1 vs R2R3	-	VGG16	GR	92.43	Sum rule

Accuracy is one of the more obvious metrics, since it measures all the cases correctly identified. If all classes have equal importance, it is most applicable. But sometimes model accuracy can be high while it will perform poorly by negative cases. In those cases, we use f1score, it is the harmonic mean of Precision and Recall, which gives a better indication of inaccurate classifications than Accuracy Metric. From Table 6.3 below we can see that f1score of our model is better than accuracy which makes our method more reliable for detection of diabetic retinopathy in retinal images. DF in Table 6.3 refers to decision fusion.

Table 6. 3: Evaluation Of the Model Using F1score.

Class	**CNN for I**	**CNN for G**	**Method**	**F1score(%)**	**DF**
R0R1 vs R2R3	VGG16	-	I	83.68	-
R0R1 vs R2R3	-	VGG16	G0	80.82	-
R0R1 vs R2R3	-	VGG16	G30	82.90	-
R0R1 vs R2R3	-	VGG16	G60	83.06	-
R0R1 vs R2R3	-	VGG16	G90	82.87	-
R0R1 vs R2R3	-	VGG16	G120	81.31	-
R0R1 vs R2R3	-	VGG16	G150	82.79	-
R0R1 vs R2R3	-	VGG16	G180	81.91	-
R0R1 vs R2R3	VGG16	VGG16	IGR	91.75	Product rule
R0R1 vs R2R3	VGG16	VGG16	IGR	**92.22**	Sum rule
R0R1 vs R2R3	-	VGG16	GR	91.28	Product rule
R0R1 vs R2R3	-	VGG16	GR	90.81	Sum rule

Similarly, Table 6.4 shows the performance of our method based on recall metrics. Recall is the measure of correctly identified positive cases from actual positive. It is important performance metrics while working with medical images because we do not want to miss the positive case. Table 6.4 indicates that our method gives better recall.

Table 6. 4: Evaluation With Performance Metrics Recall.

Class	**CNN for I**	**CNN for G**	**Method**	**Recall(%)**	**DF**
R0R1 vs R2R3	VGG16	-	I	77.77	-
R0R1 vs R2R3	-	VGG16	G0	78.78	-
R0R1 vs R2R3	-	VGG16	G30	80.8	-
R0R1 vs R2R3	-	VGG16	G60	76.7	-
R0R1 vs R2R3	-	VGG16	G90	75.7	-
R0R1 vs R2R3	-	VGG16	G120	74.7	-
R0R1 vs R2R3	-	VGG16	G150	77.7	-
R0R1 vs R2R3	-	VGG16	G180	77.7	-
R0R1 vs R2R3	VGG16	VGG16	IGR	89.89	Product rule
R0R1 vs R2R3	VGG16	VGG16	IGR	**89.89**	Sum rule
R0R1 vs R2R3	-	VGG16	GR	84.89	Product rule
R0R1 vs R2R3	-	VGG16	GR	89.89	Sum rule

Furthermore, Table 6.5 shows the performance of our method based on precision metrics. Precision metrics has importance while working on medical images because we do not want to miss out the positive case. It is evident from the Table 6.5 that our model achieves best precision among other models.

Table 6. 5: Performance Of Our Method Based On Precision Metrics.

Class	**CNN for I**	**CNN for G**	**Method**	**Precision(%)**	**DF**
R0R1 vs R2R3	VGG16	-	I	90.58	-
R0R1 vs R2R3	-	VGG16	G0	82.97	-
R0R1 vs R2R3	-	VGG16	G30	85.10	-
R0R1 vs R2R3	-	VGG16	G60	90.47	-
R0R1 vs R2R3	-	VGG16	G90	91.46	-
R0R1 vs R2R3	-	VGG16	G120	89.15	-
R0R1 vs R2R3	-	VGG16	G150	88.50	-
R0R1 vs R2R3	-	VGG16	G180	86.51	-
R0R1 vs R2R3	VGG16	VGG16	IGR	93.68	Product rule
R0R1 vs R2R3	VGG16	VGG16	IGR	**94.6**	Sum rule
R0R1 vs R2R3	-	VGG16	GR	92.70	Product rule
R0R1 vs R2R3	-	VGG16	GR	91.75	Sum rule

Furthermore, we evaluated our model based on specificity metrics which is measure of model for correctly identifying the negative cases. Table 6.6 shows the performance of our model based on specificity (TNR).

Table 6. 6: Performance Of Our Method Based On Specificity Metrics.

Class	**CNN for I**	**CNN for G**	**Method**	**Specificity(%)**	**DF**
R0R1 vs R2R3	VGG16	-	I	94.24	-
R0R1 vs R2R3	-	VGG16	G0	88.48	-
R0R1 vs R2R3	-	VGG16	G30	89.92	-
R0R1 vs R2R3	-	VGG16	G60	94.24	-
R0R1 vs R2R3	-	VGG16	G90	94.96	-
R0R1 vs R2R3	-	VGG16	G120	93.52	-
R0R1 vs R2R3	-	VGG16	G150	92.80	-
R0R1 vs R2R3	-	VGG16	G180	91.36	-
R0R1 vs R2R3	VGG16	VGG16	IGR	95.68	Product rule
R0R1 vs R2R3	VGG16	VGG16	IGR	**96.40**	Sum rule
R0R1 vs R2R3	-	VGG16	GR	94.96	Product rule
R0R1 vs R2R3	-	VGG16	GR	94.24	Sum rule

At the end we evaluate our model based on Area Under ROC curve which is plot of recall and False positive rate. It provides the aggregate performance of the model across all the classification being done. Table 6.7 below shows the performance of our model based on Area under ROC curve. It is evident from the table that our method shows promising results.

Table 6. 7: Model Performance Based On AUC.

Class	**CNN for I**	**CNN for G**	**Method**	**AUC(%)**	**DF**
R0R1 vs R2R3	VGG16	-	I	94.23	-
R0R1 vs R2R3	-	VGG16	G0	91.11	-
R0R1 vs R2R3	-	VGG16	G30	94.10	-
R0R1 vs R2R3	-	VGG16	G60	93.15	-
R0R1 vs R2R3	-	VGG16	G90	93.80	-
R0R1 vs R2R3	-	VGG16	G120	91.72	-
R0R1 vs R2R3	-	VGG16	G150	92.15	-
R0R1 vs R2R3	-	VGG16	G180	92.48	-
R0R1 vs R2R3	VGG16	VGG16	IGR	97.42	Product rule
R0R1 vs R2R3	VGG16	VGG16	IGR	**98.2**	Sum rule
R0R1 vs R2R3	-	VGG16	GR	97.60	Product rule
R0R1 vs R2R3	-	VGG16	GR	98.00	Sum rule

A public dataset Messidor with approximately 1200 color fundus images was used to assess the performance of the proposed methods. 749 images from this dataset are used as train and 238 images are used for testing the model. In the dataset, the images were divided into 4 classes: No DR(R0), mild case of DR (R1), severe case of DR(R2), proliferative case of DR (3). Mild case of DR is hard to classify since its almost to no DR level. Therefore, we have transformed this problem to binary classification by combing R0, R1 together making it class0 and R2, R3 together by making it class1. Class0 means no diabetic retinopathy and class1 corresponds to diabetic retinopathy.

Confusion matrix and ROC for different models are shown in Figure 6.4, Figure 6.5 Figure 6.6 and Figure 6.7.

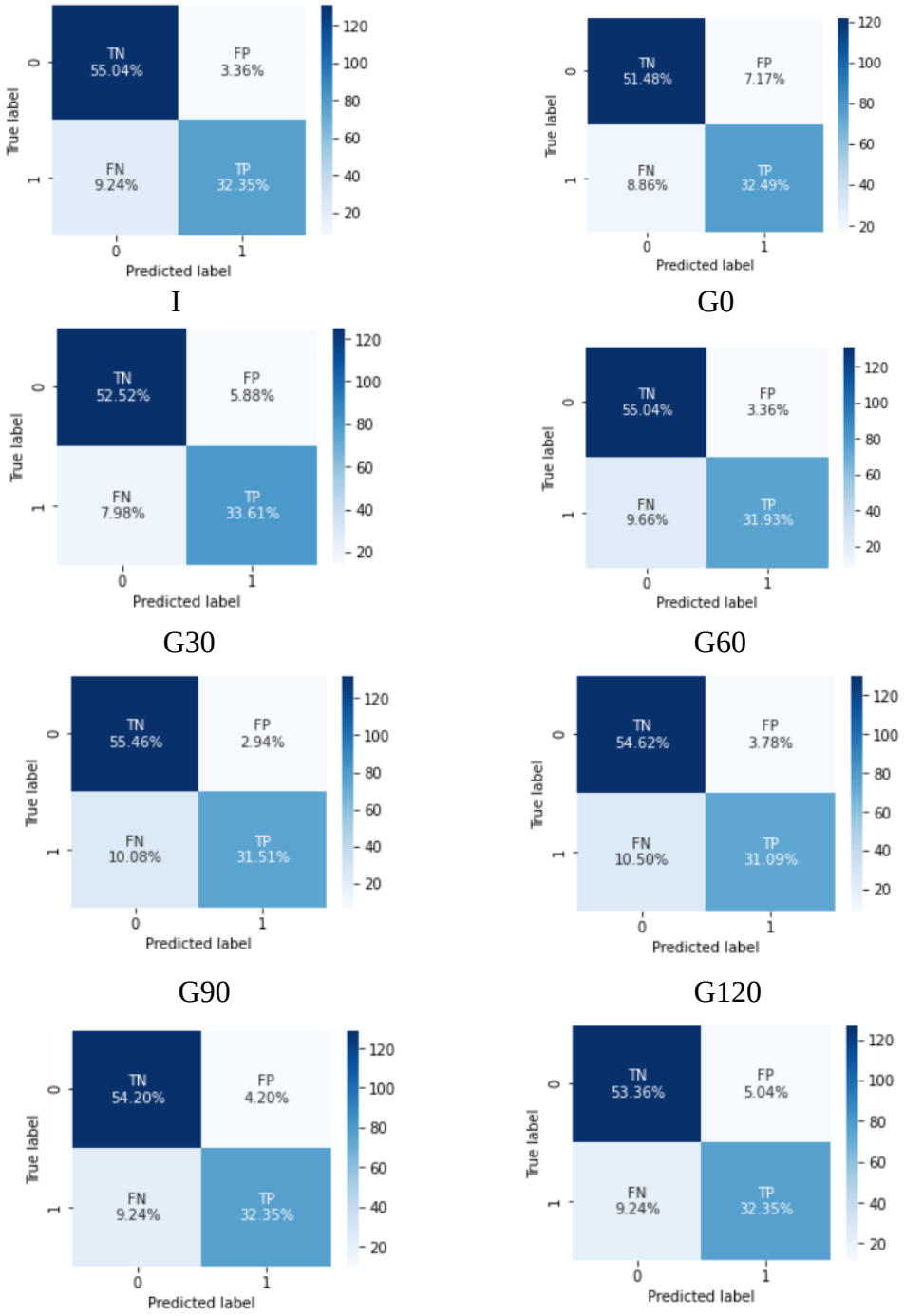
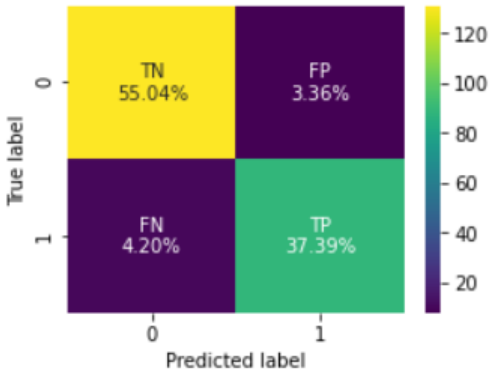
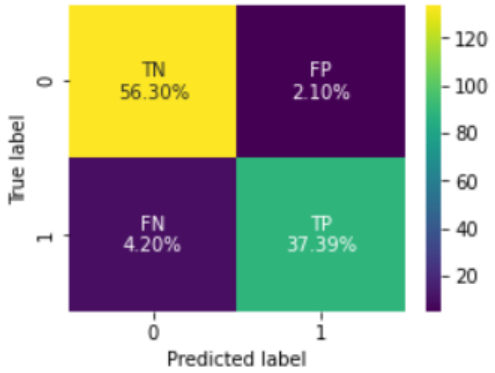


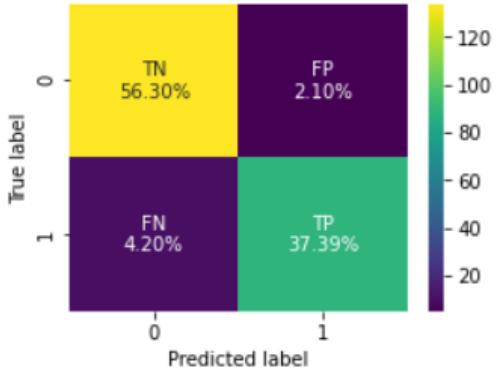
Figure 6. 4: Confusion Matrix Of Single CNN Models.



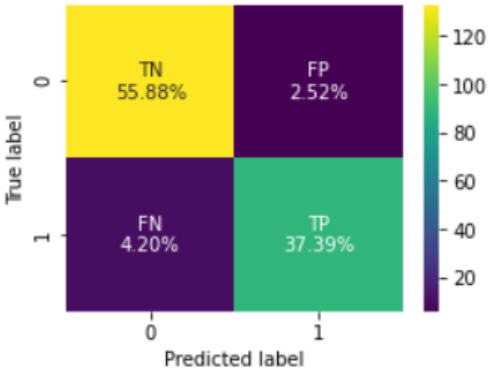
Confusion Matrix Of GR (SR)



Confusion Matrix Of GR (PR)



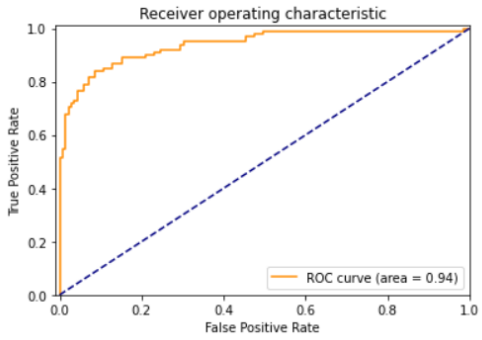
Confusion Matrix Of IGR (SR)



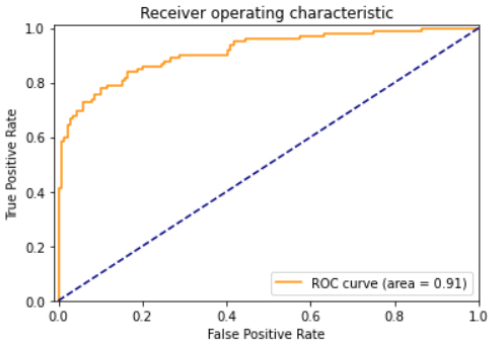
Confusion Matrix Of IGR (PR)

Figure 6. 5: Confusion Matrix of Fused Models.

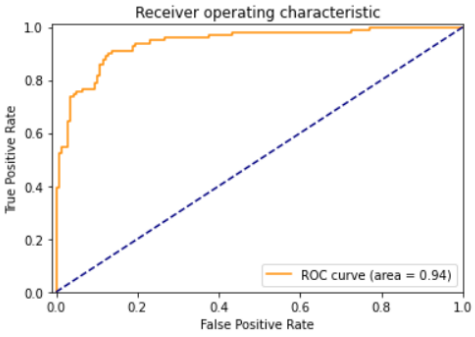
From the Figure 6.4 and 6.5 it is evident that our method IGR (SR) reduces the false positive and false negatives percentage as compared to the I model. Therefore, our method performs better than alternative methods.



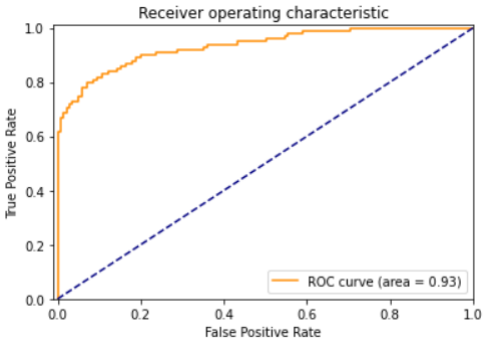
ROC Of Model I



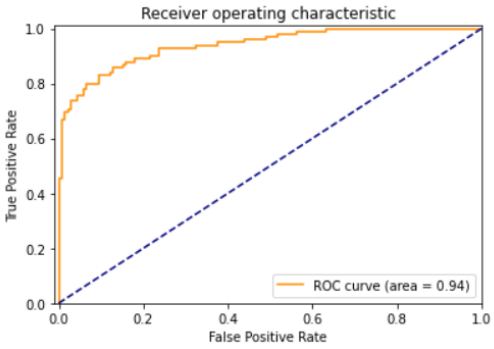
ROC Of Model G0



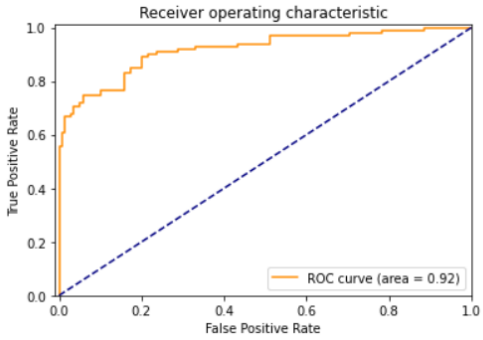
ROC Of Model G30



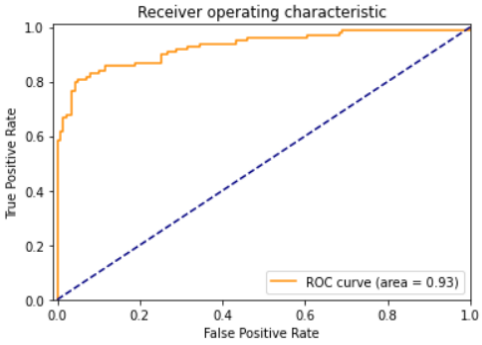
ROC Of Model G60



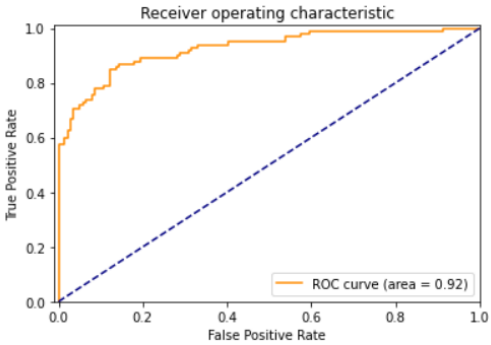
ROC Of Model G90



ROC Of Model G120

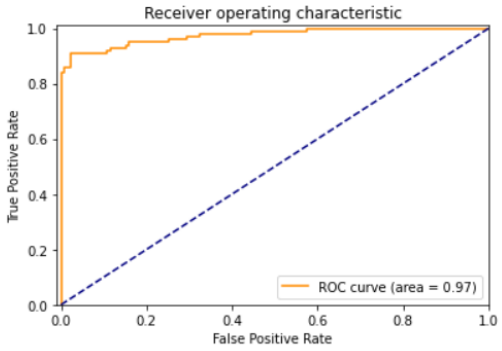


ROC Of Model G150

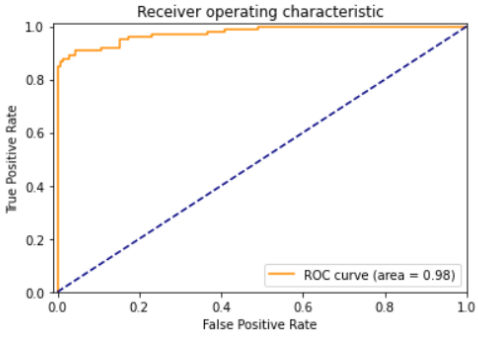


ROC Of Model G180

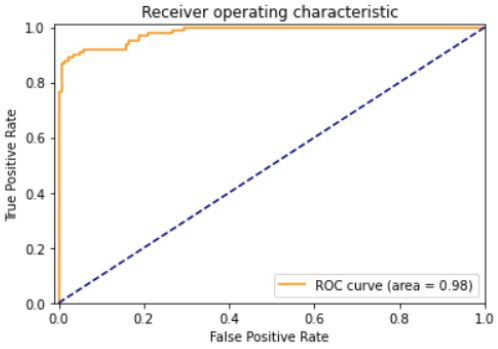
Figure 6. 6: ROC Analysis of Different Single Models.



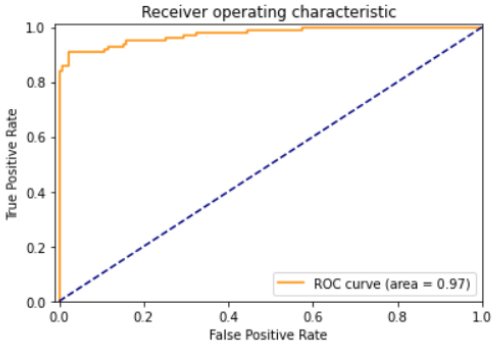
ROC Of Model GR (Sum Rule)



ROC Of model GR (Product Rule



ROC Of Model IGR (Sum Rule)



ROC Of Model IGR (Product Rule)

Figure 6. 7: ROC Analysis Of Different Fused Models.

From the Figure 6.6 and 6.7, it can be concluded that ROC of IGR (Sum Rule) is better than other models. It is evident that Area Under Curve is for IGR (Sum Rule) is more than other models making it more reliable method.

6.7 Performance Evaluation (R0 vs R3)

In this section we have trained our model on 641 color fundus retinal images of Messidor dataset and model performance is calculated using 159 color fundus images from Messidor dataset. In Messidor dataset there are four different classes as:

class0 (R0): NO DR (healthy retina)

Class1 (R1): Mild case of DR

Class2 (R2): Severe case of DR

Class3 (R3): Proliferative case of DR

In this section we have considered only class0 (No DR) and class3 (DR) for detection of DR, making it binary classification. The reason behind taking extreme class0 and class3 is to check the how it effects the performance of our method. The performance of the proposed model is evaluated for model accuracy, Area under ROC (Receiver Operating Characteristics) curve, sensitivity, specificity, f1score, precision, and false positive rate. Table 6.8 shows the performance of our model based on Accuracy metrics. IGR (Sum Rule) is the proposed method throughout the evaluation section.

Table 6. 8: Performance Evaluation Based On Accuracy Metrics.

Class	**CNN for I**	**CNN for G**	**Method**	**Accuracy(%)**	**DF**
R0 vs R3	VGG16	-	I	96.22	-
R0 vs R3	-	VGG16	G0	93.7	-
R0 vs R3	-	VGG16	G30	95.59	-
R0 vs R3	-	VGG16	G60	93.71	-
R0 vs R3	-	VGG16	G90	92.45	-
R0 vs R3	-	VGG16	G120	91.8	-
R0 vs R3	-	VGG16	G150	93.5	-
R0 vs R3	-	VGG16	G180	91.82	-
R0 vs R3	VGG16	VGG16	IGR	96.85	Product rule
R0 vs R3	VGG16	VGG16	IGR	**96.85**	Sum rule
R0 vs R3	-	VGG16	GR	96.85	Product rule
R0 vs R3	-	VGG16	GR	96.85	Sum rule

Accuracy is one of the more obvious metrics, since it measures all the cases correctly identified. If all classes have equal importance, it is most applicable. But sometimes model accuracy can be high while it will perform poorly by negative cases. In those cases, we use f1score, it is the harmonic mean of Precision and Recall, which gives a better indication of inaccurate classifications than Accuracy Metric. From Table 6.9 we can see that f1score of our model is better than accuracy which makes our method more reliable for detection of diabetic retinopathy in retinal images.

Table 6. 9: Performance Evaluation Of Method Based On F1score.

Class	**CNN for I**	**CNN for G**	**Method**	**F1score(%)**	**DF**
R0 vs R3	VGG16	-	I	93.17	-
R0 vs R3	-	VGG16	G0	88.37	-
R0 vs R3	-	VGG16	G30	91.35	-
R0 vs R3	-	VGG16	G60	87.78	-
R0 vs R3	-	VGG16	G90	85.46	-
R0 vs R3	-	VGG16	G120	83.53	-
R0 vs R3	-	VGG16	G150	86.58	-
R0 vs R3	-	VGG16	G180	85.38	-
R0 vs R3	VGG16	VGG16	IGR	93.82	Product rule
R0 vs R3	VGG16	VGG16	IGR	**93.82**	Sum rule
R0 vs R3	-	VGG16	GR	93.82	Product rule
R0 vs R3	-	VGG16	GR	93.82	Sum rule

Similarly, Table 6.10 shows the performance of our model based on recall metrics. Recall is the measure of correctly identified positive cases from actual positive. It is important performance metrics while working with medical images because we do not want to miss the positive case. Table 6.10 indicate that our model gives comparable recall compared to other models.

Table 6. 10: Performance Evaluation Of Our Method Based On Recall Metrics.

Class	**CNN for I**	**CNN for G**	**Method**	**Recall(%)**	**DF**
R0 vs R3	VGG16	-	I	95.34	-
R0 vs R3	-	VGG16	G0	88.37	-
R0 vs R3	-	VGG16	G30	86.04	-
R0 vs R3	-	VGG16	G60	83.72	-
R0 vs R3	-	VGG16	G90	81.39	-
R0 vs R3	-	VGG16	G120	76.74	-
R0 vs R3	-	VGG16	G150	79.06	-
R0 vs R3	-	VGG16	G180	88.37	-
R0 vs R3	VGG16	VGG16	IGR	88.37	Product rule
R0 vs R3	VGG16	VGG16	IGR	**88.37**	Sum rule
R0 vs R3	-	VGG16	GR	88.37	Product rule
R0 vs R3	-	VGG16	GR	88.37	Sum rule

Furthermore, Table 6.11 below shows the performance of our model based on precision metrics. Precision metrics has importance while working on medical images

because we don't want to miss out the positive case. It is evident from the Table 6.11 that our model achieves comparable results with alternate methods.

Table 6. 11: Performance of Our Method Based on Precision Metrics.

Class	**CNN for I**	**CNN for G**	**Method**	**Precision(%)**	**DF**
R0 vs R3	VGG16	-	I	91.11	-
R0 vs R3	-	VGG16	G0	88.37	-
R0 vs R3	-	VGG16	G30	97.36	-
R0 vs R3	-	VGG16	G60	92.30	-
R0 vs R3	-	VGG16	G90	89.97	-
R0 vs R3	-	VGG16	G120	91.66	-
R0 vs R3	-	VGG16	G150	94.44	-
R0 vs R3	-	VGG16	G180	82.60	-
R0 vs R3	VGG16	VGG16	IGR	100	Product rule
R0 vs R3	VGG16	VGG16	IGR	**100**	sum rule
R0 vs R3	-	VGG16	GR	100	Product rule
R0 vs R3	-	VGG16	GR	100	sum rule

Furthermore, we evaluated our model based on specificity metrics which is measure of model for correctly identifying the negative cases. Table 6.12 shows the performance of our model based on specificity (TNR). Our model is able to predict all the negatives cases correctly. Therefore, our method outperforms the other alternate methods.

Table 6. 12: Performance Of Our Method Based On Specificity Metrics.

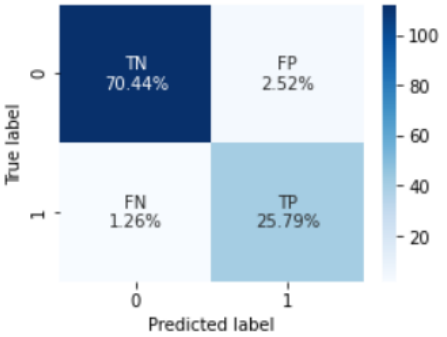
Class	**CNN for I**	**CNN for G**	**Method**	**Specificity(%)**	**DF**
R0 vs R3	VGG16	-	I	96.55	-
R0 vs R3	-	VGG16	G0	95.68	-
R0 vs R3	-	VGG16	G30	99.13	-
R0 vs R3	-	VGG16	G60	97.41	-
R0 vs R3	-	VGG16	G90	96.55	-
R0 vs R3	-	VGG16	G120	97.41	-
R0 vs R3	-	VGG16	G150	98.27	-
R0 vs R3	-	VGG16	G180	93.10	-
R0 vs R3	VGG16	VGG16	IGR	100	Product rule
R0 vs R3	VGG16	VGG16	IGR	**100**	Sum rule
R0 vs R3	-	VGG16	GR	100	Product rule
R0 vs R3	-	VGG16	GR	100	Sum rule

At the end we evaluate our model based on Area Under ROC curve which is plot of recall and False positive rate. It provides the aggregate performance of the model across all the classification being done. Table 6.13 below shows the performance of our model based on Area under ROC curve. It is evident from the table that our method shows comparable results.

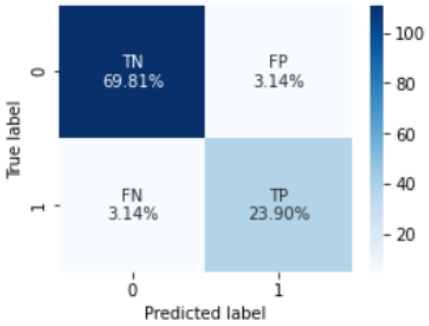
Table 6. 13: Performance Of Our Method Based On AUC.

Class	**CNN for I**	**CNN for G**	**Method**	**AUC(%)**	**DF**
R0 vs R3	VGG16	-	I	98.22	-
R0 vs R3	-	VGG16	G0	98.59	-
R0 vs R3	-	VGG16	G30	98.75	-
R0 vs R3	-	VGG16	G60	98.57	-
R0 vs R3	-	VGG16	G90	94.78	-
R0 vs R3	-	VGG16	G120	95.44	-
R0 vs R3	-	VGG16	G150	96.44	-
R0 vs R3	-	VGG16	G180	95.54	-
R0 vs R3	VGG16	VGG16	IGR	99.63	Product rule
R0 vs R3	VGG16	VGG16	IGR	**99.88**	Sum rule
R0 vs R3	-	VGG16	GR	99.12	Product rule
R0 vs R3	-	VGG16	GR	99.29	Sum rule

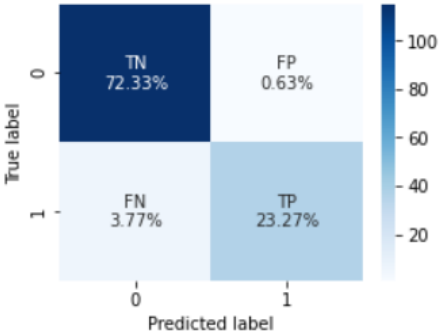
Confusion matrix and ROC for different models are shown in Figure 6.8, Figure 6.9 and Figure 6.10, Figure 6.11 respectively. It is evident from Figure 6.9 below that our method is better than alternative methods and gives the best results.



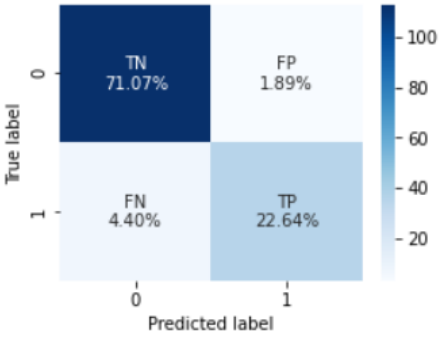
I



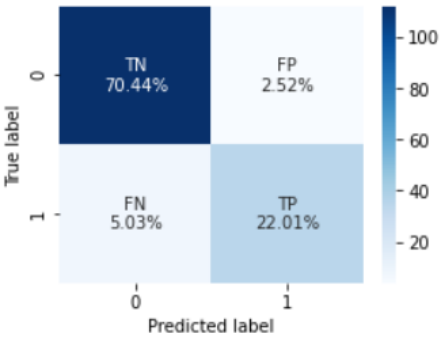
G0



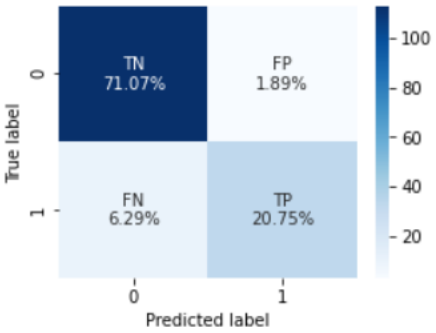
G30



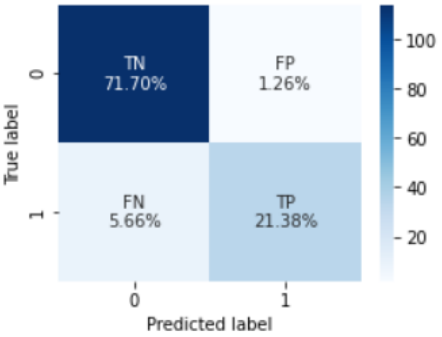
G60



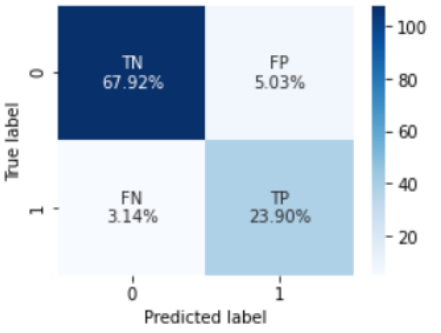
G90



G120

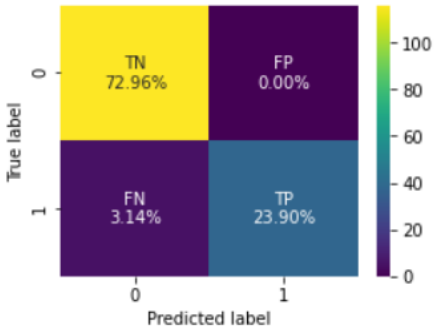


G150

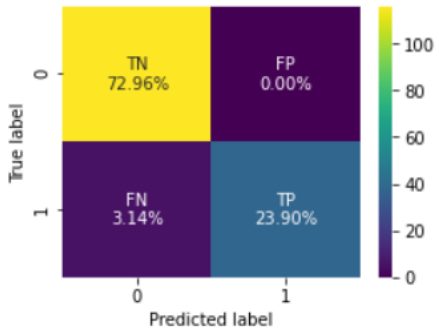


G180

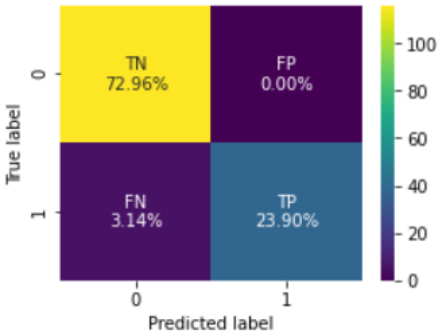
Figure 6. 8: Confusion Matrix Of Different Single Models.



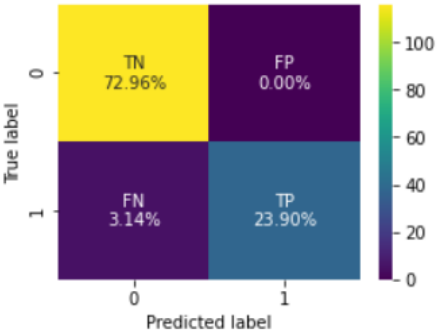
Confusion Of Model GR (SR)



Confusion Of Model GR (PR)



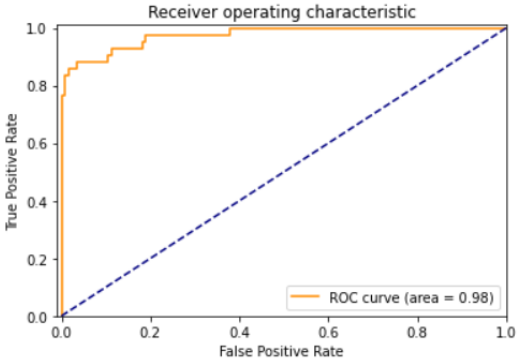
Confusion Of Model IGR (SR)



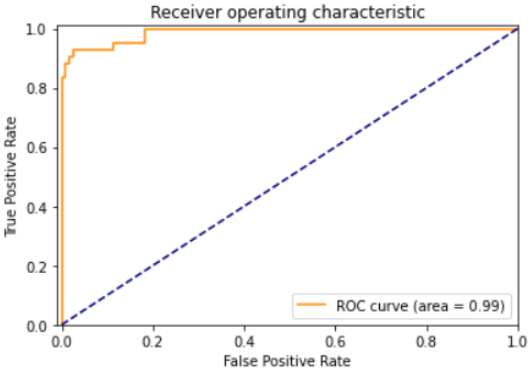
Confusion Of Model IGR (PR)

Figure 6. 9: Confusion Matrix of different Fused Models.

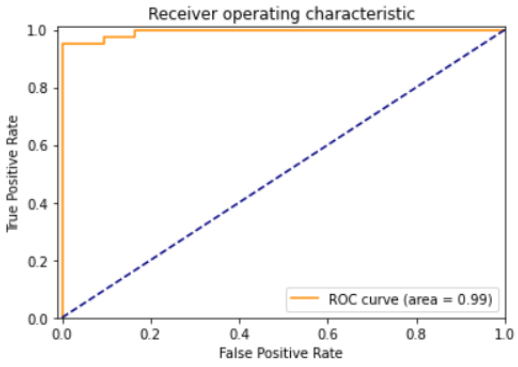
In Figure 6.9, we can see that our method IGR (SR) correctly predicts all the negative cases. Also, precision is 100% as our method is able to predict all the positive correctly. Therefore, our method outperforms the all other methods.



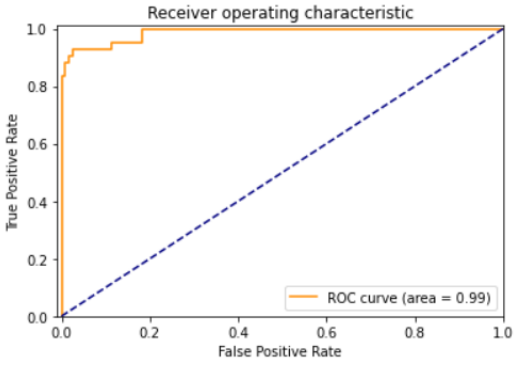
ROC of Model I



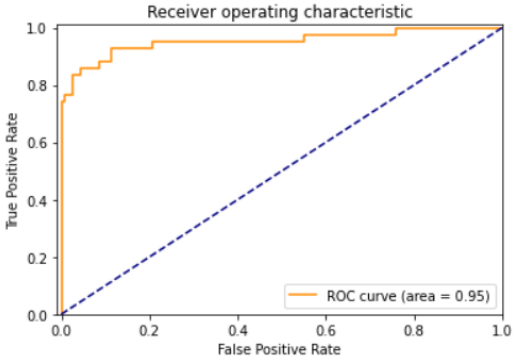
ROC of Model G0



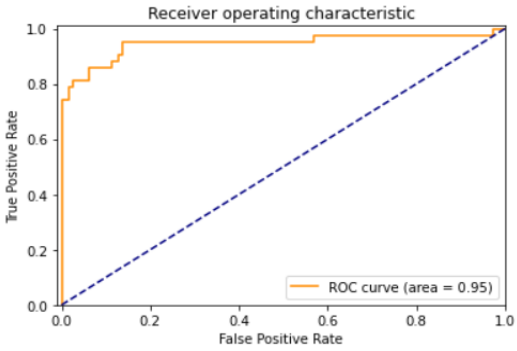
ROC of Model I



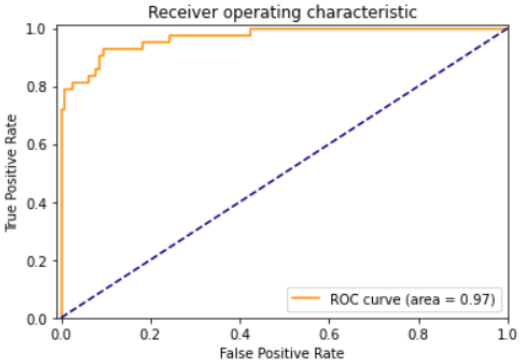
ROC of Model G0



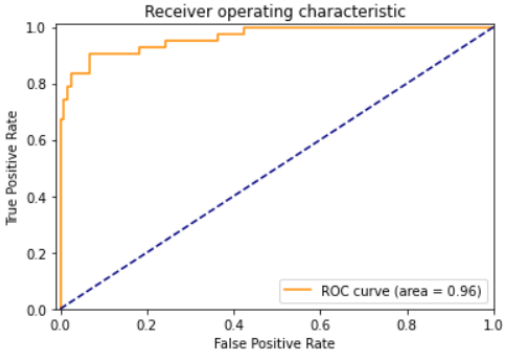
ROC of Model I



ROC of Model G0

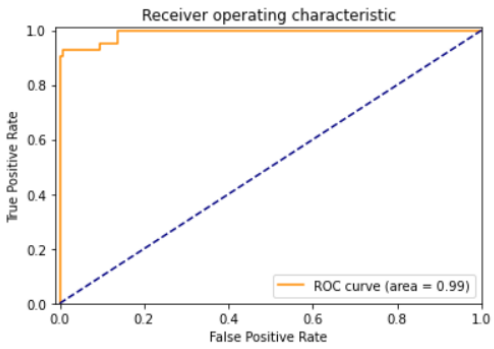


ROC of Model I

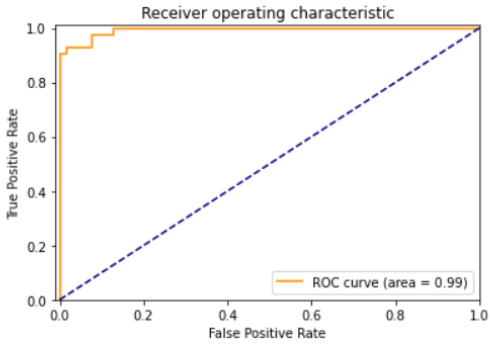


ROC of Model G0

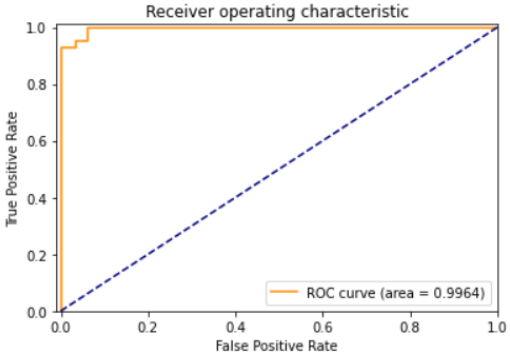
Figure 6. 10: ROC Analysis of Different Single Models.



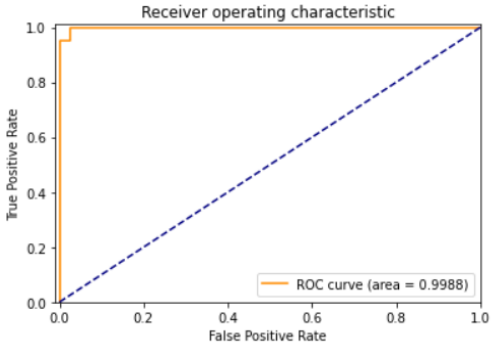
ROC of Model GR (SR)



ROC of Model GR (PR)



ROC of Model IGR (PR)



ROC of Model IGR (SR)

Figure 6. 11: ROC Analysis of Different Fused Model.

From the Figure 6.11, it is evident that our method IGR (SR) gives best result for Area under curve of ROC as compared to the method I in Figure 6.10. Therefore, it can be concluded that our method performs well then alternate methods.

6.8 Comparison of Proposed Method with Literature

In this section, we will compare our generated results with the literature. The comparisons are done based on Accuracy, Recall, Precision, F1score, Train/Test split and Validation type. We are comparing our generated results with the study [48] and [49]. The comparison of our method with literature is shown in the Table 6.14.

Table 6. 14: Comparison of Proposed Method with Alternative Methods and literature.

Method	Accuracy (%)	AUC (%)	Recall (%)	Precision (%)	F1score (%)	Train/ Test (%)	Validation Type
CNN(modified Alexnet) [48]	96.6	-	95.65	90.65	-	60/40	-
CANet [49]	92.6	96.3	92	90.6	91.2	90/10	10-Fold cross validation
Proposed method using R0 vs R3	**96.85**	**99.88**	**100**	**95.86**	**97.89**	**80/20**	**-**
Proposed method using R0R1 vs R2R3	93.69	98.20	96.40	93.00	94.6	80/20	-

In the study [48], they have trained their model on 60% of data and tested on 40% of data. They have used the modified Alex Net architecture to perform the classification task. In their study they were able to classify all the four stages of diabetic retinopathy separately. The used the Messidor dataset to train and test their model.

In the study [49], they have employed 10-Fold cross validation to check the performance of the model. They have done it using CADNet architecture. Their study focused on binary classification of the diabetic retinopathy. They also used the same Messidor database to train and test the model.

From the Table 6.14, it is evident that our method generates comparable results with respect to the methods discussed in the study [48] and [49]. Gabor filter has successfully done its job to enhance the output of the CNN network by taking care of small details in the picture which traditional CNN would miss.

Chapter 7

CONCLUSION AND FUTURE WORK

7.1 Conclusion

In this thesis, we propose to use Gabor filters with CNN for detection of Diabetic Retinopathy from the color fundus retinal images. Our method uses 7 Gabor filters with each Gabor filter separated by the 30^0 of orientation. We used fusion of Gabor based CNN and traditional CNN to detect the diabetic retinopathy. The output of each CNN is given into the decision fusion block which employs sum of probabilities rule to give final decision as to whether retinal image has diabetic retinopathy or not. Our method performance was tested based on Accuracy, Area under curve of ROC, recall, and f1score which are metrics to check the model performance while dealing with medical images. Results obtained from the implemented approach are compared with the literature-available results, and the findings express that the results generated from the proposed method are comparable with those in the literature.

7.2 Future Work

In this thesis, we implemented a method for binary classification of diabetic retinopathy. In the future we will try to implement our method for multi-class classification problem and on a larger dataset. Detecting diabetic retinopathy at early stage could save person from vision loss. So by implementing our method as a multi-classification approach, we will be able to detect it at very early stage.

REFERENCES

[1] Faust, O., Acharya, R., Ng, E. Y. K., Ng, K. H., & Suri, J. S. (2012). Algorithms for the automated detection of diabetic retinopathy using digital fundus images: a review. *Journal of medical systems*, *36*(1), 145-157.

[2] Gomes Filho, E., Pinheiro, P. R., Pinheiro, M. C. D., Nunes, L. C., & Gomes, L. B. G. (2019). Heterogeneous methodology to support the early diagnosis of gestational diabetes. *IEEE Access*, *7*, 67190-67199. doi: 10.1109/ACCESS.2019.2903691

[3] Alyoubi, W. L., Shalash, W. M., & Abulkhair, M. F. (2020). Diabetic retinopathy detection through deep learning techniques: A review. *Informatics in Medicine Unlocked*, *20*, 100377.

[4] Shankar, K., Zhang, Y., Liu, Y., Wu, L., & Chen, C. H. (2020). Hyperparameter tuning deep learning for diabetic retinopathy fundus image classification. *IEEE Access*, *8*, 118164-118173.

[5] Richa, R., Linhares, R., Comunello, E., Von Wangenheim, A., Schnitzler, J. Y., Wassmer, B., ... & Taylor, R. (2014). Fundus image mosaicking for information augmentation in computer-assisted slit-lamp imaging. *IEEE transactions on medical imaging*, *33*(6), 1304-1312.

[6] Shoughy, S.S., Kozak. I. (2016). Selective and complementary use of Optical Coherence tomography and fluorescein angiography in retinal practice.

Eye and Vision **3,** 26. doi.org/10.1186/s40662-016-0058-2.

[7] Oloumi, F., Dhara, A., Rangayyan, R., & Mukhopadhyay, S. (2014). Detection of blood vessels in retinal fundus images. *Computer Science Journal of Moldova*, *65*(2), 155-185.

[8] Rodrigues, P., Guimaraes, P., Santos, T., Simao, S., Miranda, T., Serranho, P., & Bernardes, R. C. (2013). Two-dimensional segmentation of the retinal vascular network from optical coherence tomography. *Journal of biomedical optics*, *18*(12), 126011.

[9] Liu, Z., Opas, C., & Krishnan, S. M. (1997, November). Automatic image analysis of fundus photograph. In *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.'Magnificent Milestones and Emerging Opportunities in Medical Engineering'(Cat. No. 97CH36136)* (Vol. 2, pp. 524-525). IEEE.

[10] Niemeijer, M., Van Ginneken, B., Staal, J., Suttorp-Schulten, M. S., & Abràmoff, M. D. (2005). Automatic detection of red lesions in digital color fundus photographs. *IEEE Transactions on medical imaging*, *24*(5), 584-592.

[11] Freeman, W. R., Bartsch, D. U., Mueller, A. J., Banker, A. S., & Weinreb, R. N. (1998). Simultaneous indocyanine green and fluorescein angiography using a confocal scanning laser ophthalmoscope. *Archives of Ophthalmology*, *116*(4), 455-463.

[12] Hassan, T., Akram, M. U., Hassan, B., Nasim, A., & Bazaz, S. A. (2015, September). Review of OCT and fundus images for detection of Macular Edema. In *2015 IEEE International Conference on Imaging Systems and Techniques (IST)* (pp. 1-4). IEEE.

[13] Salazar, J. J., Ramírez, A. I., De Hoz, R., Salobrar-Garcia, E., Rojas, P., Fernández-Albarral, J. A., ... & Ramírez, J. M. (2018). Anatomy of the human optic nerve: Structure and function. *Optic Nerve*.

[14] Demir, Ö., Doğan, B., BAYEZİT, E. Ç., & YILDIZ, K. (2019). Automatic Detection and Calculation of Drusen Areas in Retinal Fundus Fluorescein Angiography Images. *Marmara Fen Bilimleri Dergisi*, *30*(2), 126-132.

[15] Shanthi, R., & Prabakaran, S. (2017). Fundus Abnormalities and Image Acquisition Techniques–A Survey. *Artic J Chem Pharm Sci*.

[16] Kar, S. S., & Maity, S. P. (2017). Automatic detection of retinal lesions for screening of diabetic retinopathy. *IEEE Transactions on Biomedical Engineering*, *65*(3), 608-618.

[17] Xu, K., Feng, D., & Mi, H. (2017). Deep convolutional neural network-based early automated detection of diabetic retinopathy using fundus image. *Molecules*, *22*(12), 2054.

[18] Esfahani, M. T., Ghaderi, M., & Kafiyeh, R. (2018). Classification of diabetic and

normal fundus images using new deep learning method. *Leonardo Electron. J.Pract. Technol*, *17*, 233-248.

[19] Jiang, H., Yang, K., Gao, M., Zhang, D., Ma, H., & Qian, W. (2019, July). An interpretable ensemble deep learning model for diabetic retinopathy disease classification. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 2045-2048). IEEE.

[20] Zago, G. T., Andreão, R. V., Dorizzi, B., & Salles, E. O. T. (2020). Diabetic retinopathy detection using red lesion localization and convolutional neural networks. *Computers in biology and medicine*, *116*, 103537.

[21] Quellec, G., Charrière, K., Boudi, Y., Cochener, B., & Lamard, M. (2017). Deep image mining for diabetic retinopathy screening. *Medical image analysis*, *39*, 178-193.

[22] Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. *IEEE Access*, *7*, 53040-53065.

[23] Du, X., Cai, Y., Wang, S., & Zhang, L. (2016, November). Overview of deep learning. In *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)* (pp. 159-164). IEEE.

[24] Sharma, P., & Singh, A. (2017, July). Era of deep neural networks: A review. In *2017 8th International Conference on Computing, Communication and Networking*

Technologies (ICCCNT) (pp. 1-5). IEEE.

[25] Nguyen, H. D., Lloyd-Jones, L. R., & McLachlan, G. J. (2016). A universal approximation theorm for mixture-of-experts models. *Neural computation*, *28*(12), 2585-2593.

[26] Chauhan, R., Ghanshala, K. K., & Joshi, R. C. (2018, December). Convolutional neural network (CNN) for image detection and recognition. In *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)* (pp. 278-282). IEEE.

[27] Tabian, I., Fu, H., & Sharif Khodaei, Z. (2019). A convolutional neural network for impact detection and characterization of complex composite structures. *Sensors*, *19*(22), 4933.

[28] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.

[29] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, *15*(1), 1929-1958.

[30] Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.

[31] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[32] Orhan, A. E., & Pitkow, X. (2017). Skip connections eliminate singularities. *arXiv preprint arXiv:1701.09175*.

[33] Hill, P. C. (2007, September). Dennis Gabor-Contributions to Communication Theory & Signal Processing. In *EUROCON 2007-The International Conference on" Computer as a Tool"* (pp. 2632-2637). IEEE.

[34] Turner, M. R. (1986). Texture discrimination by Gabor functions. *Biological cybernetics*, *55*(2), 71-82.

[35] Benoit, A., Caplier, A., Durette, B., & Hérault, J. (2010). Using human visual system modeling for bio-inspired low level image processing. *Computer vision and Image understanding*, *114*(7), 758-773.

[36] Bastiaans, M. J. (1980, August). Sampling theorem for the complex spectrogram, and Gabor's expansion of a signal in Gaussian elementary signals. In *1980 Intl Optical Computing Conf I* (Vol. 231, pp. 274-280). International Society for Optics and Photonics.

[37] Sagiv, C., Sochen, N. A., & Zeevi, Y. Y. (2008). Two-dimensional affine frames for image analysis and synthesis. *Applied and Computational Harmonic Analysis*, *24*(2), 174-194.

[38] Shajihan, N. (2020). Classification of stages of Diabetic Retinopathy using Deep Learning. *Bournemouth University United Kingdom*.

[39] Devaiah, D, Brinda., A, Kumar. (2019). Gabor filters in fingerprint application. *International Journal of Engineering Science and Computing*, 9(11).

[40] Ubul, K., Tursun, G., Aysa, A., Impedovo, D., Pirlo, G., & Yibulayin, T. (2017). Script identification of multi-script documents: a survey. *IEEE Access*, *5*, 6546-6559.

[41] De la Torre Gallart, J. (2019). *Diabetic retinopathy classification and interpretation using deep learning techniques* (Doctoral dissertation, Universitat Rovira i Virgili).

[42] Mushtaq, Z., Yaqub, A., Hassan, A., & Su, S. F. (2019, February). Performance analysis of supervised classifiers using pca based techniques on breast cancer. In *2019 international conference on engineering and emerging technologies (ICEET)* (pp. 1-6). IEEE.

[43] Benmeziane, H. (2020). Comparison of Deep Learning Frameworks and Compilers.

[44] Kumar, A. (2020, September). Deep Learning Explained Simply in Layman Terms. *Data Analytics*.

[45] Sahla, N. (2018). A deep learning prediction model for object classification.

[46] Farokhian, F. (2013). Blood vessels detection and segmentation in retina using Gabor filters. *Eastern Mediterranean University*.

[47] Messidor: Digital color fundus images for diabetic retinopathy. [online]. Available: https://www.adcis.net/en/third-party/messidor/.

[48] Shanthi, T., & Sabeenian, R. S. (2019). Modified Alexnet architecture for classification of diabetic retinopathy images. *Computers & Electrical Engineering*, *76*, 56-64.

[49] Li, X., Hu, X., Yu, L., Zhu, L., Fu, C. W., & Heng, P. A. (2019). CANet: cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading. *IEEE transactions on medical imaging*, *39*(5), 1483-1493.