An-Najah National University

Factory of Graduate Studies

Using Mathematical Modeling and the Linear Discriminant Analysis to Classify Retinal Vessels into Arteries / Veins

By

Alaa Mohammad Bawaqneh

Supervisor

Dr. Hadi Hamad

This Thesis is Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Computational Mathematics, Factory of Graduate Studies, An-Najah National University, Nablus, Palestine.

2018

Using Mathematical Modeling and the Linear Discriminant Analysis to Classify Retinal Vessels into Arteries / Veins

By

Alaa Mohammad Bawaqneh

This Thesis was defended successfully on 23/ 12/2018 and approved by:

Defense Committee Members	**Signature**
– **Dr. Hadi Hamad / Supervisor**	**…….……**
– **Dr. Saed Mallak / External Examiner**	**……..……....**
– **Prof. Naji Qatanani/ Internal Examiner**	**……..……...…….**

Dedication

I dedicate this thesis and give special thanks to my father who always supported me and has been my source of inspiration, to my mother who continually provides her moral, spiritual and emotional support.

I dedicate this thesis to my husband, Yahia, who has been a constant source of support and encouragement, I am truly thankful for having you in my life.

Acknowledgment

First of all, thanks and praises to Allah for blessing me much more I deserve.

I would like to express my deep gratitude to my supervisor Dr. Hadi Hamad for his understanding, guidance, support and patience from the initial to the final stage.

Very special thanks go to my beloved parents, husband and family who supported and encouraged me during stressful moments.

الإقرار

أنا الموقعة أدناه مقدمة الرسالة التي تحمل العنوان:

Using Mathematical Modeling and the Linear Discriminant Analysis to Classify Retinal Vessels into Arteries / Veins

أقر بأن ما اشتملت عليه هذه الرسالة إنما هي نتاج جهدي الخاص، باستثناء ما تمت الإشارة إليه حيثما ورد، وأن هذه الرسالة ككل، أو أي جزء منها لم يُقدم لنيل أية درجة أو لقب علمي أو بحثي لدى أي مؤسسة تعليمية أو بحثية أخرى.

Declaration

The work provided in this thesis, unless otherwise referenced, is the researcher's own work, and has not been submitted elsewhere for any other degree or qualification.

Student's name: **اسم الطالبة: الاء محمد بواقنة**

Signature: **التوقيع:**

Date: **التاريخ: 23/12/2018**

Table of Contents

Dedication III
Acknowledgment IV
Declaration V
Table of Contents VI
List of Tables IX
List of Figures X
List of Appendices XIII
List of Abbreviations XIV
Abstract XV
Introduction 1
Chapter One 8
Background 8
1.1 Eye Anatomy 9
1.2 Retinal Vessels 12
1.3 History of Retinal Imaging and Processing 15
1.4 Fundus Images 17
1.5 Some Retinal Diseases 20
1.5.1 Diabetes 21
1.5.2 Vascular Occlusion 23
1.5.3 Retinopathy Of Prematurity (ROP) 25
1.5.4 Hypertension 26
1.6 Retinal Vessels Segmentation 27
1.7 Retinal Vessels Classification 30
1.8 Related Work 32
Chapter Two 35
Linear Discriminant Analysis (LDA) 35
2.1 Fundamental Definitions of Classification 36

2.1.1 Pattern 37
2.1.2 Classes 38
2.1.3 Classification 39
2.1.4 Supervised / Unsupervised Methods 40
2.2 Linear Discriminant Analysis (LDA) 41
2.2.1 The idea Behind LDA 43
2.2.2 LDA for Two Classes in Two Dimensions 43
2.2.3 LDA for Higher Dimensions 48
2.3 Singularity Problem 52
2.3.1 Pseudo-Inverse 52
2.3.2 LDA Extensions 54
2.4 Principal Component Analysis (PCA) 56
2.5 The Implemented Classifiers 60
2.5.1 Support Vector Machine (SVM) 60
2.5.2 Bayesian Classifier 63
2.5.3 K-Nearest Neighborhood (K-NN) 64
2.5.4 K-Means 66
Chapter Three 69
Implementing LDA in Vessels Classification 69
3.1 Introduction 69
3.2 The DRIVE Database 70
3.3 Methodology 72
3.3.1 Preprocessing 73
3.3.2 Features 81
3.3.3 Feature Extraction (Dimensionality Reduction) 86
3.3.4 Classification 86
3.3.5 Post-processing 88
Chapter Four 91

Results and Discussion 91
4.1 Introduction 91
4.2 Evaluation Metrics 92
4.3 Comparing Results of the Different Approaches 93
4.4 Difficulties Facing the Methodology. 105
4.5 Conclusion and Future Work. 106
References 109
Appendices 120
الملخص ..... ب

List of Tables

Table 1.1: Blood vessels structure 13

Table 1.2: Manifestation of some retinal diseases that affect the blood vessels 20

Table 1.3 : Summary of retinal vessels classification techniques..... 32

Table 3.1: All features tested in the algorithm..... 81

Table 4.1: Ground truth table..... 93

Table 4.2: The average performance for the classifiers after implementing LDA. 98

Table 4.3: Performance after applying PCA at different dimensions with majority voting technique. 99

Table 4.4: The average performance of the classifiers after implementing two-stage method. 100

Table 4.5: Performance for applying the reduction techniques to classify center line pixels. 102

Table 4.6: Performance for the reduction techniques after apply majority rule 104

Table 4.7: Performance comparison between using all features in classification and after reducing the dimension by two-stage method..... 107

Table 4.8: Performance comparison between using LDA and Two-stage method..... 107

List of Figures

Fig. 1.1: Anatomy of human eye 11
Fig. 1.2: Human retinal image 11
Fig. 1.3: Retinal vessels with central light reflex 14
Fig. 1.4: First retinal image taken by Van Trigt in 1853 16
Fig. 1.5: Non-Mydriatic fundus camera 16
Fig. 1.6: Normal human retina. 18
Fig. 1.7: a) Color fundus image, b) RGB image with green filter. 19
Fig. 1.8: (a) Normal retina. (b) Retina with diabetic retinopathy. 22
Fig. 1.9: (a) CVO with heavy bleeding, (b) few bleedings and engorgement of the veins, (c) vitreous bleeding, (d) dot and plot bleeding 24
Fig. 1.10: CAO after few hours of occlusion 25
Fig. 1.11: a) Normal retinal eye, b) retina with plus ROP showing dilation and tortuosity. 26
Fig. 1.12: (a) Normal retina, (b) retina with hypertension. 27
Fig. 1.13: (a) Color retinal image, (b) segmented retinal image 28
Fig. 2.1: A three dimensional data (A and B), reduced into two-dimensional data making it easier to separate into different classes (C) 36
Fig. 2.2: (a) Two dimensional space, (b) three dimensional space. 37
Fig. 2.3: The weighted vector w projects the data: (a) non-conveniently separated, (b) well separated. 44
Fig. 2.4: Scatter variance matrices $S1, S2$ and projected means $m1, m2$ ... 45
Fig. 2.5: Scatter plot for a two classes data in two-dimensional space, illustrating Sw, Sb , $\mu1, \mu2$ and μ. 47
Fig. 2.6: (a) Input data space, (b) feature space. 60
Fig. 2.7: Three different lines to separate data into two classes. 61
Fig. 2.8: SVM classifier. 62
Fig. 2.9: K-NN classifier 65
Fig. 3.1: Classifying retinal vessels block diagram. 70

Fig. 3.2: (a) Colored retinal image, (b) manual vessels segmentation, (c) FOV mask. 71
Fig. 3.3: Artery/Vein ground truth. 72
Fig. 3.4: A block diagram of the implemented retinal vessels classification algorithm. 73
Fig. 3.5: Clear sign of retinopathy in a retinal DRIVE image. 74
Fig. 3.6: (a) RGB image, (b) the neighborhood of the marked pixel in (a) 75
Fig. 3.7: Applying average mean kernel on a pixel. 76
Fig. 3.8: Applying Gaussian kernel on a pixel..... 77
Fig. 3.9: (a) Original image, (b) 7×7 mean filter, (c) 7×7 Gaussian filter with $\sigma = 1$ 77
Fig.3.10: (a) Original image, images after erosion with: (b) *strel('disk',1),* (c) *strel('disk',2),* (d) *strel('disk',3).* 79
Fig. 3.11: (a) Image after erosion with *se= strel('disk',2)* (b) then *bwareaopen* for areas less than 30. 79
Fig. 3.12: (a) The original segment image, (b) the image after the morphological processes. 80
Fig. 3.13: Retinal image (a) red, (b) blue, (c) green, channels 82
Fig. 3.14: Gray scale (intensity) retinal image 83
Fig. 3.15: Apply *"adapthisteq"* for (a) blue, (b) green and (c) red channels 83
Fig. 3.16: Red channel (a) without, (b) with, adaptive histogram equalization. 84
Fig. 3.17: Retinal image in HSV color space. 84
Fig. 3.18: Retinal image: (a) Hue (b) saturation and (c) value channels. ... 85
Fig. 3.19: Upper main vessels after classification. 87
Fig. 3.20: (a) Upper vessels, (b) after removing the nodes. 88
Fig. 3.21: Classification result (a) before, (b) after applying majority rule. 88
Fig. 4.1: (a) The classified image, (b) ground truth image. 92

Fig. 4.2: Sensitivity rate for all testing images using the three classifiers before reduction. 95
Fig. 4.3: Specificity rate for all testing images using the three classifiers before reduction. 95
Fig. 4.4: Accuracy rate for all testing images using the three classifiers before reduction. 96
Fig. 4.5: Sensitivity rate for all testing images using the three classifiers after implement LDA. 97
Fig. 4.6: Specificity rate for all testing images using the three classifiers after implement LDA. 97
Fig. 4.7: Accuracy rate for all testing images using the three classifiers after implement LDA. 98
Fig. 4.8: A comparison between sensitivity rates for the reduction techniques under the majority voting. 101
Fig. 4.9: A comparison between specificity rates for the reduction techniques under the majority voting. 101
Fig. 4.10: A comparison between accuracy rates for the reduction techniques under the majority voting. 102
Fig. 4.11: Result of our algorithm. 103
Fig. 4.12: Sensitivity rate for the reduction technique before and after applying enhancement. 104
Fig. 4.13: Specificity rate for the reduction technique before and after applying enhancement. 104
Fig. 4.14: Accuracy rate for the reduction technique before and after applying enhancement. 105

List of Appendices

Appendix (A): Algebraic Background 120
Appendix (B): Statistical Background 124
Appendix (C): Classifiers 129

List of Abbreviations

OD	Optic Disc
AVR	Artery to Vein Ratio
ROP	Retinopathy of Prematurity
NPDR	Non-Proliferative Diabetic Retinopathy
PDR	Proliferative Diabetic Retinopathy
CVO	Central Vein Occlusion
BVO	Branched Vein Occlusion
CAO	Central Artery Occlusion
SVM	Support Vector Machine
RGB	Red, Green and Blue color spaces
ROI	Region of Interest
MLP	Multilayer Perceptron
K-NN	K-Nearest Neighborhood
DRIVE	Digital Retinal Images for Vessel Extraction
LDA	Linear Discernment Analysis
PCA	Principle Component Analysis
GMM-EM	Gaussian Mixture Model using Expectation Maximization
QDA	Quadratic Discriminant Analysis
NN	Neural Network
RBF	Radial Basis Function
ROC	Receiver Operator Characteristic
CAD	Computer Aided Design
CRL	Coefficient of Racial Likeness
FA	Factor Analysis
RLDA	Regularized Linear Discernment Analysis
MLDA	Modified Linear Discernment Analysis
PLDA	Penalized Linear Discernment Analysis
OLDA	Orthogonal Linear Discernment Analysis
ULDA	Uncorrelated Linear Discernment Analysis
FOV	Field of View
GT	Ground Truth
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative

Using Mathematical Modeling and the Linear Discriminant Analysis to Classify Retinal Vessels into Arteries / Veins

By
Alaa Mohammad Bawaqneh
Supervisor
Dr. Hadi Hamad

Abstract

Retinal vessels classification into artery / vein is an important issue; since it helps in early detection of serious retinal diseases, this can avoid blindness. Many features can be extracted from fundus images to distinguish between vessels, but some of these features are redundant, disturbing and time consuming, hence leading to inefficient classification results.

In this work, methods for dimensionality reduction are discussed: Linear Discernment Analysis (LDA) and Principal Component Analysis (PCA). We also modified and implemented a two stage method in which PCA is used first to project data into a lower dimension, then LDA is implemented on the resulted projected data to obtain a one dimensional data. Through building a program using MATLAB, all these steps and other image processing are executed to classify vessels. Images from DRIVE database are implemented in some preprocessing steps and features are read from the vessels, then the dimension is reduced using dimensionality reduction technique. Finally, major voting for three classifiers (K-NN, SVM, and Naïve Bayes rule) are used to classify projected data.

In the final stage of this work, results are measured and evaluated through sensitivity, specificity and accuracy metrics. Results of the different approaches showed that both LDA and the two stage method are able to

reduce the feature space into one dimensional data space and the process time. Although both techniques enhance the classification but Two-stage method showed slightly better results. The proposed two stage method classified the vessels with average accuracy 86.4% on the 14 testing images.

Introduction

Retina is a very important part of the human body which might be affected by diseases, early detection of the retinal diseases avoids the risk of blindness. Classifying retinal vessels into Artery/Vein helps ophthalmologists to notice early signs of some retinal diseases which affect the vessels itself, one of the symptom is artery becomes narrower in case of hypertension, so the artery/vein ratio (AVR) decreases which is considered as an early sing for this disease [52, 53].

Over the years, researchers worked in retinal vessels classification, they used different features and different classifiers in order to decrease the misclassification error. In the following paragraph a set of works relevant to this issue.

C. Muramatsu et al. (2010) considered a method for automatic measurement of the artery to vein diameter ratio (AVR) in order to evaluate the hypertension in retinal fundus images. They first apply two methods to segment images which are black top-hat transformation and double ring filter, they developed a new method by combining these two methods simply by taking the average of output, they compared results and find that the developed method gave best results. Then RGB color information is extracted from major vessels which are located in Region of Interest (ROI) which is quarter-disc to one disc diameter. Then the centerline pixels are classified using LDA. Finally, each sub-segment is assign to be artery or vein using majority rule. For each pair of vessels that runs parallel to each

other, AVR was calculated. Using test images from DRIVE dataset, 88.2% of centerline pixels were correctly assigned as artery or vein [46].

Maheswari et al. (2015) proposed an automated method to classify retinal vessels into artery or vein, the images are obtained from available database like DRIVE database. Green or red channels are used with normalized intensity since vessels in these channels are visible better. Their method consists of four main steps. First step, is to segment the image using morphological processes. Second, they extracted the center line using thinning algorithm. Third step, is to extract features from center line pixels, these features are: a) red, green and blue intensities in addition to their means and standard deviation, b) saturation, hue and intensity in addition to their standard deviation. Finally they implement these features in different classifiers: Linear Discriminant Analysis (LDA), Quadratic discriminant analysis (QDA) and k-nearest neighbor (k-NN). [40].

Divya et al. (2014) believed that using only intensity features is not sufficient for vessels classification so they used the fact that veins rarely cross veins and arteries rarely cross arteries in their approach. Their method consists of three main phases: First phase, generating the graph in which they segmented the vessels and then extracted vessels center line. After that, the graph is extracted and some modifications are applied. Second phase, analyzing the graph, four types of intersection points are defined and then node classification algorithm are used in order to decide the node type. In final phase, commonly used classifiers (LDA, QDA and K-NN) are implemented to classify vessels. They noticed that using graph-based

method with LDA outperform the accuracy of LDA classifier with intensity features [9].

Another work for Divya with a different team (2015) performed an automatic method for artery/vein classification by incorporating graph based algorithm. First they segmented the retinal vessels and then extracted the graph to analyze it by determining intersection points and vessels segments. After that a features vector was extracted and finally they implemented the neural network method for pattern recognition classification. They compared the result of this work with Divya's previous work and found that accuracy performance is improved from 93% by K-NN to 95% by neural network [10].

Kondermann et al. (2007) compared between two methods to build features vector. First they applied three preprocessing steps: image enhancement, extract the skeleton and segment the vessels tree. Then applied two approaches to build features using either Region of Interest (ROI) in which they defines a quadratic region around skeleton pixel, or vessel profile in which they read the vessels profile to a vector along a line crossing the skeleton vessel. After that, they reduced the dimension of feature vector using combined multiclass PCA; finally two classification methods are applied which are Support Vector Machines (SVM) and Neural Network (NN). This methodology was tested on four 1024×1280 retinal images containing 10,132 centerline pixels. For SVM four kernels were used which are: linear, polynomial, radial basis function (RBF) and sigmoidal. They found that using RBF kernel gave the best results. Finally, they use

the Knowledge that all pixels between any two crossing points belong to same class which improved their final result by 6.44 % . Using ROI to extract feature and then applying NN gave the best classification results (95.32 % of major vessels were correctly assigned to be artery or vein) [33, 42].

Niemeijer et al. (2009) proposed a supervised, automatic method for classifying retinal vessels. They obtained the segmented images from DRIVE database. First thinning algorithm is used for detecting center line pixels, then the intersection points are removed. After that features are extracted from each center line pixels and a soft label is assigned to each centerline indicating it's being a vein pixel, then the dimension is reduced by selecting the best features using sequential forward floating in which we start in empty set of features and then we add or remove feature to enhance the performance of classifier. Some of these features are: intensity, normalized vessels' contrast, normalized average hue, saturation and intensity. 12 from 24 feature were selected. Finally classifiers such as LDA, PCA, K-NN and SVM are tested using the fact that connected centerline pixels in vessel segment are belonging to the same type. They found that K-NN provided best performance by evaluating the results using ROC (Receiver Operator Characteristic) where the area under ROC curve was 0.88 [49].

Classify retinal vessels still an important issue, since small error in classification may lead to misdiagnosis, so the researchers worked heavily in this domain to get more and more accurate results.

Many challenges face this classification, particularly the inhomogeneous lighting and low contrast in retinal images. Also vessels themselves disturb the classification since different vessels in different regions in the same or images might be identical. Moreover, choosing good features is considered as a basic challenge that faces the classification.

Many features can be used for this classification but some of them are redundant and disturb the classification results. To keep the most effective features, either features selection or feature extraction (dimensionality reduction) methods are used; in this work different dimensionality reduction techniques are implemented which are Linear Discriminant Analysis (LDA) and Principle Component Analysis (PCA). Moreover we modified the Two-stage method (PCA+LDA) as a new dimensionality reduction technique for retinal vessels classification.

This thesis is organized as follows:

Chapter 1 presents a medical background and an introduction to retinal vessels segmentation and classification. Moreover it includes an overview of the related works.

Chapter 2 provides a description for the used dimensionality reduction techniques which are: LDA, PCA and Two-stage method, in addition to the following classifiers: Support Vector Machine (SVM), Bayes rule, K-means and K-Nearest Neighborhood (K-NN).

Chapter 3 covers our methodology for retinal vessels classification with detailed steps, MATLAP software was used and a program was built to read different features form the training images, then dimensionality

reduction techniques are implemented in order to project the data into lower dimensional space. In the next step, different classifiers are executed and evaluated.

Chapter 4 reports comparisons between different dimensionality reduction techniques and the conclusion of this work, in addition to discuss the main contributions of this research work.

Chapter One

Background

Chapter One

Background

The eye is a very important and sensitive part of the human's body; hence studying diseases that affect this part is an important issue. As some diseases lead to the risk of blindness, an early detection of diseases will help in treatment. Ophthalmologists may not be able to observe slight changes in the retina; consequently, it's very hard to detect the disease in the early stage; although researchers suggest methods for this purpose. Fundus images play a vital role in studying retina since it gives non-invasive visualization to obviously see the retina and its vessels, it gives an acceptable amount of information that ophthalmologists may need.

One can say that there is a relationship between eye diseases and changes appearing in vessel's structure; this makes studying retinal vessels a very important task. Many researches interested in retina, some of them study vessels segmentation: manual, semi-automatic and automatic methods are proposed over the years. Also, morphological features of vessels are evaluated; one of the most important parameters that can be measured from studyimg vessels structure is Artery-Vein ratio (AVR).

AVR is found to be a very useful indicator for early detection of retinal diseases. Hence, to achieve accurate diagnosis, an accurate measurement of AVR is required. Classifying vessels into an artery or vein is an important step for AVR measurement which is an essential, accurate and critical step that faces many challenges making it a difficult task. Classifying vessels

into an artery or vein is a critical step since a slight error in classification makes large influence in final AVR value which in turn leads to misdiagnosing of eye disease.

1.1 Eye Anatomy

Human eye consists of three major layers: external, intermediate and internal layer; the first layer consists of sclera (white part of the eye) and cornea. Cornea is a transparent layer that has two functions: protects the eyeball and refracts the light, it covers both iris (colored circle) and pupil (dark circle). Pupil's color is dark due to absorbing most of light by pigment epithelium. The second layer is divided into two parts: anterior part contains iris and ciliary body, and posterior part which is choroid. The last layer is retina, see Fig. 1.1. The space between the lens and retina filled by colorless, transparent mass called vitreous.

Iris is a membrane that controls the size of the pupil diameter and hence the amount of light that reaches retina. In bright light, iris reduces the pupil size, hence decreasing the amount of light entering the eye, while in dim-light it enlarges the pupil size, which increases the amount of light entering the eye [32, 37, 2, 65].

Retina is a sensitive tissue layer located at the back of the eyeball and it's only 1 mm in thickness and in which microscopic blood vessels can be seen distinctly and evaluated. This important part of the eye is served by a second blood supply which is the choroid. Retina plays an important role in visualization, and it has two kinds of vessels which are Arteries and Veins.

In retinal center there is a circular to oval white area with about 1.5×2 mm called optic disc (optic nerve), which is the place where retinal vessels originate and converge, it connects the retina to the brain. From the center of the optic disc radiate the major blood vessels of the retina. Also, there is a small extra sensitive area located in the center of retina called macula, which is responsible for central vision and it has high light sensitive cells that helps to see small objects clearly. In the middle of macula there is a circular indentation of the retina of approximately $1.5\ mm$ in diameter and it has no blood vessels which is called fovea; and it is responsible for sharp vision, see Fig. 1.2 [32, 54].

There are two sources of blood supply to the retina: central retinal artery and the choroidal vessels. The choroidal vessels supply the outer retinal layers since they receive the greatest blood flow (65-85 %) while the remaining (20-30 %) flows to retina through the central retinal artery from the optic nerve to supply the inner retinal layers [48].

Retina contains photoreceptors that transform captured light ray into neural signals. There are two types of photoreceptors in retina: *Cones* which are primary located in the macula and they are high sensitive to color vision and they also deal with white and black vision. The other type is *Rods* which are sensitive to low level of illumination so they are useful in dim light and they receive no color information [37, 64].

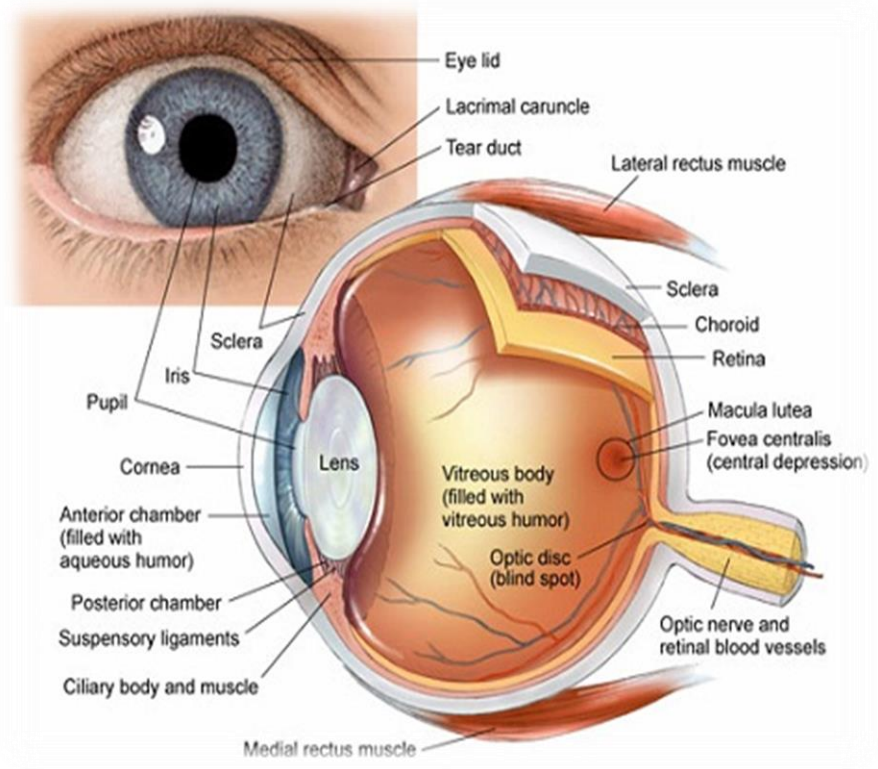


Fig. 1.1: Anatomy of human eye

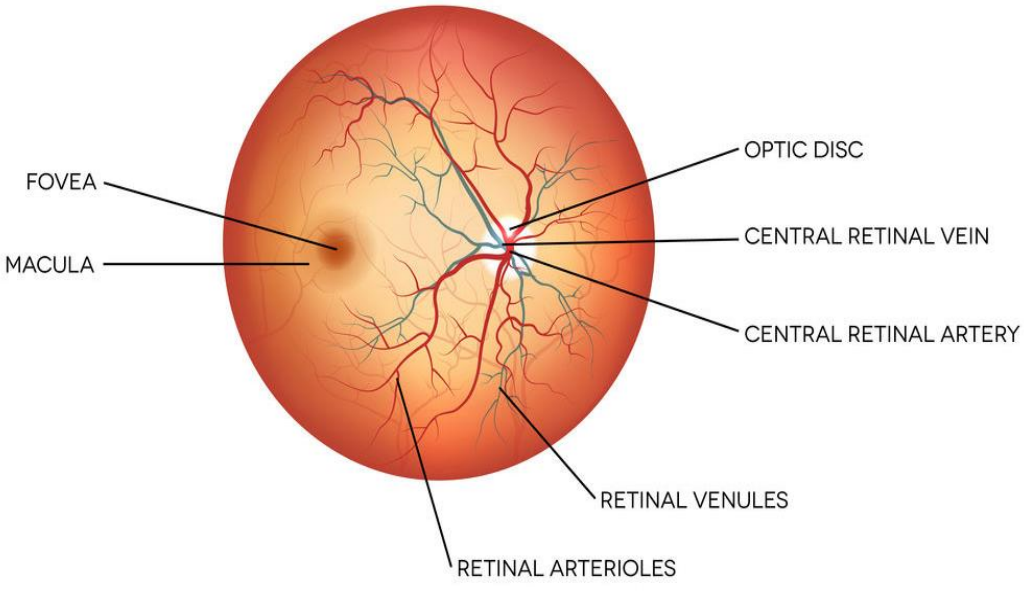


Fig. 1.2: Human retinal image

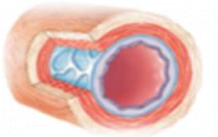
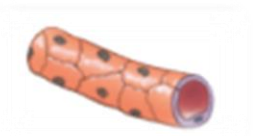
1.2 Retinal Vessels

There are five types of vessels in vivo: *Arteries* which are very strong vessels that carrying blood with oxygen and nutrients away from heart under high pressure. They are subdivided into thinner tubes that give rise to branched, finer vessels which are called *Arterioles*. Smaller-dimeter blood vessels are called *Capillaries* which connect smallest arterioles to smallest venules; capillary walls allow the diffusion of blood with high levels of oxygen and nutrients and allow high levels of carbon dioxide and other wastes to move from tissue to capillaries. *Venules* are microscopic vessels that link capillaries to another type of vessels called veins. *Veins* carry blood with carbon dioxide and other wastes again to the heart, see Table 1.1 [44, 60].

Retinal blood vessels are the only part of human circulation that can be visualized non-invasively. Arteries have a wall that consists of three distinct layers, also veins wall are similar but they have thinner walls that are less elastic than artery walls.

Blood pressure is strongest when the blood leaves the heart, so it is highest in arteries, less in the arterioles, and lowest in the capillaries. Unlikely the arteries, veins don’t have sufficient pressure to keep blood moving so veins have valves which open if blood flow toward the heart and close if it reverses [44].

Table 1.1: Blood vessels structure

	Mean diameter	Mean wall thickness	Shape
Artery	4.0 mm	1.0 mm	
Arteriole	30.0 μm	6.0 μm	
Capillary	8.0 μm	0.5 μm	
Venule	20.0 μm	1.0 μm	
Vein	5.0 mm	0.5 mm	

*** (mm: millimeters, μm: micrometer)**

In retina, we need to classify vessels into their two major types: artery or vein, this step is essential for detecting some retinal diseases. There are differences between arteries and veins that will be helpful in distinguishing the vessels, some of the most important ones are:

1- Veins are darker and more reddish whereas arteries are brighter than veins.

2- Arteries are thinner than the adjacent veins.

3- Near the optic disc veins and arteries usually alternate to each other before branching out. Far from the optic disc one can find a vein next to two adjacent arteries.

4- At least near the optic disc an intersection between two arteries or two veins is rarely, but both can bifurcate into smaller vessels.

5- Some retinal blood vessels include a light streak, which clearly appears in high resolution images known as center light reflex that is found along blood vessel's center. For Arteries the central reflex is more distinctive than veins, it is more common in younger retinas. It appears as a result of light reflection from the surface of blood wall. See Fig. 1.3.

Without using structural information which is mentioned above, it is almost impossible to classify small vessels correctly [2, 55, 35, 59].

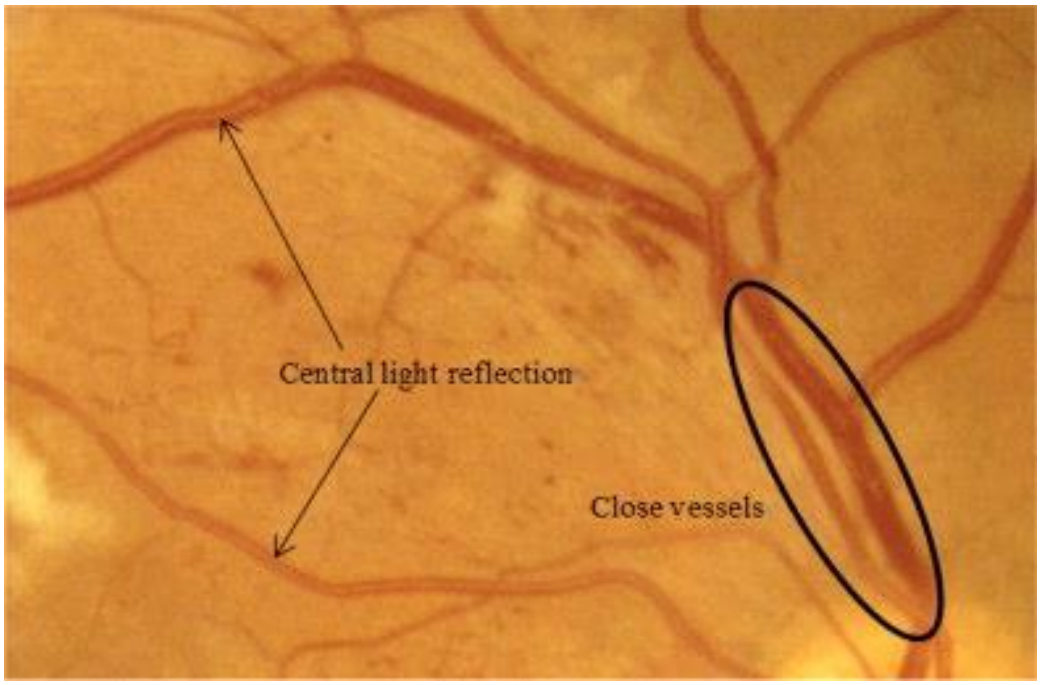


Fig. 1.3: Retinal vessels shows with center light reflex

1.3 History of Retinal Imaging and Processing

Since many important eye diseases appear in retina, researchers were interested in obtaining retinal images and performing image analysis which help in studying these diseases. First attempt to image retina was by the French physician Jean Mery, who showed that: if you immersed a live cat in water, its retinal vessels will be visible from outside. But this experiment is impractical for human, so in 1823 Jan Evangelista Purkyně invented the principle of ophthalmoscope, and it was first developed in 1845 by Charles Babbage .Then again developed and reported in 1851 by Von Helmholtz. The first image of retina was published in 1853 by Van Trigt, see Fig. 1.4 [1].

Ophthalmoscope is considered as a direct method to obtain images for the retina but it's often inaccurate due to non-standardized illumination, low magnification and subject bias. In addition, ophthalmoscope requires the physician to come close to the patient face and infectious disease was prevalent at that time, hence it was necessary to image the eye photographically. The first useful photographic retinal image was obtained in 1891 by Gerloff. In 1910, Gullstrand developed the fundus camera which is still a concept used to image the retina today; it is a complex optical system used for imaging the retina of the eye and it is capable of illuminating and imaging the retina simultaneously [37, 1]. Nowadays, it is easier to capture retinal digital images due to the development of digital fundus cameras. See Fig. 1.5.

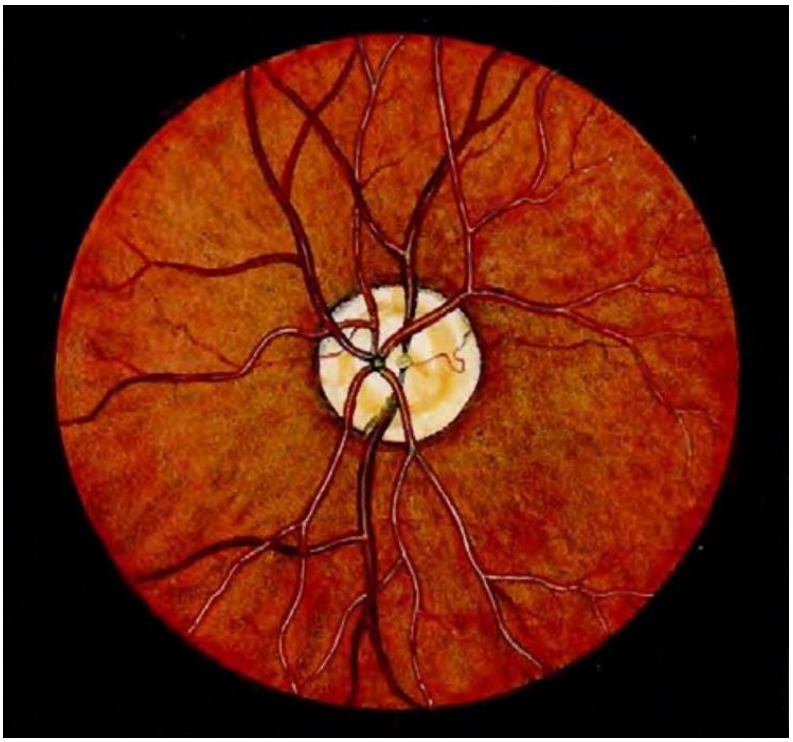


Fig. 1.4: First retinal image taken by Van Trigt in 1853

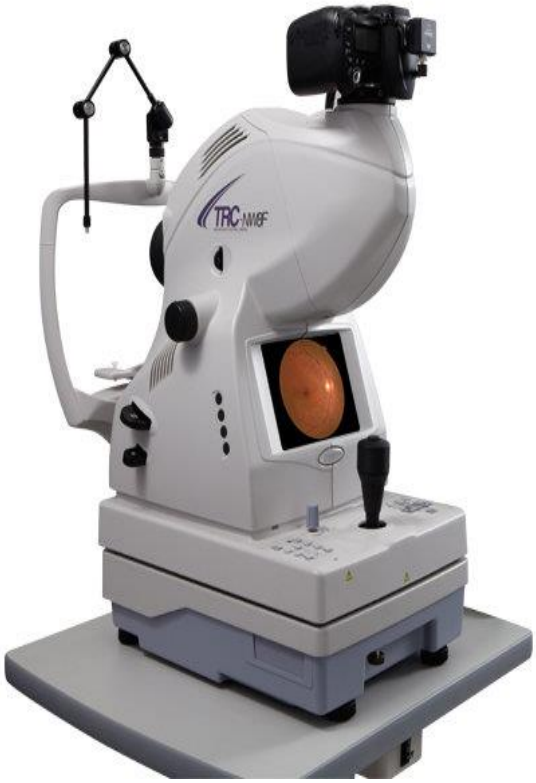


Fig. 1.5: Non-Mydriatic fundus camera

Ophthalmology is a specialized field of medicine, where we can directly visualize the pathology of the eye and treat it accordingly.

In 1973, Matsui et al. published a method for retinal image analysis that focused primary on vessels segmentation. Their approach was based on mathematical morphology and they used digitized slides of fluorescein angiograms of the retina. In 1984, Baudoin et al. described an image analysis method for detecting microaneurysms, a characteristic lesion of diabetic retinopathy, it was the first method to detect and segment abnormal structures. In 1990s the field of retinal image processing dramatically changed as a result of the development of digital retinal imaging and the expansion of digital filter-based image analysis techniques [1].

Fundus imaging helps in documenting the disease process which further helps in classifying, treating and following up the diseases effectively.

1.4 Fundus Images

Retinal photography or funduscopy or fundus imaging is a technique for 2D projection of 3D retinal tissues onto an imaging plane by using reflected light rays. The amount of reflected rays appears as intensity values on the image plane. Fundus image shows retina, optic nerve and fovea, also one can see the retinal blood vessels in fundus image with no need for surgery. The fundus imaging helps in diagnosing various ocular and systemic disorders like diabetic retinopathy and hypertensive retinopathy. The fundus imaging techniques are classified into several

categories depending on the type of image, emitted light and acquisition procedure [1, 2, 65].

Retinal fundus images play an important role in many applications, for example, they help in diagnosing several eye diseases even it helps in early detection, also it is used in human recognition. Retinal fundus images are used in many popular methods to detect vessels structure both manually and automatically. Vessels structure such as thickness, curvature and artery-vein ratio (AVR), are used for early diagnosis. A normal fundus human eye is reddish and clean as it's shown in Fig. 1.6.

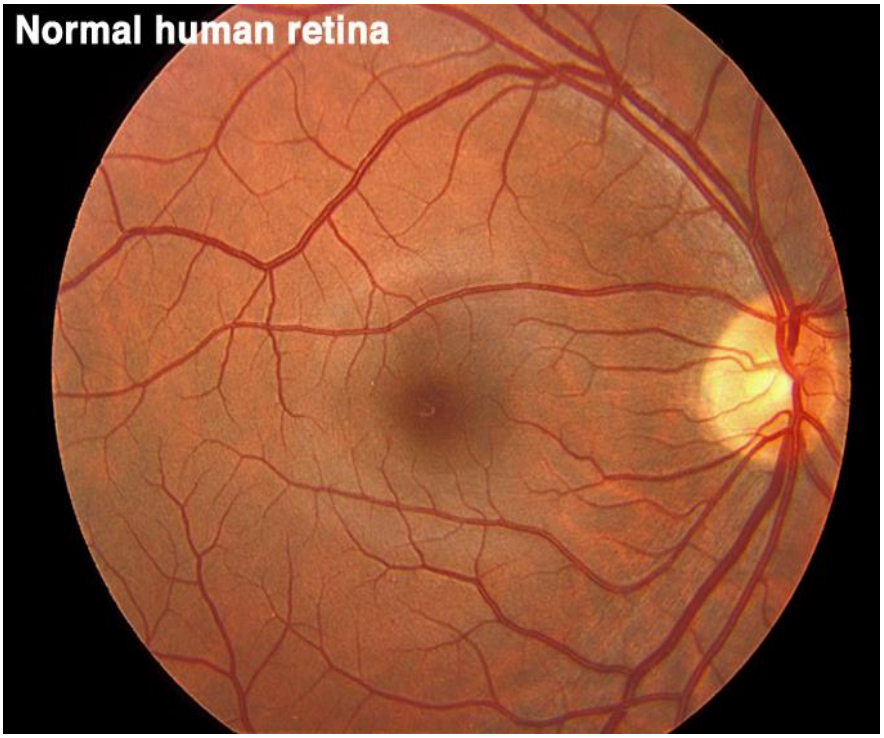
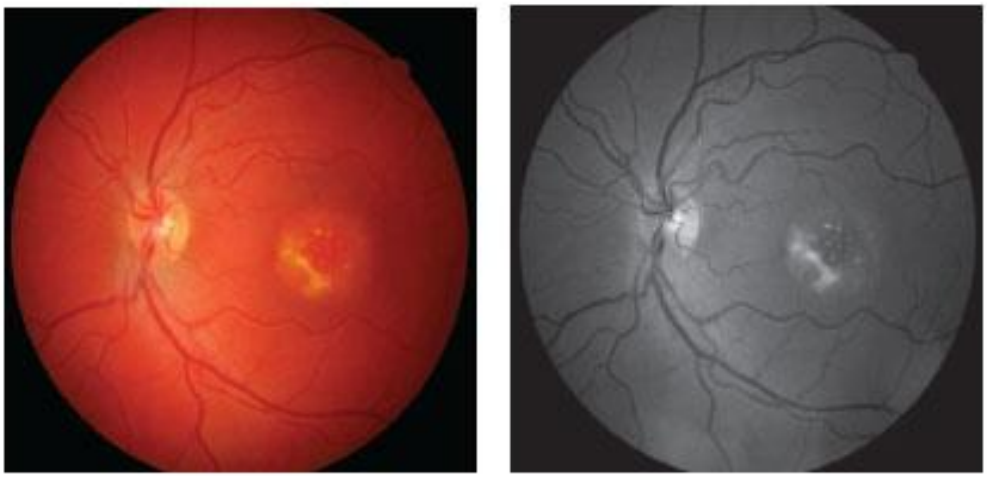


Fig. 1.6: Normal human retina.

Fig. 1.6 shows optic disc which appears as bright region in the image where all blood vessels are converged, while the dark region in the image is called macula. The aforementioned properties for retinal vessels can be seen in

the image, for example artery looks brighter than vein, center light reflex is clearer in artery etc.

Basically there are two main categories of fundus images, we will concentrate our work on digital fundus photography, as its easier to obtain for the physician and the patient which is taken in a short time [5, 42]. See Fig. 1.7.



(a) (b)

Fig. 1.7: a) Color fundus image, b) RGB image with green filter.

Retinal fundus images are considered as visual records that document opthalmoscopic appearance of patient's retina. They also allow the physician to study patient's retina in order to help them to identify retinal changes on follow-up. Due to its safety and cost effective, fundus imaging remains the basic method of retinal imaging [51].

Retinal images are captured with different field of view and the resolution can vary from 768x576 pixels to 2948x1536 pixels for different cameras, high resolution are recommended but it requires large storage capabilities.

In general, image quality can be affected by a number of factors including movement of eye, poorly dilated pupils and/or small pupils [37, 64].

1.5 Some Retinal Diseases

Normal retina, as seen in Fig. 1.6, has clean space between vessels and the background, which as seen the majority of the retinal fundus image is reddish, clean and does not have any hemorrhages, dots, blots or occlusion in vessels. The retinopathy or diseases that affect the retina such as diabetic, glaucoma, stroke and other cardiovascular diseases in adult, and also retinopathy of prematurity (ROP) in infants, affects the retina, see Table 1.2. However, ophthalmologists don't have time to observe abnormality signs in fundus images, so automatic analyzing of fundus images is essential.

Table 1.2: Manifestation of some retinal diseases that affect the blood vessels [48]

Disease	Manifestation							
	NVE	NVD	Artery Color	Vein Color	Artery Narrowing	Artery Dilation	Vein Narrowing	Vein Dilation
Choroidal Neovascularization	✖							
Diabetic Retinopathy	✖	✖	✖		✖		✖	✖
Hypertensive Retinopathy				✖	✖			
Branch Retinal Vein Occlusion		✖		✖	✖	✖		✖
Central Retinal Vein Occlusion		✖		✖	✖	✖		✖
Hemi-Central Retinal Vein Occlusion				✖	✖			✖
Branch Retinal Artery Occlusion		✖	✖		✖		✖	
Cilio-Retinal Artery Occlusion		✖			✖		✖	
Arteriosclerotic Retinopathy			✖					

*** (NVD = Neovascularization of the optic disc, NVE = Neovascularization elsewhere).**

Since most of retinal diseases affect the vessels, automated calculation to find the ratio of artery caliber to vein caliber (AVR) helps ophthalmologists in diagnosing retinal abnormalities. Accurate diagnoses of retinal diseases are critical issue, and since part of the diagnoses depend on the AVR, accurate classification of retinal vessels into arteries or veins is essential. There are different protocols for AVR calculation, for example, in Japan they consider the optic disc centered image and the largest adjacent pair of vessels in the area within 0.25-1 of the optic disc diameter from its margin are used for AVR calculation [46, 48, 42].

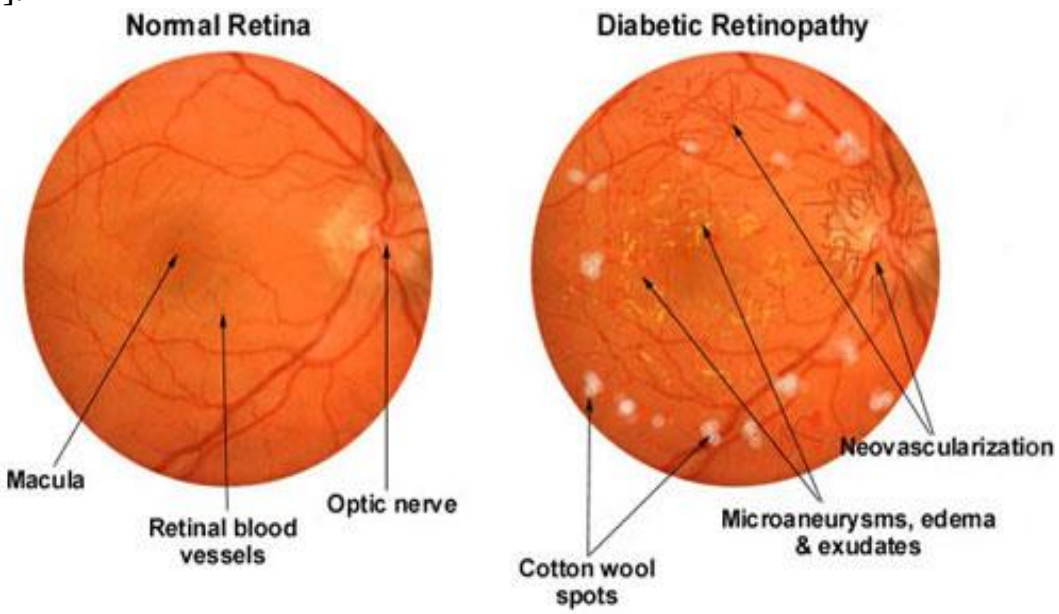
AVR ratio involves several problems such as: localization of the optic disc, vessels segmentation, accurate measurement for vessels diameter and classification of vessels into artery / vein. Classification of vessels is very essential step for AVR and accurate result is really important since misclassification errors may lead to large errors in final AVR measurement [42].

Some retinal diseases are mentioned below:-

1.5.1 Diabetes

It’s one of the most influential retinal diseases, it affects the body with different asymptotes, for example it makes changes in retinal vessels. If diabetes affects the vascular area of retina, it is named as diabetic retinopathy which is caused by complication of diabetes that causes blindness. In general, there are two types of diabetic retinopathy: non-proliferative and proliferative. In non-proliferative (NPDR) the dilation and

tortuosity of retinal veins are increased and intra-retinal hemorrhages is appeared .While in proliferative (PDR) abnormal vessels begin to grow on the surface of retina or the optic disc, this is called neovascularization, the wall of these new abnormal vessels are weaker and it can bleed or cause retinal detachment, see Fig. 1.8. Diabetic retinopathy in early stage can be treated by laser so that it is very important to regularly test and follow up retina by ophthalmologist, this reduces the risk of blindness [65, 48, 42, 28].



(a) (b)

Fig. 1.8: (a) Normal retina. (b) retina with diabetic retinopathy.

1.5.2 Vascular Occlusion

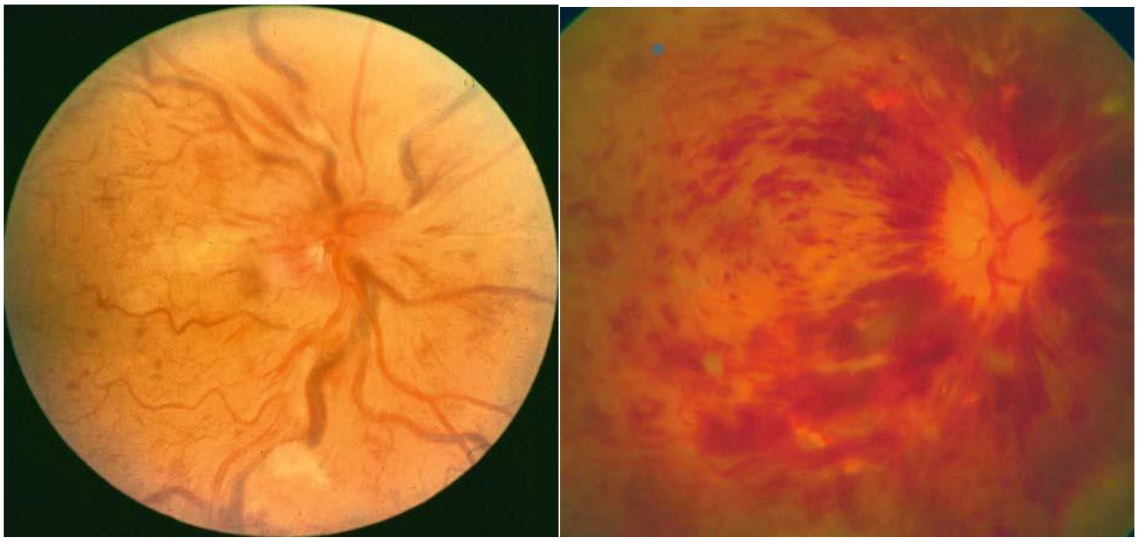
There are two types of vascular occlusion: central or branched vein occlusion and central or branched artery occlusion.

Central/ branch vein occlusion

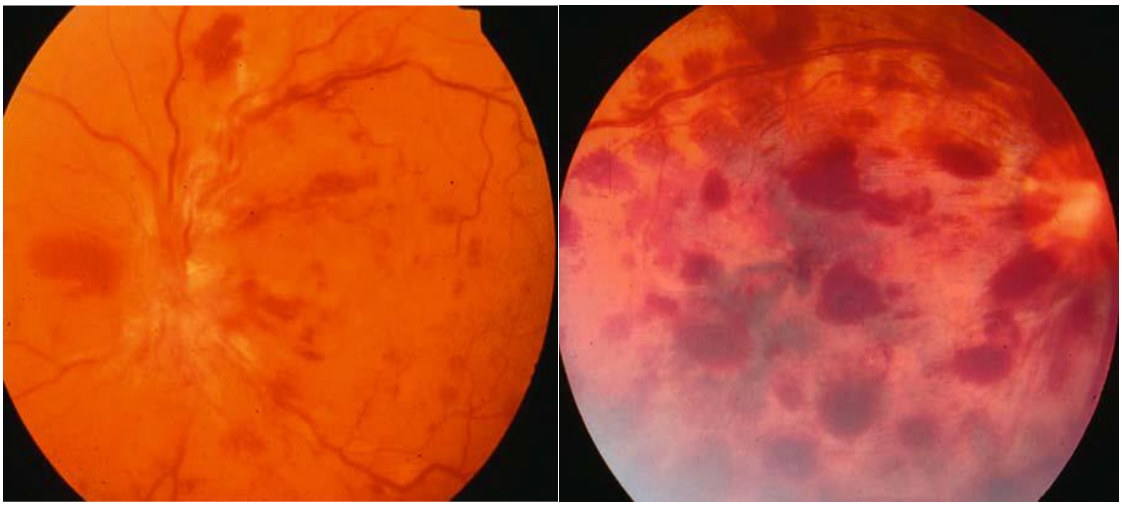
Central vein occlusion (CVO) affect whole retina since the central vein is blocked where branched vein occlusion (BVO) affect only a portion of the retina where the small branch of vein is blocked.

CVO in early stage has the following symptoms: flame-shaped bleedings or in some cases deep bleeding, slight engorgement of the veins, hyperemia and swelling of the optic nerve head. See Fig. 1.9.

While the earliest sign for branch vein occlusion (BVO) is slightly hemorrhages in the region where artery cross the vein or vice versa, hemorrhages appear in different degrees, often it takes long time (months up to year) to absorb. Away from this region the veins look dilated, congested and tortuous. In general, as occlusion occurs nearer to the optic disc, the size of affected retina will be larger.



(a) (b)



(c) (d)

Fig. 1.9: (a) CVO with heavy bleeding, (b) few bleedings and engorgement of the veins, (c) vitreous bleeding, (d) dot and plot bleeding

Central/ branch artery occlusion

In central artery occlusion (CAO) fiber layer becomes thickened and the region around the macula become whitening, see Fig. 1.10. Arteries become thinner and weaker, and cherry red spot develops at fovea.

Fig. 1.10: CAO after few hours of occlusion

After weeks, the optic disc become atrophic, neovascularization appear in iris and rarely appear at the optic disc [28].

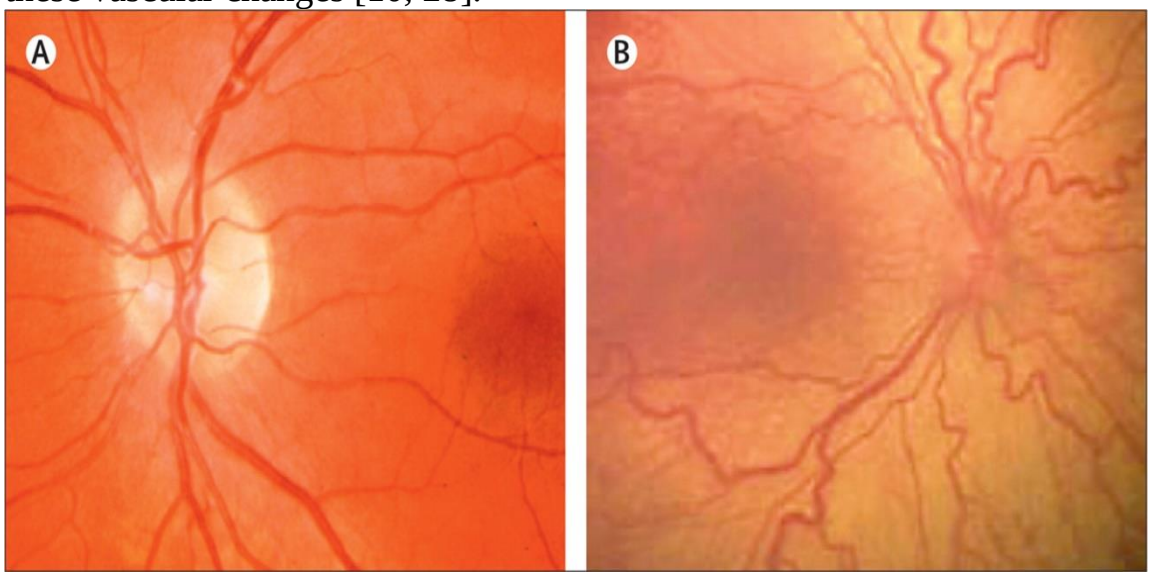
1.5.3 Retinopathy Of Prematurity (ROP)

ROP is a disease occurred with premature childhood babies, it can cause blindness. ROP isn't only caused by single reason, it is caused by a combination of several factors. There are two phases of ROP, in the first stage vessels grow slowly and even some vessels are lost, while in second phase a neovascularization occurs at the junction between the non-vascularized retina and vascularized retina, these new vessels are leaky.

ROP affect the posterior pole vascular, it initial caused mild venous dilation and it followed by tortuosity in arteriole, these vascular changes developed over years.

There is a virulent type of ROP which is retinopathy of prematurity plus, this type shows more dilation and tortuosity in arteries. See Fig. 1.11.

Early detection of diseases protects the child from future blindness, so if we can classify retinal vessels into artery and vein it becomes easier to notify these vascular changes [10, 28].

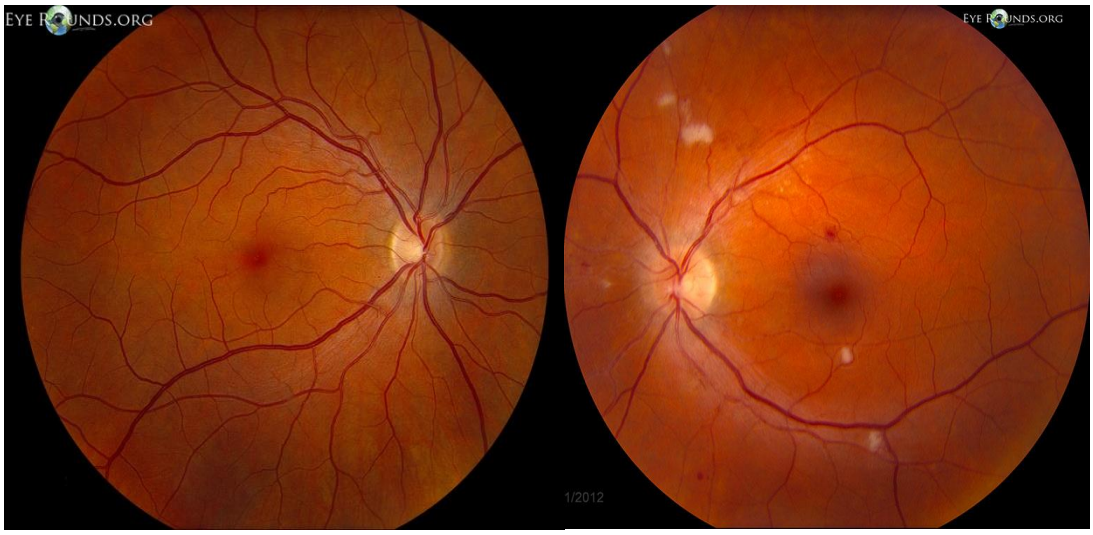


(a) (b)

Fig. 1.11: a) Normal retinal eye, b) retina with plus ROP showing dilation and tortuosity.

1.5.4 Hypertension

In hypertension, artery become narrower which can be discovered by calculating the AVR ratio for pair of parallel vessels at specific distance from the optic disc, this distance change in different protocols e.g., in USA this distance is defined from half-disc to one disc diameter from the optic disc margin and from quarter-disc to one disc diameter from the optic disc margin in Japan, i.e. decreasing in the AVR is an early sign for hypertension, also vessels tends to be tortuous, the edge of the optic disc become blurred [46, 54, 28]. See Fig. 1.12.



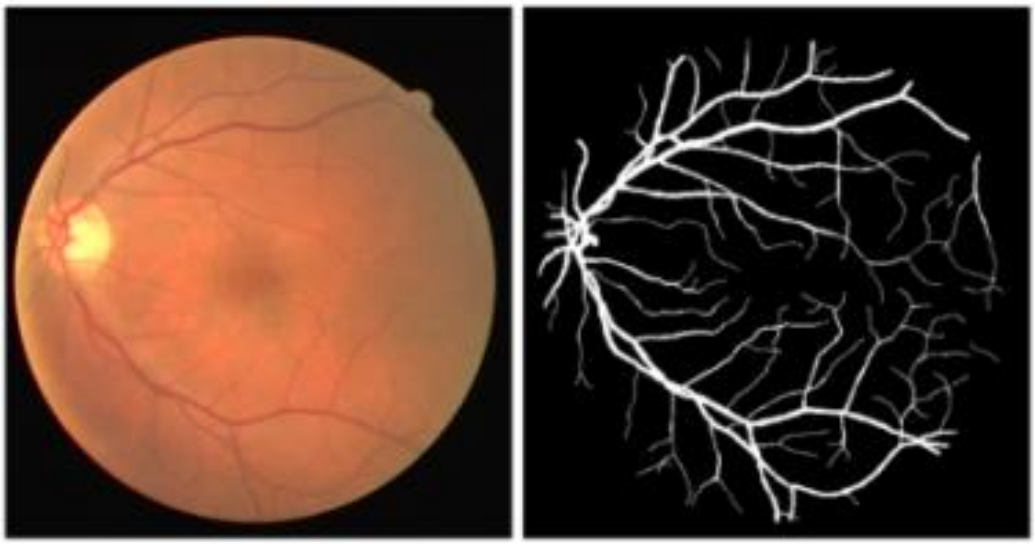
(a) (b)

Fig. 1.12: (a) Normal retina, (b) retina with hypertension.

1.6 Retinal Vessels Segmentation

Before classifying the vessels into artery or vein there is an essential step called segmentation of retinal blood vessels which have been heavily researched over the years. Segmentation can be considered as the heart of retinal image processing, it plays an important role in retinal disease detection also as mentioned it is an essential step for retinal vessels classification, it allows detecting geometrical changes in vessels such as tortuosity, diameter and/or length. Vessel segmentation is converting the colored retinal image into binary image by assign each pixel in the retinal image as vessel or background, see Fig. 1.13, which helps in recognizing the background in order to dismiss it from the calculation. In binary image, pixels have only two values either 0 which refers to background pixels or 1 which represent vessels pixels. Retinal vessel segmentation can be obtained manually or automatically, manual segmentation is very hard and time-

consuming especially since vascular network is complex and large number of images that expert need to segment so it was essential to find an automatic method or at least semi-automatic ones [39].



(a) (b)

Fig. 1.13: (a) Color retinal image, (b) segmented retinal image

Automatic methods provide consistency and accuracy, they also save time and effort, and they are very useful since they help in segmenting large number of retinal images in little time. Moreover, it reduces the workload on doctors in practical applications, even non-experts can use these methods, also automatic segmentation can provide good results or even better than manual ones in some cases [39].

Accurate blood vessels segmentation is a difficult task for several reasons: low contrast, the presence of noise, irregular widths of vessels and large variation in retinal images structures such as the optic disc size and location. The above difficulties and others deteriorate automatic blood

vessel segmentation. Several factors affect the segmentation such as choosing the appropriate images for training, feature detection method, using the global and local features models and prior knowledge.

On the other hand, the safety limitation that used to obtain the image making the boundary of vessels unclear, also small depth of focus in camera as well as the motion of the eye cause different degrees of blur. Moreover, the vessels have a phenomenon called center light reflex which may make vessels misunderstood as two vessels even retinal lesions which caused by retinal disease can disturb vessels segmentation. Several image enhancements such as normalization and denoising techniques have been developed to tackle these complications [22, 63, 13, 39].

Segmentation of blood vessels is a required step for further analysis that allows measuring attributes such as length, width, branching factor and tortuosity, with these indicators it’s possible to classify those vessels into artery or vein, also it’s possible to diagnose and evaluate the evolution of several eye diseases [22].

Segmentation methods had attention from a lot of researchers and there are several techniques to segment retinal blood vessels which are vary in the results performance and the complexity of the method, for example:

Pattern recognition techniques: each pixel is classified to a vessel or non-vessel class using some features [39].

Matched filtering: it bases on structural properties for classified objects, in 1989 this method developed for the first time, it uses the knowledge that the intensity profile for cross sectional vessel can be approximated by

Gaussian curve. Matched filter method uses 12 different kernels to detect vessels with different orientations. Most of researchers implement this algorithm in green channel; that's because this channel gives the best contrast between vessels and background [39].

Mathematical morphology: In general, morphology approaches that deal with structure and form of something, but mathematical morphology is the part that deals with mathematical theory to extract useful image components such as features, boundaries and skeleton. It uses some vessels properties such as the fact that vessels are connected and linear also smooth vessels curvature and other properties. In medical image segmentation the popular morphological operators are top-hat transformation and watershed transformations [39, 64].

Many other methods for images segmentation such as rule-based method, model based method, vessel tracing/tracking and pixel classification approach can be used. For more details see [39, 64, 57].

1.7 Retinal Vessels Classification

Retinal vessels classification into artery or vein is a very important topic since it helps ophthalmologist in diagnosing several eye diseases and it also represents an essential step to calculate AVR which is considered as an indicator for some retinal diseases, for instance, diabetic retinopathy. On the other side, automated classification for vessels has received less attention than vessels segmentation; that's because there are many

challenges that face the classification, such as low contrast and inhomogeneous lighting which are considered as basic challenges.

Classification of the retinal vessels is based on visual and geometrical features that help in distinguishing between vessels; the differences between artery and vein are mentioned previously. But in many cases, these differences are not enough, for example:

1- The outer regions of image are darker due to inhomogeneous lighting and this disturbs the classification since arteries look very like veins.
2- Center light reflex might disappear from the outer regions in low contrast images.
3- Near the optic disc (OD), veins and arteries alternate and rarely cross each other. But as they branch and go far away from the OD it is possible that there are two adjacent veins or arteries, it is even possible that artery cross an artery or vein cross a vein.

So that classification using only these features doesn’t guarantee good results [42].

Many researchers in this field try to simplify the problem by classifying only the main vessels, which helps in avoiding the confusion caused by small ones. Small vessels are very hard to classify as they look very similar to each other, using tracking method in addition to graph analysis may help in classifying these small vessels.

1.8 Related Work

Medical image processing is getting more and more important. As vessel classification is part of it, many researchers working on vessel classification has published their work and others are trying to improve methodologies to have high performance. Table 1.3 shows a summary for researchers work.

Table 1.3: Summary of retinal vessels classification techniques

Reference	Year	Features	Classification method	Database	Evaluation method-Results
Grisan and Rugger [19]	2003	Variance red + mean hue	Fuzzy clustering + vessel tracking	35 fundus images	Accuracy − 87.6%
Li et al. [36]	2003	Central reflex in Green	Minimum Mahalanobis distance	505 segments from different fundus images	True positive rate Arteries:2.46% Veins:89.03%
Jelinek et al. [25]	2005	Mean and std of RGB, H	Naive-Bayes	20 fundus images	Mean accuracy of 70% over 8 images
Ruggeri et al. [56]	2007	Variance red + mean hue	Fuzzy clustering + vessel tracking	14 fundus images	AVR −Correlation: 0.73–0.83
Narasimha Iyer et al. [47]	2007	Structural: central Reflex Functional: the ratio of the vessel optical densities	SVM	25 dual wavelength fundus images	True positive rate −Arteries: 97% Veins: 90%
Kondermann [33]	2007	RGB − profile-based and ROI-based	MLP	4 high quality retina images	Accuracy −95.32% (by ROI-based features +MLP)
Tramontan et al. [62]	2008	R contrast	Thresholding + vessel tracking	20 fundus images (DCCT)	AVR −Correlation: 0.88

Niemeijer et al. [49]	2009	HSI + width, contrast, steered Gaussian derivatives (on G channel)	KNN	DRIVE	AUC − 0.88
Muramatsu et al. [46]	2010	RGB	LDA	DRIVE	Accuracy − 88.2%
Muramatsu et al. [45]	2011	RGB + R,G contrast values	LDA	DRIVE	Accuracy − 92.8%
Niemeijer et al. [50]	2011	HIS + R,G (colors, averages, and maxima)	LDA	65 fundus images from University of Iowa	AUC − 0.84Mean AVR − 0.67
Rothaus and Jiang [55]	2011	HSI + 5 model features	k-Means clustering	MARS	Classification error less than 30% in 80% of the test images
Zamperini et al. [69]	2012	16 color and Positional features	Linear normal Bayes	42 fundus images from Ninewells Hospital, Dundee	Accuracy − 93.1%
Mirsharif et al. [45]	2013	Eight color features (R, G, LAB) + structural features in post-processing	LDA	DRIVE + 13 images from Khatam-Al-Anbia Eye Hospital	Accuracy DRIVE: 90.16% Khatam: 88.18%
Relan et al. [54]	2013	Mean of (R, G, H) and variance of R	GMM-EM	35 fundus images from ORCADES	Accuracy − 92% (13.5% unlabeled)
Joshi et al. [27]	2014	Mean and variance of G, H	Fuzzy C-means clustering	EYECHECK	Accuracy −91.44%
Hatami and Goldbaum [21]	2016	Nine feature extraction methods (RGB, PCA, LBP, etc.)	Multiple classifier systems (MCSs)	STARE	Accuracy − 90.7% AUC − 0.97 (by Rand. Committee + MS-RI LBP)

Chapter Two

Linear Discriminant Analysis

Chapter Two

Linear Discriminant Analysis (LDA)

To classify objects or observations into different classes you need to define some variables or features to help classification technique and understand the nature of this observation. Large number of variables (features) of an observation doesn't necessary mean good discriminant, some variables are redundant and irrelevant or maybe noisy, and the solution will be more effective if we get rid of them, this step is called dimensionality reduction. Dimensionality reduction becomes a necessary step before applying any classification algorithm or even it may be considered as preprocessing step especially if we deal with high dimensional data, see Fig. 2.1, which represents a three dimensional data with redundant features and how dimensionality reduction helps in separating the data efficiently. Dealing with a large number of features requires more storage space and it becomes computationally more complex, also it requires larger training data as the number of features is increased [30]. However, overlapping is also a problem which may interrupt classification in which the inputs of different classes are mixed.

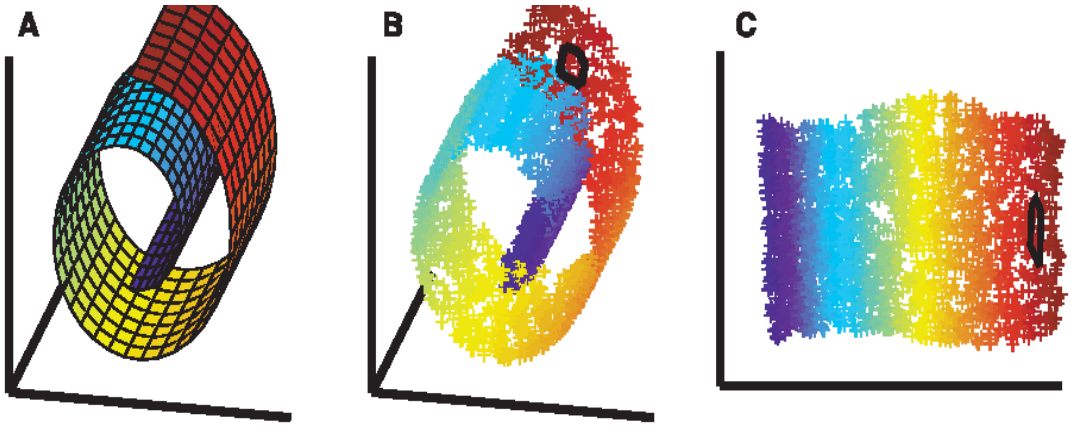


Fig. 2.1: A three dimensional data (A and B), reduced into a two-dimensional data making it easier to separate into different classes (C)

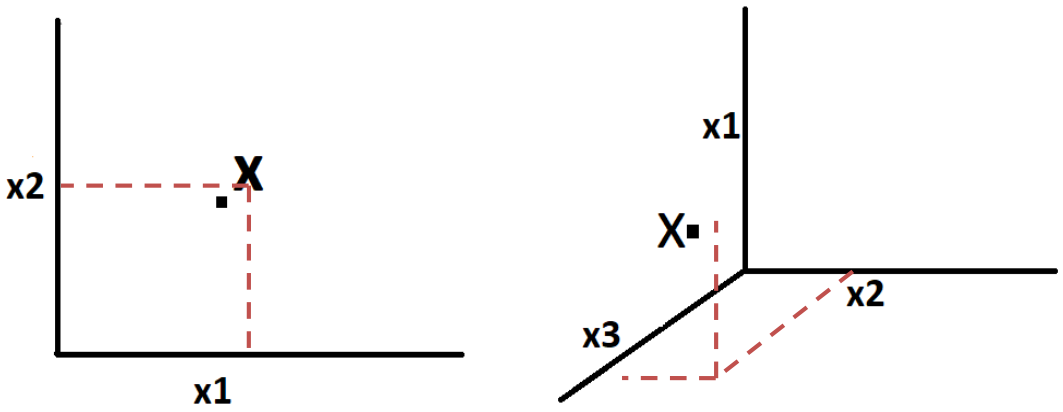
To avoid these problems, the data is transformed into a lower dimensional space (reduce features or variables). There are many methods which deal with dimensionality reduction for classification; some of these methods are supervised approaches such as Linear Discriminant Analysis (LDA) and unsupervised approaches such as Principle Component Analysis (PCA) [35, 53, 66, 11, 26].

2.1 Fundamental Definitions of Classification

Classification is getting more and more important with the increasing number of information in the world in many disciplines, to convert this process into an automatic one we need to define a mathematical model that manages to define those items with common features in the same group using distance metrics. We are introducing now some relevant terms to this topic.

2.1.1 Pattern

Pattern is defined as quantitative or structural description of objects, it is represented by a vector in d-dimension, $\boldsymbol{x} = [x_1, x_2, \dots, x_d]$, where $x_1, \dots, x_d$ are considered as features which describe the important characteristics of a given object or data point, a set of extracted features is called feature vector. Any object can be represented as a point in feature space, Fig. 2.2 represents a two-dimensional and 3-dimensional feature space respectively [30].



(a) (b)

Fig. 2.2: (a) Two dimensional space, (b) three dimensional space.

Pattern recognition is considered as a classification problem where the inputs are assigned to classes, it can be supervised or unsupervised. Pattern recognition is used in many applications, such as speech recognition, computer aided design (CAD), finger print, face detection and retinal vessel classification and segmentation [30, 39].

A set of patterns that fed the algorithm to find specific category of an output is called the inputs; this input basically consist of feature vectors,

where the length of these vectors is considered as the dimension of this input data.

If we consider a pattern $\boldsymbol{x}_i = [x_{i,1}, x_{i,2}, \dots, x_{i,d}]$, where d is the number of features for a given pattern, then the input will be represented in matrix form as follows:-

$$A = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,d} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \dots & x_{n,d} \end{bmatrix}$$

where n represents the number of patterns and d is the dimension of input data matrix [30].

2.1.2 Classes

Each set of patterns that share common properties are called a category or class. The input data matrix is divided into pre-known classes. The goal of pattern classification is to classify new patterns into distinct classes and hence the distribution of the classes should be separable, which means each pattern belong to a unique category [30].

2.1.3 Classification

Classification consists of two main stages:

Feature selection or feature extraction: features are considered as distinct characteristic of an image, and extracting these features helps in classifying the image. Feature extraction and selection methods are applied to decide which features are useful to get the best performance. For different feature extraction or selection methods, different set of feature vectors are produced, but they always try to minimize the within-class variance and maximize between-class variance. The input image contains large number of features which may be redundant or complex information which lead to errors and misclassification problem, so that it is important to decide which of these features are useful and give best discriminant between the two classes.

Feature extraction is considered as a especial type of dimensionality reduction, in which the most important features are kept by projecting the data into a lower dimensional feature space, it is very important to choose appropriate extraction method according to the input dataset. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are considered as feature extraction methods.

Feature selection methods select a subset of features from the original set of features; these methods omit those features with no effective information. There are three methods for feature selection which are: wrappers, filters and embedded.

In the next stage, the classifier receives a d-dimensional feature vector for a particular observation then assigns it to one labeled class from a set of predefined classes. [30, 34, 15].

Pattern classifier is trained using a set of training vectors, (called the training dataset), and the performance for this classifier is evaluated by implementing the classifier in some feature vectors from another dataset called test dataset and evaluate the results.

2.1.4 Supervised / Unsupervised Methods

Mistakes may take place during data analysis or setting the relationship between multi features. These mistakes and more are efficiently solved using machine learning in which most of datasets can be represented using some known features.

Machine learning algorithms are divided into two main types: supervised and unsupervised learning. Supervised learning assumes that classifiers have a prior knowledge about the corresponding classes, this knowledge is integrated in the process. The supervised algorithm learns using a data which consists of feature vectors and corresponding labels, this data is called training dataset. But why we learn the algorithm if we already know the answer? That is because our goal is generalization by discovering patterns which help in predicting future unknown inputs.

Unlikely; unsupervised learning doesn't use any prior knowledge, that means the features vector hasn't any corresponding labels, it uses the redundancy in the data to find a best representation that describes them, the

goal for unsupervised learning is basically to discover the groups for input data or to predict the class for new data [11].

2.2 Linear Discriminant Analysis (LDA)

Discriminant Analysis (DA) is a statistical technique that provides an understanding of the differences between objects that belong to different groups with respect to several variables simultaneously. DA is used when we have a prior knowledge about classes, and it is used to classify objects into predefined groups or classes in addition to dimensionality reduction problems.

The idea behind DA goes back to around 1920 when Karl Pearson proposed a method to compute between-class distance called coefficient of racial likeness (CRL). In the 1930s, P. C. Mahalanobis studied another method to calculate between class distance. R. A. Fisher was the first one who translated the problem to a linear combination of related features and it was renamed as linear discriminant analysis (LDA). After Fisher many extensions and refinements appeared [3].

LDA is used for classification and dimensionality reduction, as a classifier it assumes normal distribution for distinct groups, identical covariance matrix for each class and features are independent. But LDA for dimensionality reduction can break these assumptions. It finds a linear combination of features, in order to classify observations into two or more predefined classes, that combination gives the optimal separation between classes [3, 35, 66, 26].

Dimensionality reduction is a very important task in pattern recognition, it aimed to discover meaningful structures and unexpected relationships in the multivariate data. There are many methods that represent high dimensional data into lower dimensional space, it can be categorized into feature selection methods and feature extraction methods. In feature selection methods only few useful features are kept and the less important ones are discarded. While in feature extraction methods the original dimension is reduced by projecting high dimensional data into lower dimensional space to constrict fewer features from large number of original features. Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA) and Factor Analysis (FA), which can be described as an extension of PCA, are considered as popular methods for feature extraction [41].

There are many advantages for dimensionality reduction, for example, in low dimensional space the complexity of computation is lower which reduces the time and space, also the low dimensional data is more reliable to show actual properties of the original data.

The LDA is used in many different areas; researchers used LDA in face recognition, bioinformatics, text recognition, speech recognition and multimedia information retrieval including vessels classification. For instance in the literature, S.Balakrishnama et al. used LDA to classify forestry images based on their beauty scenic [7]. H. Yu et al. implemented direct LDA algorithm in face recognition [68]. C. Muramatsu implemented LDA in vessels classification to classify center line pixels into artery or vein, 30 out of 40 pairs of artery and vein was correctly determined [46].

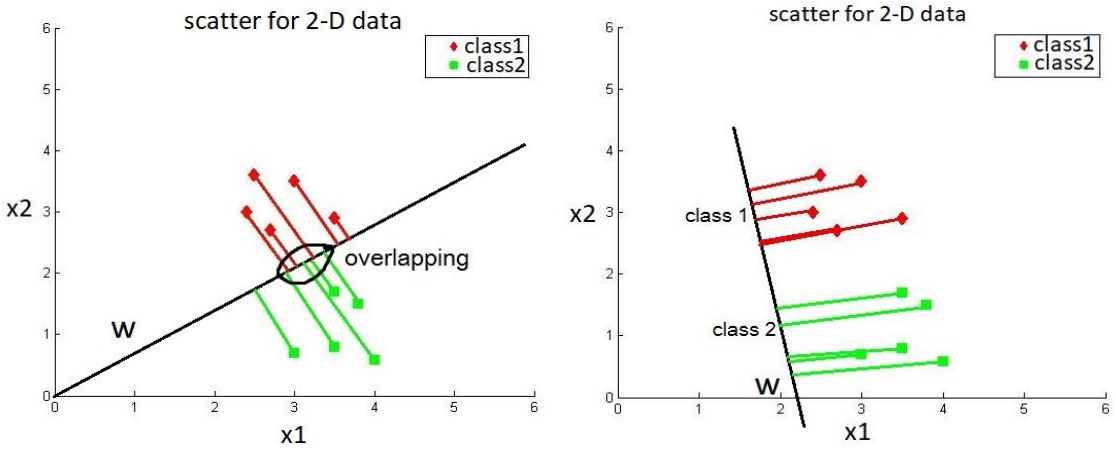
2.2.1 The idea Behind LDA

LDA is a statistical technique that tries to separate input data by reducing its dimension; it looks for the best projection to transfer the given data from high dimensional space to lower one in order to separate the data easily. The basic idea of LDA is to find an orientation w (weight matrix) that projects d-dimensional space to l-dimensional space (where $d > l$) such that the classes are well separated in the lower dimensional space, and then apply a classification technique in l dimensional space.

2.2.2 LDA for Two Classes in Two Dimensions

Given an $n\ by$ 2-dimensional dataset $A = [\boldsymbol{x_1}, \boldsymbol{x_2}]$ of two classes with labels C_1(class1) and C_2(class2), where $\boldsymbol{x_i}$ is a vector of n dimension, our goal is to find the direction of a weighted column vector $\boldsymbol{w}$ which separates the data as well as possible when the data is projected onto $\boldsymbol{w}$, see Fig. 2.3. The projection of data point $\boldsymbol{x} \in \mathbb{R}^{\mathbf{2\times 1}}$ onto $\boldsymbol{w}$ is written by

$$y = \boldsymbol{w}^T \boldsymbol{x} \tag{2.1}$$



(a) (b)

Fig. 2.3: The weighted vector $\boldsymbol{w}$ projects the data: (a) non-conveniently separated, (b) well separated.

So to find the direction of the weighted vector, define $\boldsymbol{\mu}_1 \; and \; \boldsymbol{\mu}_2$ as the dataset mean vectors of $C_1 \; and \; C_2$ respectively:

$$\boldsymbol{\mu}_1 = \sum_{\boldsymbol{x_i} \in C_1} \frac{\boldsymbol{x_i}}{n_1} \;, \qquad \boldsymbol{\mu}_2 = \sum_{\boldsymbol{x_i} \in C_2} \frac{\boldsymbol{x_i}}{n_2}$$

where n_i is the sample size for class i. The projection means $\boldsymbol{m}_1 and \; \boldsymbol{m}_2$ are defined as:

$$\boldsymbol{m}_i = \boldsymbol{w}^T(\boldsymbol{\mu}_i) \tag{2.2}$$

and the scatter variance matrices for C_1, C_2 after projection are defined as

$$S_1^2 = \sum_{\boldsymbol{x_i} \in C_1} (\boldsymbol{w}^T(\, \boldsymbol{x_i} - \boldsymbol{\mu}_1))^2$$

$$S_2^2 = \sum_{\boldsymbol{x_i} \in C_2} (\boldsymbol{w}^T(\, \boldsymbol{x_i} - \boldsymbol{\mu}_2))^2$$

Fig. 2.4 represents the projected means and the scatter variance matrices. To get best separability between the two classes, we would like to make the distance between two projected means as large as possible $|\boldsymbol{m}_1 - \boldsymbol{m}_2|$ and make both scatter variances as small as possible $(S_1^2 + S_2^2)$, our problem becomes to maximize $|\boldsymbol{m}_1 - \boldsymbol{m}_2|$ and minimize $(S_1^2 + S_2^2)$ simultaneously, i.e. maximize the following ratio:

$$J(\boldsymbol{w}) = \frac{(\boldsymbol{m}_1 - \boldsymbol{m}_2)^2}{(S_1^2 + S_2^2)} \tag{2.3}$$

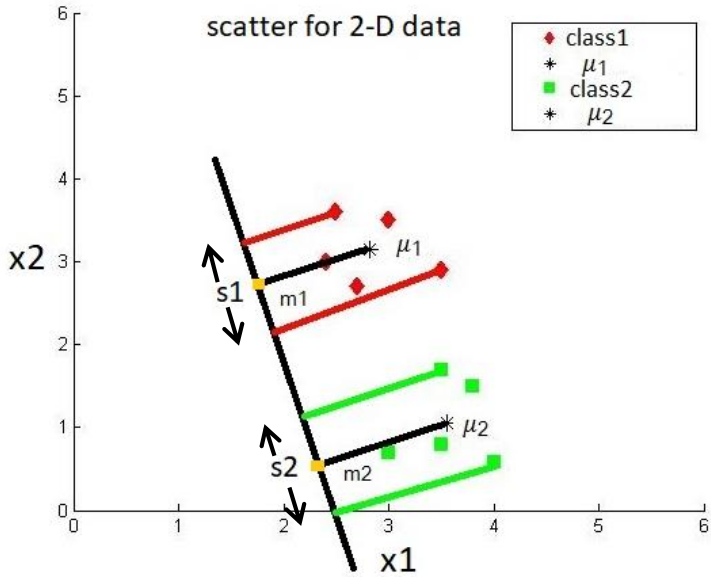


Fig. 2.4: Scatter variance matrices $\boldsymbol{S_1}, \boldsymbol{S_2}$ and projected means $\boldsymbol{m_1}, \boldsymbol{m_2}$

The vector $\boldsymbol{w}$ that maximizes the function (J) is called *Fisher's linear discriminant*, rewrite the numerator and denominator using equation (2.1)

$$(\boldsymbol{m}_1 - \boldsymbol{m}_2)^2 = (\boldsymbol{w}^T\boldsymbol{\mu}_1 - \boldsymbol{w}^T\boldsymbol{\mu}_2)^2$$

$$= \boldsymbol{w}^T(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\boldsymbol{w}$$

$$= \boldsymbol{w}^T S_b \boldsymbol{w} \qquad (2.4)$$

where $S_b = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T$ is defined as the between-scatter matrix. For the denominator we can rewrite it as:

$$S_1^2 = \sum_{\boldsymbol{x_i} \in C_1} (\boldsymbol{w}^T\,(\boldsymbol{x_i} - \boldsymbol{\mu}_1))^2$$

$$= \sum_{\boldsymbol{x_i} \in C_1} \boldsymbol{w}^T(\boldsymbol{x_i} - \boldsymbol{\mu}_1)\,(\boldsymbol{x_i} - \boldsymbol{\mu}_1)^T\boldsymbol{w}$$

$$= \boldsymbol{w}^T\,\Sigma_1\boldsymbol{w}$$

where Σ_1 is within-scatter for C_1, similarly $S_2^2 = \boldsymbol{w}^T\,\Sigma_2\boldsymbol{w}$ and Σ_2 is within-scatter for C_2, hence

$$S_1^2 + S_2^2 = \boldsymbol{w}^T (\Sigma_1 + \Sigma_2) \boldsymbol{w}$$
$$= \boldsymbol{w}^T S_w \boldsymbol{w} \quad (2.5)$$

where $S_w = (\Sigma_1 + \Sigma_2)$ is the within-scatter matrix, see Fig. 2.5, our objective function will be:

$$J(w) = \frac{\boldsymbol{w}^T S_b \boldsymbol{w}}{\boldsymbol{w}^T S_w \boldsymbol{w}} \quad (2.6)$$

Since $\boldsymbol{w}^T S_w \boldsymbol{w}$ is constant our problem becomes a constrained optimization problem, so we want to maximize $\boldsymbol{w}^T S_b \boldsymbol{w}$, assume $\boldsymbol{w}^T S_w \boldsymbol{w} = 1$, the solution can be obtained using Lagrangian:

$$L(\boldsymbol{w}, c) = \boldsymbol{w}^T S_b \boldsymbol{w} - c(\boldsymbol{w}^T S_w \boldsymbol{w} - 1)$$

where c is the Lageangian constant. Now derive the function with respect to w and set it equal to zero.

$$\frac{d}{d\boldsymbol{w}}\big(L(\boldsymbol{w}, c)\big) = 2S_b \boldsymbol{w} - 2cS_w \boldsymbol{w} = 0$$

$$S_b \boldsymbol{w} = cS_w \boldsymbol{w}$$

$$S_w^{-1} S_b \boldsymbol{w} = c\boldsymbol{w}$$

Since $S_b = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T$ we get

$$S_w^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{w} = c\boldsymbol{w}$$

$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T$ is 1×2 and $\boldsymbol{w}$ is 2×1, so the multiplication result is a constant,

$$S_w^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \frac{c}{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{w}} \boldsymbol{w} = \lambda \boldsymbol{w}$$

where

$$\lambda = \frac{c}{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{w}}$$

But we are interested in the direction of $\boldsymbol{w}$ no matter what the magnitude is (let $\lambda = 1$), so

$$\boldsymbol{w} = S_w^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \tag{2.7}$$

[4, 41]

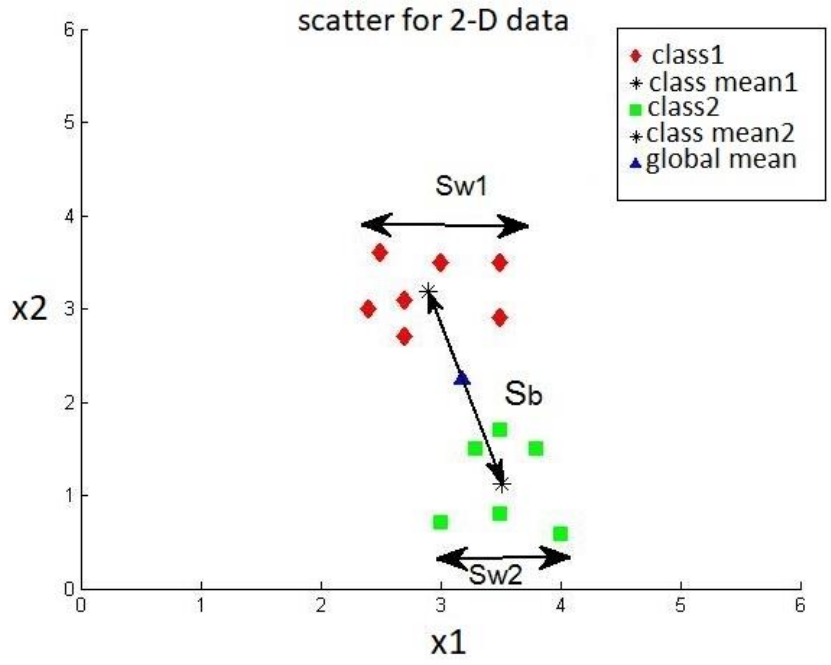


Fig. 2.5: Scatter plot for a two classes data in two-dimensional space, illustrating $\boldsymbol{S_w}$, $\boldsymbol{S_b}$, $\boldsymbol{\mu_1}, \boldsymbol{\mu_2}$ and $\boldsymbol{\mu}$.

2.2.3 LDA for Higher Dimensions

Given a data matrix $A = [\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}]$ with n by d-dimensional training samples (feature vectors) divided into k disjoint classes $C_1, C_2, \ldots, C_k$, each class is a subset of A with mean $\boldsymbol{\mu}_i = \frac{\sum \boldsymbol{x_i}}{n_i}$, where n_i is the sample size for class i, μ represent the mean for a given data matrix A (global mean), m_i represents the projection for each μ_i respectively. Each class C_i is a subset of A and $C_1 \cup C_2 \cup \ldots \cup C_k = A$.

The within class scatter for class C_i is defined as:

$$\Sigma_i = \sum_{\boldsymbol{x_i} \in C_i} (\boldsymbol{x_i} - \boldsymbol{\mu}_i)(\boldsymbol{x_i} - \boldsymbol{\mu}_i)^T$$

hence within-scatter matrix can be defined as

$$S_w = \frac{1}{n} \sum_{i=1}^{k} \Sigma_i$$

$$S_w = \frac{1}{n} \sum_{i=1}^{k} \sum_{\boldsymbol{x_i} \in C_i} (\boldsymbol{x_i} - \boldsymbol{\mu}_i)(\boldsymbol{x_i} - \boldsymbol{\mu}_i)^T \qquad (2.8)$$

and the between-classes scatter matrix S_b is defined as:

$$S_b = \frac{1}{n} \sum_{i=1}^{k} n_i (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T \qquad (2.9)$$

Another scatter matrix can be defined; which is called the total scatter matrix S_t, where $S_t = S_w + S_b$

$$S_t = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{x_i} - \boldsymbol{\mu})(\boldsymbol{x_i} - \boldsymbol{\mu})^T \qquad (2.10)$$

Also we can simplify the formulation of the scatter matrices in equations (2.8, 2.9, 2.10) through precursors H_w , H_b and H_t. These formulations can save memory and simplify eigen- analysis, and they are defined as follows:

$$H_w = \frac{1}{\sqrt{n}} [C_1 - \boldsymbol{\mu}_1(e^1)^T, \dots , \qquad C_k - \boldsymbol{\mu}_k(e^k)^T]$$

$$H_b = \frac{1}{\sqrt{n}} \left[\sqrt{n}_1(\boldsymbol{\mu}_1 - \boldsymbol{\mu}), \dots , \sqrt{n}_k(\boldsymbol{\mu}_1 - \boldsymbol{\mu})\right]$$

$$H_t = \frac{1}{\sqrt{n}} (X - c\boldsymbol{e}^T)$$

where $\boldsymbol{e}^i = (1,1, \dots ,1)^T \in \mathbb{R}^{n_i}$ and $\boldsymbol{e} = (1,1, \dots ,1)^T \in \mathbb{R}^n$

The within scatter and between scatter matrices can be expressed as follows:

$S_w = H_w \times H_w^T$, $S_b = H_b \times H_b^T$, $S_t = H_t \times H_t^T$

The projection of the data in lower dimension can be represented by $y = W^T\boldsymbol{x}$, where $W \in \mathbb{R}^{d\times l}$, let $\|W\| = 1$ since the magnitude of W doesn't effect the decision, so projected data can be computed along W for every point, we need to choose W that gives best separation of data. The means for original data will also be projected into lower dimension, it becomes $\boldsymbol{m}_i = W^T\boldsymbol{\mu}_i$, within and between scatter matrices will be changed as follows:

$$S_w^l = \frac{1}{n}\sum_{i=1}^{k}\sum_{\boldsymbol{x}\in A_i}(W^T\boldsymbol{x} - W^T\boldsymbol{\mu}_i)(W^T\boldsymbol{x} - W^T\boldsymbol{\mu}_i)^T$$

$$= W^T\left(\frac{1}{n}\sum_{i=1}^{k}\sum_{\boldsymbol{x}\in A_i}(\boldsymbol{x} - \boldsymbol{\mu}_i)(\boldsymbol{x} - \boldsymbol{\mu}_i)^T\right)W$$

$$S_w^l = W^T S_w W \tag{2.11}$$

$$S_b^l = \frac{1}{n}\sum_{i=1}^{k} n_i\big((W^T\mu_i - W^T\mu)(W^T\mu_i - W^T\mu)^T\big)$$
$$= W^T\left(\frac{1}{n}\sum_{i=1}^{k} n_i\big((\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \mu)^T\big)\right) W$$

$$S_b^l = W^T S_b W \tag{2.12}$$

$$S_t^l = \frac{1}{n}\sum_{i=1}^{k}\big((W^T\boldsymbol{x}_i - W^T\boldsymbol{\mu})(W^T\boldsymbol{x}_i - W^T\boldsymbol{\mu})^T\big)$$
$$= W^T\left(\frac{1}{n}\sum_{i=1}^{k}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T\right) W$$

$$S_t^l = W^T S_t W \tag{2.13}$$

Using the matrix trace properties, we have

$$tr(S_w) = \frac{1}{n}\sum_{i=1}^{k}\sum_{\boldsymbol{x}\in C_i}\|\boldsymbol{x} - \boldsymbol{\mu}_i\|_2^2$$

$$tr\,(S_b) = \frac{1}{n}\sum_{i=1}^{k} n_i\|\boldsymbol{\mu}_i - \boldsymbol{\mu}\|_2^2$$

where $tr(A)$ is the trace of matrix A, it's clear that trace of S_w measures the distance between the data point and its corresponding mean, while trace of S_b measures the distance between the classes mean and the global mean.

Optimal LDA is obtained by simultaneously maximizing the distance between classes ($tr(W^T S_b W)$) and minimizing the within class distance ($tr\,(W^T S_w W)$), thereby guaranteeing maximal separability. The following ratio is defined using equations (2.11, 2.12):

$$J(W) = \frac{tr(W^T S_b W)}{tr(W^T S_w W)} \tag{2.14}$$

the function $J(W)$ can be rewritten using S_t since $S_t = S_w + S_b$

$$J(W) = \frac{tr(W^T S_b W)}{tr(W^T S_t W)}$$

should be maximized by differentiating $J(W)$ function with respect to W and setting the derivative to be zero. In order to do that; we will use constrained optimization. Assume that our problem is to maximize $tr(W^T S_b W)$, given $tr(W^T S_w W)$ is a constant, say 1, the solution can be obtained using the Lagrangian:

$$L(W, \lambda) = tr(W^T S_b W) - \lambda(tr(W^T S_w W) - 1)$$

$$\frac{d}{dW}\big(L(W, \lambda)\big) = 2S_b W - 2\lambda S_w W = 0$$

Again, we looking for the direction of W so:

$$S_b W = \lambda S_w W \tag{2.15}$$

$$S_w^{-1} S_b \, W = \lambda W \tag{2.16}$$

Also the optimal W can be calculated by solving the following optimization problem

$$W = \underset{W}{argmax} \left\{ S_b^l \ \left(S_t^l\right)^{-1} \right\}$$

Note that the weight matrix W has size of $d \times l$ and the $rank\ (S_b)$ is at most $d - 1$ so there exist at most $d - 1$ eigenvectors corresponding to a

non-zero eigenvalues, and λ is a diagonal matrix of $d-1$ eigenvalues. The solution will be the largest l eigenvalues of $S_w^{-1}S_b$ [26, 23].

LDA can be used to determine the most significant features for the new data to be classified, in our case we will separate/classify the vessels into two classes, artery or vein classes.

2.3 Singularity Problem

There are several applications that we can use LDA to solve them, but there are some applications such as text mining, face recognition, medical image analysis and microarray data classification comes from a high dimensional space and all scatter matrices have opportunity to be singular matrices, this case is called *singularity* or *undersampled problem* which is a major limitation of classical LDA [53, 23].

Over the years, researchers proposed many algorithms to solve the singularity problem they are considered as LDA extensions, for example, Regularized LDA (RLDA), Penalized LDA (PLDA), Orthogonal LDA (OLDA), Uncorrelated LDA (ULDA) and Modified LDA (MLDA), for more details see [66, 67, 26]

Another method to solve singularity problem is to use more general inverse which is pseudo-inverse that can be computed even if S_w is singular.

2.3.1 Pseudo-Inverse

It is known that the matrix is said to be invertible if it is a non-singular matrix. Recently, applied mathematics requires the inverse for singular

matrices, so that mathematicians worked hard until they discovered that even when the matrix is not invertible it still has either left or right inverse.

Let $A \in \mathbb{R}^{m\times n}$ be a right invertible matrix (or left invertible matrix), then there exists a matrix $R \in \mathbb{R}^{n\times m}$ or $L \in \mathbb{R}^{n\times m}$ for right/left invertible matrix, such that:

$$AR = I_m \qquad Or \qquad LA = I_n$$

The matrix is said to have generalized inverse if it has some inverse-like matrix, in general this generalized inverse isn't used but it is considered as fundamental basis to create various generalized inverse for particular purposes. The most important generalized inverse is Moore-Penrose inverse or pseudoinverse, it is discovered by E. H. Moore (1920) and Roger Penrose (1950) [38].

Definition1. [38] If $A \in \mathbb{R}^{m\times n}$, then there exists a unique $A^{+} \in \mathbb{R}^{n\times m}$ that satisfies the Moore-Penrose conditions:-

1. $AA^{+}A = A$
2. $A^{+}AA^{+} = A^{+}$
3. $(AA^{+})^{H} = AA^{+}$
4. $(A^{+}A)^{H} = A^{+}A$

where A^{H} denote the conjugate transpose, A^{+} denote the pseudo inverse [38, 16].

If $rank(A) = r$ then one can decompose A as $A = FR$ where $F \in \mathbb{R}^{m\times r}$, $R \in \mathbb{R}^{r\times n}$, hence $rank(R) = rank(F) = r$, and A^{+} can be defined as:

$$\mathrm{A}^{+} = R^{H}(RR^{H})^{-1}(F^{H}F)^{-1}F^{H}$$

There are three special cases:

Case 1: if A is a full row rank $r = m$, then $A^{+} = A^{H}(AA^{H})^{-1}$

Case 2: if A is a full column rank $r = n$, then $A^{+} = (A^{H}A)^{-1}\,A^{H}$

Case 3: if $rank(A) = m = n$ then $A^{+} = A^{-1}$

In equation (2.15), if S_w is singular we can't use classical LDA directly, so considering more general case, where S_w is singular, one can use Moore-Pseudo invers of $S_w$which denoted by S_w^{+}, and then we can use the eigen decomposition of $S_w^{+}S_b$. Note that we can replace S_w by S_t since as we said $S_t = S_w + S_b$ [38, 16].

2.3.2 LDA Extensions

As mentioned before, the singularity problem is considered as main problem that faces LDA, the generalized inverse is considered as a solution for this problem but in spite of the simplicity of using it, it might give an unstable prediction, so that researcher go towards new methods to solve this critical problem in addition to have a stable prediction by using LDA extensions. One of the popular, effective and simple ways to get rid of singularity is to use Regularized LDA (RLDA).

Regularized LDA (RLDA)

RLDA has applications in different areas such as face recognition and medical image analysis, it stabilizes the estimation of sample covariance matrix and improves the performance of classification using LDA [67].

We add a constant $\alpha > 0$ to the diagonal elements of the within scatter matrix in order to avoid singularity problem:

$$\widetilde{S_w} = S_w + \alpha I_d \tag{2.17}$$

where I_d is the identity matrix of size d, α is the regularization parameter [26].

Since S_w is a positive semi-definite matrix then $S_w + \alpha I_d$ is positive definite and hence non-singular [26, 16]

If α is large the information in scatter matrix can be significantly disturbed, while small α may not be efficient enough to solve a singularity problem, so that choosing an appropriate α is critical issue, the common way to get it is by estimating an optimal α from finite set $\Lambda = \{\alpha_1, \alpha_2, \dots, \alpha_m\}$

The computational cost for finding optimal α can be high especially for large m. However large m is often desirabled in practice to obtain a good α [67].

Optimal RLDA can be obtained by solving the following optimization problem:

$$G^{RLDA} = \frac{trace(G^T S_b G)}{trace(G^T \widetilde{S_w} G)} \tag{2.18}$$

This equation can be solved by applying eigen-decomposition of $\left(\widetilde{S_w}\right)^{-1}S_b$ [66, 67, 26].

There are some disadvantages in RLDA, for example, regularization parameter α chosen by cross-validation might waste time especially if the data come from a high dimensional data. Moreover, for large number of features RLDA may be not effective, therefor we need to search for the optimal α .

Recently, researchers recommended empirical methods to find optimal α, but still no theoretical justification to support these methods [66].

2.4 Principal Component Analysis (PCA)

PCA is a well know multivariate method for dimensionality reduction, it can be used as supervised or unsupervised method. PCA tried to reduce the dimension into lower one in which most of essential information is preserved by extracting the important information or features and then express them in new orthogonal variables called principal components. PCA is considered as the oldest and the most common multivariate method, it is also known as Karhunen-Loeve methods [41, 20, 8]

PCA goals:

1- Extract the most important information or features from the input dataset.
2- Decrease the dataset size by keeping only the important variables
3- Simplify the dataset description.
4- Analyze the structure of the features and observation.

To achieve these goals we need to find new variables obtained from the linear combinations of features and they are called principal components.

PCA idea:

PCA looks for the direction of the largest variation of data and set the first axis in the same direction. In order to find the second axis, calculate the perpendicular line which covers the largest number of reminding points and so on, these perpendicular lines are called principal components.

Now assume that $\boldsymbol{x}$ is the given data matrix, the projection of $\boldsymbol{x}$ on the direction of $\boldsymbol{w}$, which is the first principle component, is

$$z = \boldsymbol{w}^T \boldsymbol{x} \qquad (2.19)$$

PCA maximizes the variance, and for unique solution and to make the direction the important factor, let $\|\boldsymbol{w}\| = 1$, using equation (2.19) we get

$$\begin{aligned} cov(z) &= E(z\, z^T) \\ &= E(\boldsymbol{w}^T \boldsymbol{x}\, (\boldsymbol{w}^T \boldsymbol{x})^T) \\ &= E(\boldsymbol{w}^T \boldsymbol{x}\, \boldsymbol{x}^T \boldsymbol{w}) \\ &= \boldsymbol{w}^T E(\boldsymbol{x}\, \boldsymbol{x}^T)\, \boldsymbol{w} \\ &= \boldsymbol{w}^T cov(\boldsymbol{x})\, \boldsymbol{w} \\ &= \boldsymbol{w}^T \Sigma\, \boldsymbol{w} \end{aligned}$$

It's clear that $E(\boldsymbol{w}) = \boldsymbol{w}$ since it's not a data-dependent matrix, and with $\boldsymbol{w}^T = \boldsymbol{w}^{-1}$, we seek $\boldsymbol{w}$ that maximize the $var(\boldsymbol{x})$ subject to the constrain that $\boldsymbol{w}^T \boldsymbol{w} = 1$. Writing this using Lagrangian, we have

$$\max \boldsymbol{w}^T \Sigma \ \boldsymbol{w} \ - \ \alpha \ (\boldsymbol{w}^T \ \boldsymbol{w} - 1)$$

Taking the derivative with respect to $\boldsymbol{w}$ and set it equal to zero, we have

$$\alpha \ \boldsymbol{w} = \Sigma \ \boldsymbol{w} \qquad (2.20)$$

which is an eigen-value problem. Since $\Sigma \in \mathbb{R}^{N \times N}$, then it has N eigen-values with N corresponding eigen-vectors. We choose the eigenvector corresponding to the largest eigenvalues for the variance to be maximized. The second principle component is perpendicular to the first one, and it can be easily calculated, and so on.

Another method to find the principle component:

First: centralize the given data matrix by finding the mean for each column then subtract the column element from its center data. This first step is required in order to shift the center of mass for the data points to the origin.

Second: to find the eigen-values and corresponding eigen-vectors of Σ , use the Singular Value Decomposition (SVD) for the resulted data. We get

$$S = CDC^T$$

where S is the estimator to Σ , and C is a $d \times d$ matrix whose i^{th} column is the eigenvectors and D is a diagonal matrix whose diagonal elements are the eigenvalues $\lambda_1 , \lambda_2, \dots, \lambda_d$, to reduce the dimension from d to l, where $d > l$, we pick the largest l eigenvalues with respect to corresponding eigenvectors [41].

The algorithm of PCA can be reduced in these steps:

1- Find the Covariance matrix of $\boldsymbol{x}$ by computing the centered of the dataset.

2- Find the eigen-values and corresponding eigen-vector for the computed covariance matrix, which can be find using SVD.

3- Pick the largest l eigen-values and build a matrix W which contains the eigen-vectors corresponding to the picked eigen-values.

4- Multiply $\boldsymbol{x}$ by W.

Differences between LDA and PCA are presented in the following: [7,6]

1- PCA changes both the location and the shape of the data in transformed space while LDA finds the decision boundary in order to achieve maximum separability between classes without changing the location of data.

2- PCA classify features, it transforms data from input data space into feature space see Fig. 2.6, on the other hand, LDA is used in data classification to achieve best separability between classes.

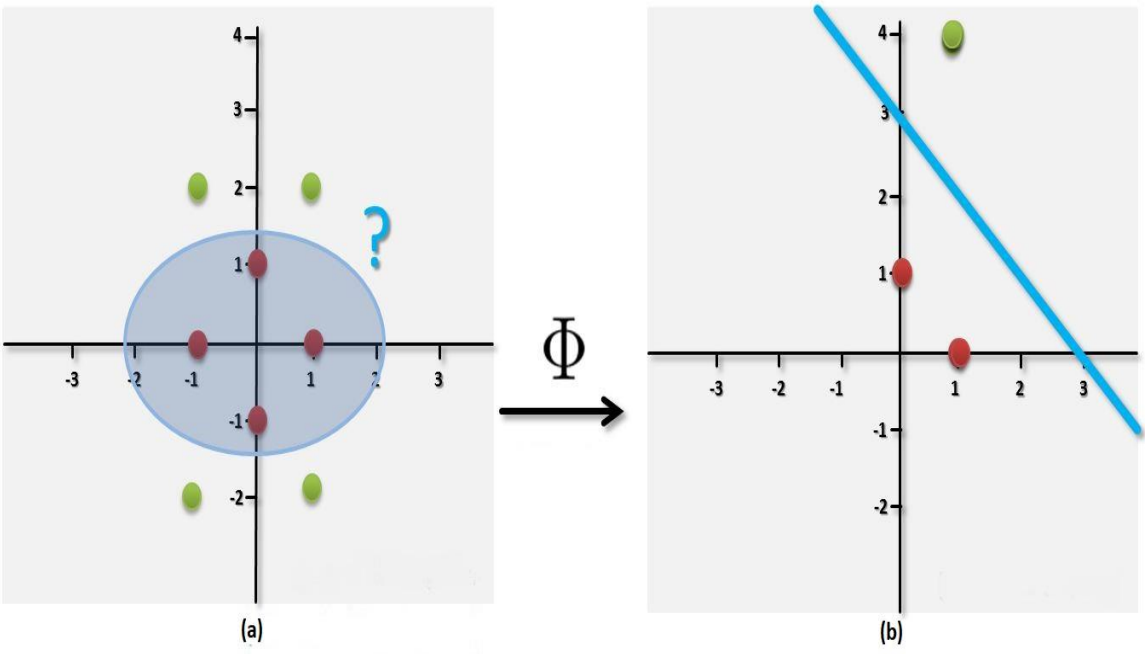


Fig. 2.6: (a) Input data space, (b) feature space.

Balakrishnama et al. tested four classifiers which are: LDA, PCA, SVM (Support Vector Machines) and ICA (Independent Components Analysis) on 478 training and 159 test forestry images, after comparing the results they found that SVM has lowest miss-classification rate and LDA achieved better results rather than PCA [7].

2.5 The Implemented Classifiers

After reducing the dimension and get rid of redundant features, one needs to implement a classification technique in order to classify vessels in the test dataset into artery or vein. Classifiers are either supervised (need prior knowledge about class labels) or unsupervised (doesn't need prior knowledge). Now we will introduce common classifiers in retinal vessels classification.

2.5.1 Support Vector Machine (SVM)

Support Vector Machine (SVM) is one of the most popular classification methods in machine learning, Vapnik was the first one who introduce SVM in 1992 and since then it came to be popular. SVM work very well in reasonable dataset and it gives better classification result than other techniques, but it becomes a high computational cost in extremely large datasets. To understand this algorithm see Fig. 2.7, it's clear that the drawn lines give good separation for the given 2-D data, but surely there is one

line which gives the best separability, SVM chose the line which separate the data best [41].

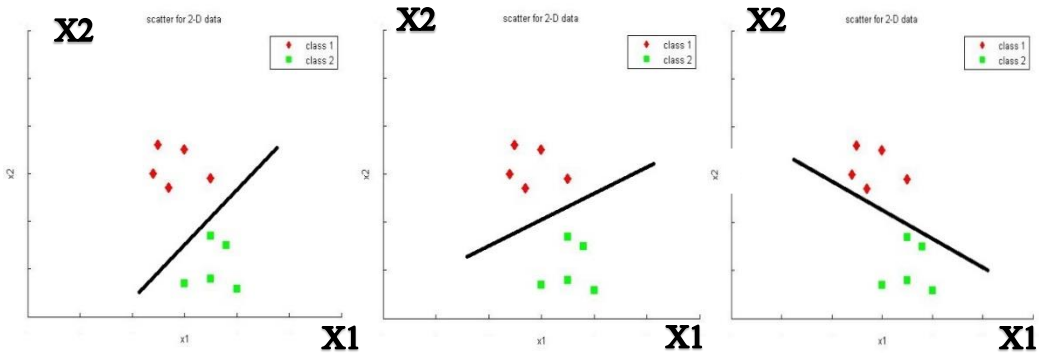


Fig. 2.7: Three different lines to separate the same data into two classes.

SVM calculates the perpendicular distance from classification line to the closest data point forms both sides and then draw a region which is symmetric around the line, it form a rectangle in 2-D space, cylinder in 3-D space and hyper-cylinder in higher dimensions, see Fig. 2.8, any positive distance from the decision boundary is known as margin and the best margin is the largest one and hence SVM looks for the largest margin since that means better classifier generalization, the closest points to the classification line known as support vectors and these points are considered as the most useful points since they are the points that might be wrong.

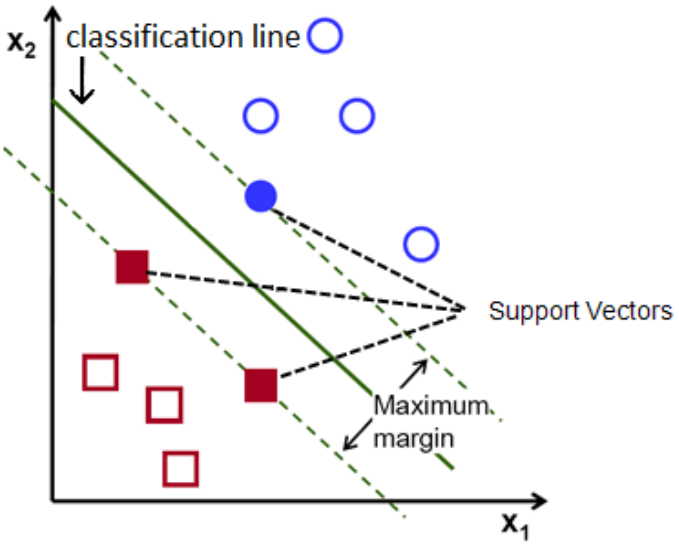


Fig. 2.8: SVM classifier.

As we note in Fig. 2.7, the middle classifier has largest margin classifier so that it is chosen as the best classification line. After training the data and determine the support vectors, we get rid of all data except these points and use them in classification.

The classification line has the following equation:

$$y = \boldsymbol{w}.\boldsymbol{x} + b \tag{2.22}$$

where $\boldsymbol{w}$ is the weight vector, $\boldsymbol{x}$ is the input vector and b is constant.

In general we say if we implement a given input vector $\boldsymbol{x}$ in the equation (2.22) and it returns a positive value then $\boldsymbol{x}$ lies above the line or belong to 'o' class and under the line otherwise.

SVM doesn't depend on whether the value is positive or not, it checks this value with respect to the margin, i.e. for a given value M, we can say, if $\boldsymbol{w}.\boldsymbol{x} + b \geq M$ then $\boldsymbol{x}$ belong to circles class while if $\boldsymbol{w}.\boldsymbol{x} + b \leq -M$ then $\boldsymbol{x}$ belong to squares' class. The vector $\boldsymbol{x}^{+}$which lies on the circles' class

boundary is a support vector and it satisfies $\boldsymbol{w}.\boldsymbol{x}^{+} = M$. If we want to find the closest point from the other class which lies on the boundary we will calculate the perpendicular distance from the boundary line of first class to boundary line of the second class, the first point we hit is the closest one, write it as $\boldsymbol{x}^{-}$. It is clear that this perpendicular line is equal to $2M$, see Fig. 2.8, so $|\boldsymbol{x}^{-} - \boldsymbol{x}^{+}| = 2M$ and hence:

$$M = \frac{1}{2|\boldsymbol{w}|} = \frac{1}{2\sqrt{\boldsymbol{w}\,.\,\boldsymbol{w}}} \tag{2.23}$$

As we said before we want M to be as large as possible, equation (2.23) says that for largest M we need $|\boldsymbol{w}|$as small as possible, if that is the only constrain we consider, then we choose $\boldsymbol{w}$ to be zero, but we need also the classification line to separate the data into two classes, so our problem is to find a decision boundary that classify the data well and minimize $|\boldsymbol{w}|$ [41, 4, 8], for more details see Appendix (C).

2.5.2 Bayesian Classifier

Prior probability is defined as the probability of classifying a specific data as class one or two regardless of the features vector value, and its denoted by $P(C = C_0)$ where $C_0 = 0$ for first class and $C_0 = 1$ for second one. Posterior probability calculates the probability of some event after having new information, it is denoted by $P(C|\boldsymbol{x})$, Bayes' rule calculates the posterior probability using prior probability as follows:

$$P(C|\boldsymbol{x}) = \frac{P(C)P(\boldsymbol{x}|C)}{P(\boldsymbol{x})}$$

where $P(C)$ is prior probability which calculates how much each class appears in our training dataset, $P(\boldsymbol{x}|C)$ denotes the conditional probability or the likelihood of C with respect to $\boldsymbol{x}$, while $P(\boldsymbol{x})$ defines the marginal probability for $\boldsymbol{x}$ and it is called evidence.

$$P(\boldsymbol{x}) = P(\boldsymbol{x}|C = 1)\, P(C = 1) + P(\boldsymbol{x}|C = 0)P(C = 0)$$

Bayes rule is very important since it allows calculating the posterior probability with much easier computations, posterior probability helps us in assigning new observation to a specific class.

For more than two classes $C_i\,, i = 1\,, 2\,, \dots, k$, where k defines number of classes and satisfying $P(C_i) \geq 0$, $\sum_{i=1}^{k} P(C_i) = 1$ then

$$P(C_i|\boldsymbol{x}) = \frac{P(\boldsymbol{x}|C_i)P(C_i)}{P(\boldsymbol{x})}$$

where

$$P(\boldsymbol{x}) = \sum_{i=1}^{k} P(C_i)P(\boldsymbol{x}|C_i)$$

We choose C_i with maximum posterior probability.

If the independence relationship between features is unknown, we assume that the features are conditionally independent which gives us a naïve Bayes rule [4], see Appendix (C).

2.5.3 K-Nearest Neighborhood (K-NN)

K-NN is one of the simplest supervised classification techniques in machine learning algorithms. K-NN tries to classify a new point which

doesn't have a label with respect to closest K points from the training dataset. If $K = 1$, we classify the test point with respect to the closest labeled point. For two classes case K must be odd number, the distance from test point to each point in training set are calculated and the commonly used distance function is Euclidean distance as follows

$$d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\| = \sqrt{(\boldsymbol{x} - \boldsymbol{y})^2} = \left(\sum_{i=1}^{m} ((x_i - y_i)^2)\right)^{1/2}$$

K-NN classifier follows two steps process, first training the classifier by reading the feature vector in training data and read its label, and then it assigned the unlabeled data with respect to nearest K neighborhood using the majority voting. See Fig. 2.9 which illustrates the K-NN classifier, as one can see we have 3 different classes and one unlabeled point (diamond), if we consider $K = 1$, the point will be classified as square class, if you take $K = 5$, by majority voting the point will be classified as circle, for $K = 10$ again using majority voting it will be considered as star class [31, 24].

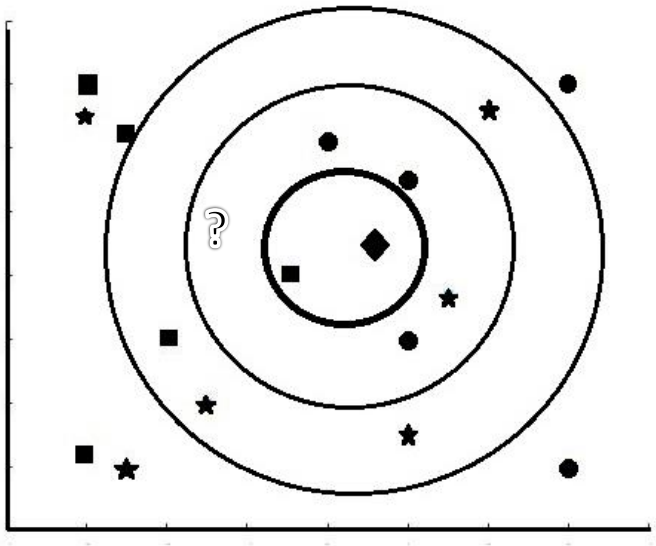


Fig. 2.9: K-NN classifier.

2.5.4 K-Means

K-means is considered as the most popular clustering algorithm, clustering divides a given dataset which contains n data point in d dimensional space into k different clusters (groups), where the data point that belongs to a one group is more similar to each other from the other points in other groups.

First of all the number of classes or clusters must be determined, let $X = \{x_1, x_2, \dots, x_n\}$ be a data set that have k clusters, $C = \{c_1, c_2, \dots, c_k\}$ be set of k clusters and $S = \{s_1, s_2, \dots, s_k\}$ where s_i is a set of element of class i .

The K-means algorithm follows the following steps

1- Choose the center of c_k using random sampling

2- Determine the member of s_k according to minimizing the following distance function

$$F = \sum_{i=1}^{n} dis((s_k)_i, c_k)$$

where $dis((s_k)_i, c_k)$ defines the Euclidean distance between elements of s_k and the cluster c_k .

3- Calculate new c_k using

$$c_k = \frac{\sum_{i=1}^{n} (s_k)_i}{|s_k|}$$

where $|s_k|$ is the number of s_k elements.

4- Redo step 2 and 3 until c_k converges.

It is not necessary to have a unique result if we use K-means since it depend on random sampling to pick up c_k for first time. If the initial cluster is close to the final solution, then K-mean algorithm gives better result. Hence many algorithms and attempts was proposed in cluster initialization

problem. One of these methods was proposed by Duda and Hart [11]. Furthermore, Pena et al. compared between different methods for the cluster initialization problem [29].

Chapter Three
Implementing LDA in Vessels Classification

Chapter Three

Implementing LDA in Vessels Classification

3.1 Introduction

Classifying vessels into artery or vein basically depends on the different properties of vessels, for instance artery is brighter with a bright center light reflex, on the other hand vein is thicker and darker, for more illustration see section (1.2). However, these properties are useful in the regions around the optic disc but they become less useful as we go farther.

Vessels classification faces many challenges influenced by their inconsistent properties, for example: the low contrast of the images as a result of the fact that different vessels have different contrast and inhomogeneous light in the background, also the two types of vessels are similar in shape and color (for instance the color of vein may by similar to the color of some artery in another image or even in the same image but in different region). Consequently, classification methods that use only this feature are expected not to obtain good results [55].

There are many approaches to classify vessels into artery or vein in fundus images; a lot of these approaches involve five steps: vessel segmentation, select the region of interest (ROI), extract features from vessels, classify the extracted feature vector and finally combine all results to classify each pixel into artery or vein, see Fig. 3.1 [42].

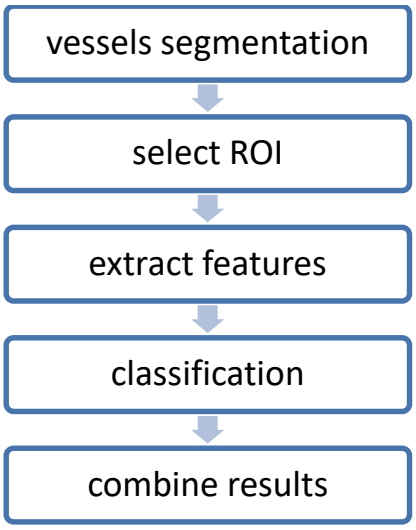


Fig. 3.1: Classifying retinal vessels block diagram.

The RGB retinal image is usually used to read features of the vessels, these features are implemented as training data for the classifier in order to separate data into groups. Using some decision rule in which an observation is classified as class one if the feature vector lays above the decision boundary and as class two otherwise [11]. The most important issue in this problem is to choose the optimal decision rule which makes the misclassification error as small as possible.

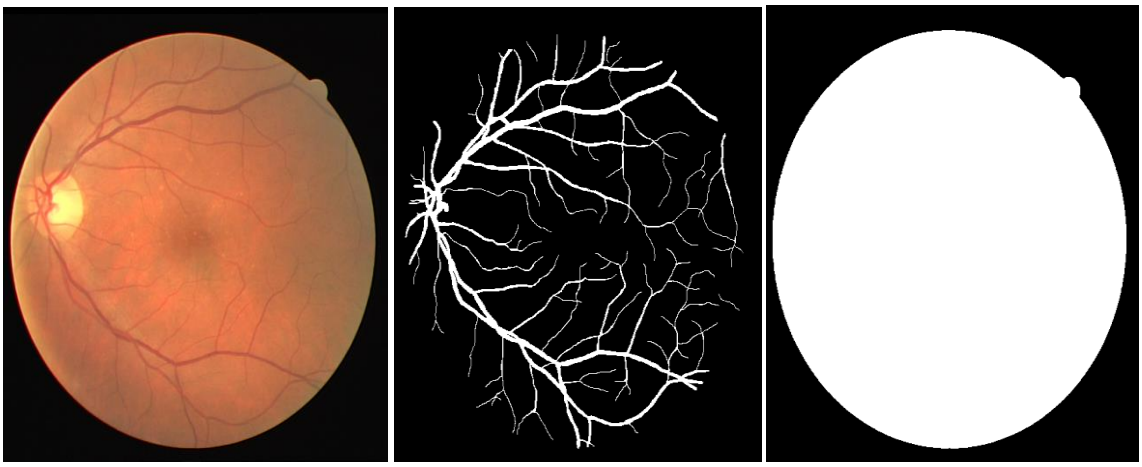
After introducing the used database in our research, we will discuss the implemented algorithm in details in the following sections.

3.2 The DRIVE Database

In this work images from DRIVE (Digital Retinal Images for Vessel Extraction) database are used, which is available for public. It contains 40 colored retinal images, taken by CR5 non-mydriatic 3CCD Canon camera with 45 degree field of view (FOV), the FOV appears as a circle with 540 pixel diameter. Seven of these retinal images contain pathology such as

hemorrhages. The images in the data base are divided into training and testing sets, each set contain 20 images [63, 59].

The training set contains the colored retinal images in addition to their FOV masks and corresponding manual retinal vessels segmentation see Fig. 3.2. While in the testing set, there are color retinal images with corresponding FOV mask in addition to two manual retinal vessels segmentation done by two different ophthalmologists for each image [59].



(a) (b) (c)

Fig. 3.2: (a) Colored retinal image, (b) manual vessels segmentation, (c) FOV mask.

The color images in the DRIVE database are saved in TIF format while for both FOV mask and vessels segmentation are saved in GIF format. The size for each image is 565 x 584 pixels, and they are obtained from a diabetic retinopathy screening program in the Netherlands [46].

The aim of our research is to classify vessels into artery or vein, so that we need a ground truth images in which vein appear as blue and artery as red, see Fig. 3.3, with the help of an expert ophthalmologist we obtained the

manual Artery/Vein classification of vessels for the training and testing images.

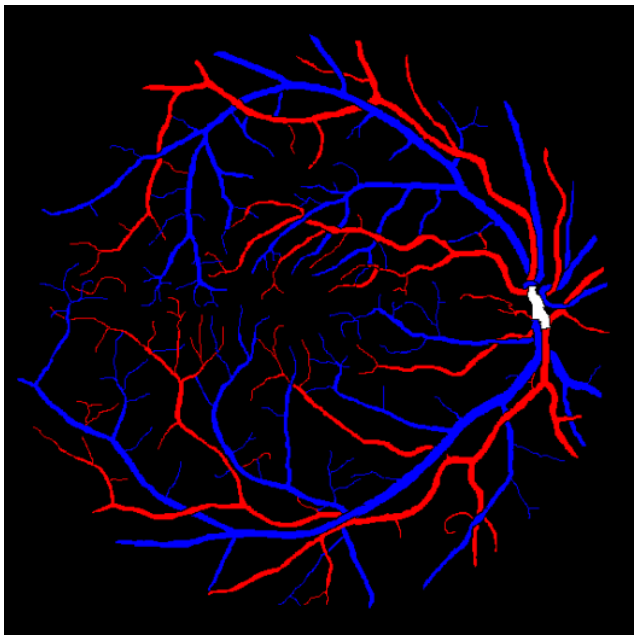


Fig. 3.3: Artery/Vein ground truth.

3.3 Methodology

In this section we will discuss in details the structure of our program, we chose MATLAB software to automatically classify vessels into artery or vein. We start reading the training images, then some enhancement techniques are implemented on these images, after that appropriate features are selected from the region of interest. In the next step, we reduce the dimension of features to be applied in different classifiers. Finally, we calculate the accuracy for each used classifier in order to define the best dimensionality reduction algorithm and the best classifier, see Fig.3.4.

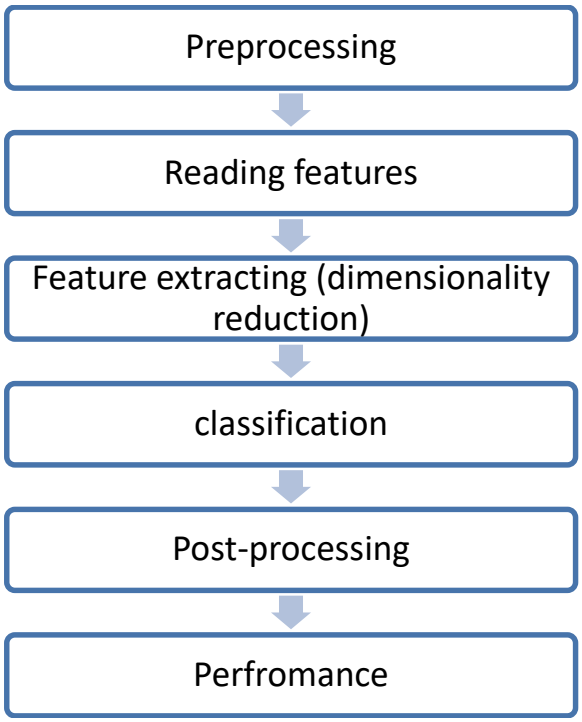


Fig. 3.4: A block diagram of the implemented retinal vessels classification algorithm.

3.3.1 Preprocessing

Preprocessing is a very essential step before any process technique in order to improve the results. The main problems that face the classification algorithm is: the low contrast between vein and artery in retinal images, non-uniform illumination problem which affects the recognition results. Also as we said before it is almost impossible to recognize small vessels correctly. These problems can be avoided using some preprocessing steps.

First, we choose a set of images as training set; 10 images from training dataset are randomly selected, the retinal images which clearly shows signs of retinopathy such as hemorrhage are excluded, see Fig. 3.5, since in these images the disease appears clearly, where the goal of the automatic classification is to discover the early signs of diseases.

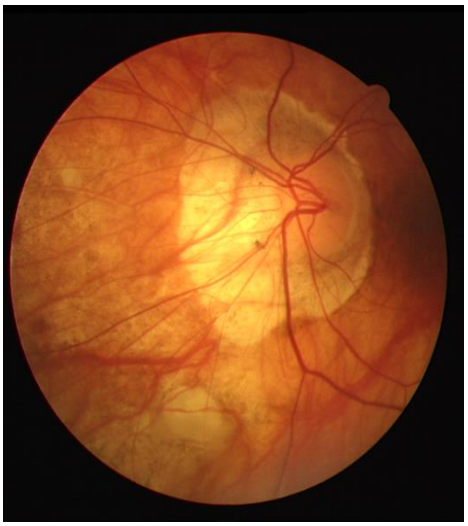


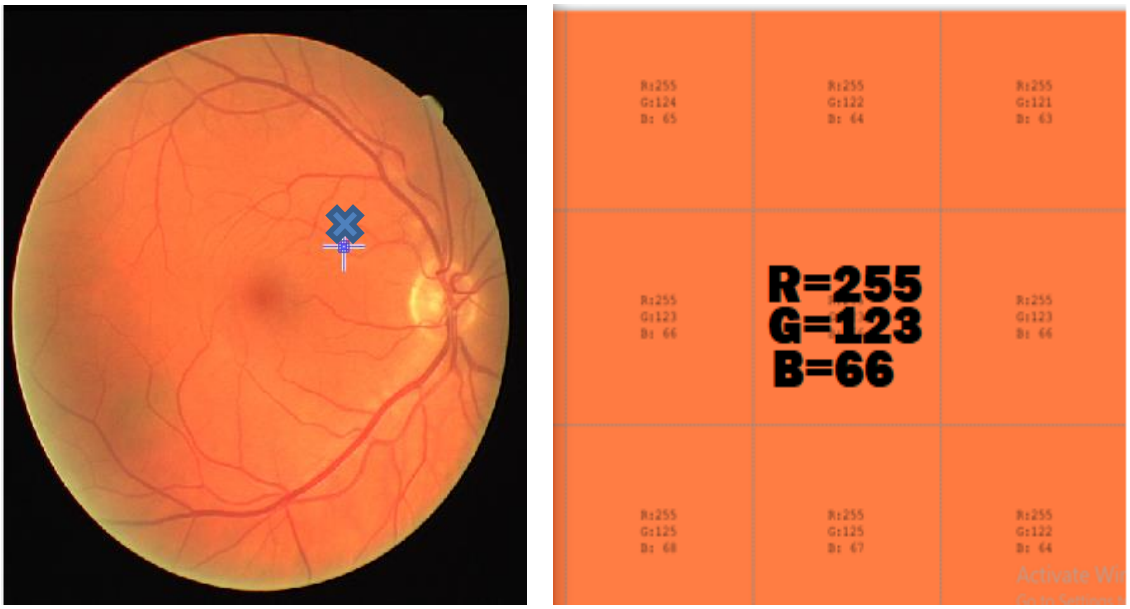
Fig. 3.5: Clear sign of retinopathy in a retinal DRIVE image.

Because of the limited number of training and testing images that we have (only 14 images) and to get more reliable result, instead of reading only 10 random images as training and 4 as testing, we decided to train the classifier using 13 images and test it on the rest one. This step is repeated for all available images, allowing us to train the classifier with larger dataset and test our classifier 14 times. Then the average results from these repeated steps are calculated in order to get the final result.

In the second preprocessing step, we try to apply some filters on color retinal images in order to uniform the illumination and increase the contrast between veins and arteries. There are many filters in MATLAB, for example, average filter and Gaussian filter. By using one of these filters we modify each pixel in an image based on some function of neighborhood,

Images are treated as matrices in MATLAB, binary or gray images are considered as 2-D matrix. The RGB image is a 3-D matrix, first matrix represent red channel, second is the green channel while the third one is the blue channel, see Fig. 3.6. We apply the filters on a specified pixel in

addition to its neighbor's pixels. Fig. 3.6 shows the 3×3 window where the center pixel has 8 neighbors.



(a) (b)

Fig. 3.6: (a) RGB image, (b) the neighborhood of the marked pixel in (a)

The average filter works in some specific window with size $m \times m$ in which it replaces the center pixel by the average of the pixels in that window by using some kernel function. For example, if we consider 3 × 3 kernel to define the average filter, the value of each pixel will change as follow:

$$f(i,j) = \sum_{k,l} I(i+k, j+l)\, h(k,l)$$

where I is the image and h is the kernel. Fig. 3.7 shows an example for applying 3×3 average filter kernel [61].

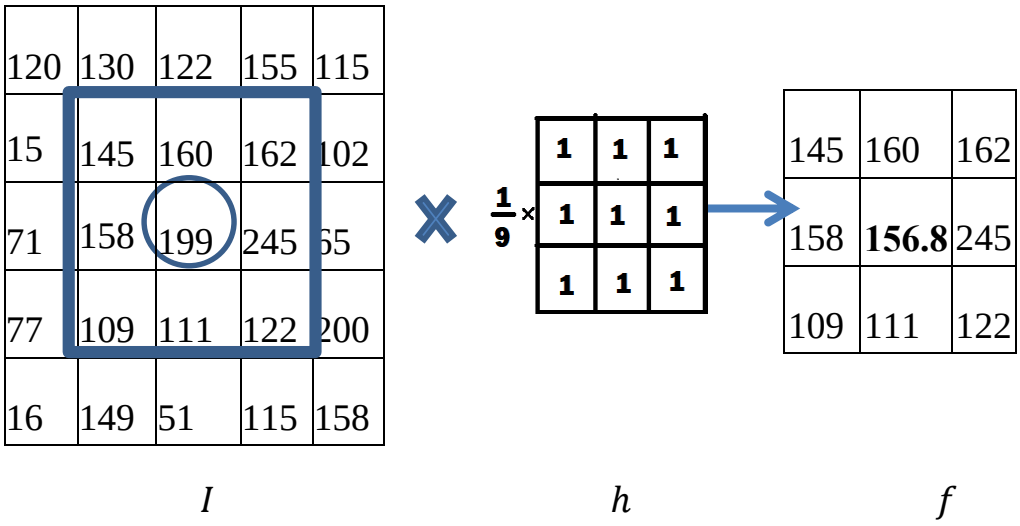


Fig. 3.7: Applying average mean kernel on a pixel.

While gaussian filter uses the following equation to find the kernel in $m \times m$ window

$$G(i,j) = \frac{1}{2\pi\sigma} e^{-\frac{(i^2+j^2)}{2\sigma^2}}$$

For example applying this function on 3×3 window with $\sigma = 1$, we get the following kernel [61]

$$h = \begin{array}{|c|c|c|} \hline 0.059 & 0.097 & 0.059 \\ \hline 0.097 & 0.159 & 0.097 \\ \hline 0.059 & 0.097 & 0.059 \\ \hline \end{array}$$

but MATLAB used the normalized kernel, so each element is divided by the sum of all elements, which gives

$$h^* = \begin{array}{|c|c|c|}\hline 0.075 & 0.124 & 0.075 \\ \hline 0.124 & 0.204 & 0.124 \\ \hline 0.075 & 0.124 & 0.075 \\ \hline \end{array}$$

The following Figure (Fig. 3.8) shows application for 3×3 Gaussian filter.

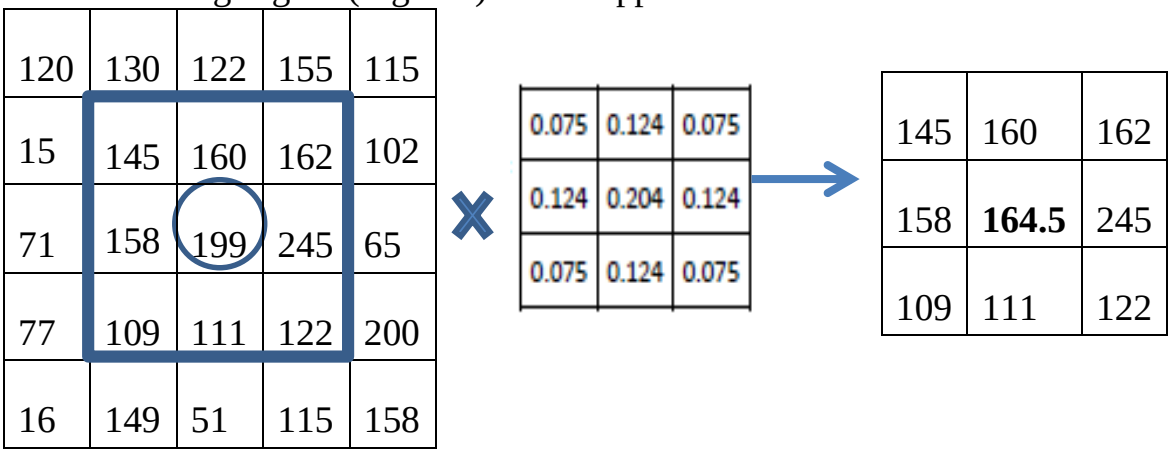


Fig. 3.8: Applying Gaussian kernel on a pixel.

An example of applying mean and Gaussian filters on a colored image is shown in Fig. 3.9.

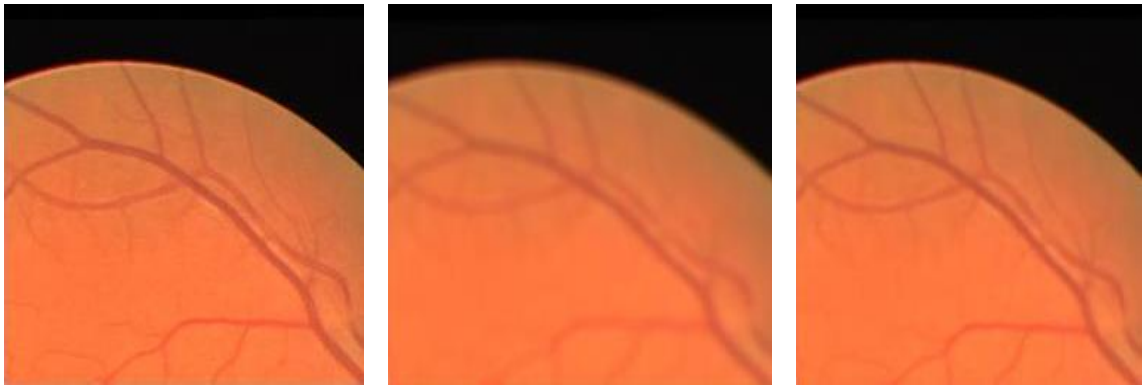


Fig. 3.9: (a) Original image, (b) **7 × 7** mean filter, (c) **7 × 7** Gaussian filter with $\boldsymbol{\sigma = 1}$.

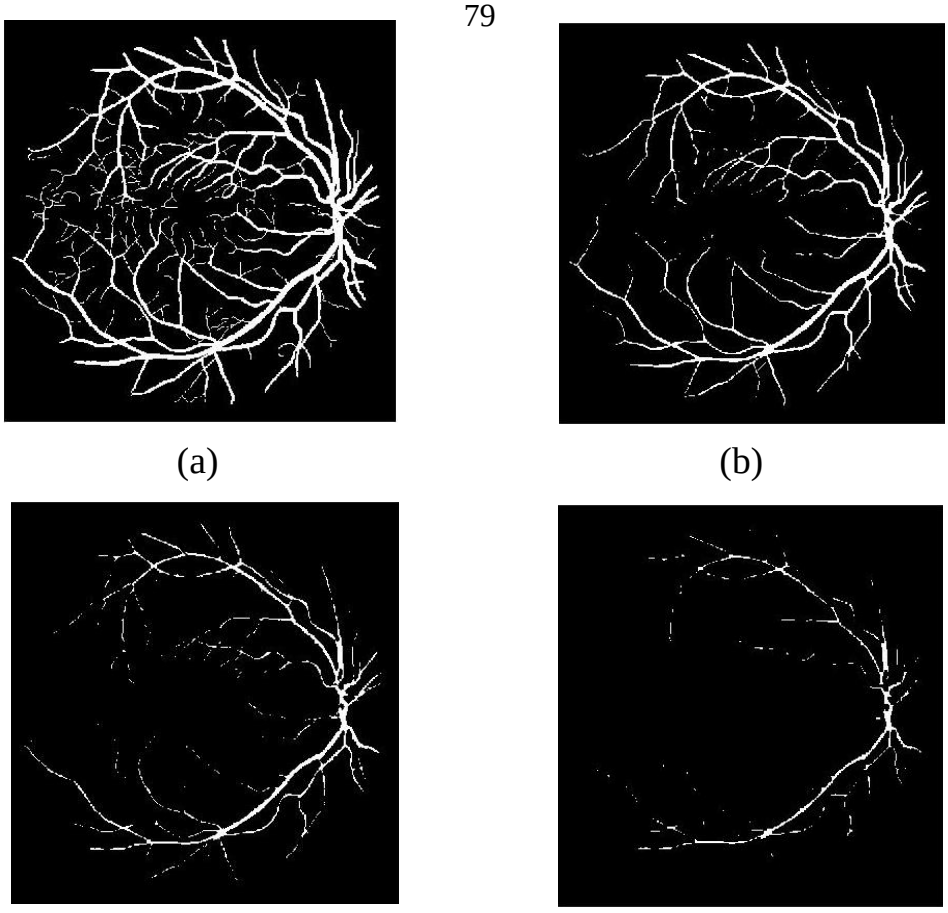
In the third preprocessing step, we remove small vessels as they are not used in the A/V ratio which is an important diagnostic parameter in ophthalmic treatment, moreover it is hard to classify small vessels as artery or vein because both look the same, and they might disturb the classifier. For removing those small vessels we use the morphological operations in MATLAB. The key idea is to use the so called **Structural elements** which can be defined as a binary templet to produce a new image from old one, the center pixel in a structural element is called origin and it determines which pixel in the image will be processed. The MATLAB function "*strel*" is used to define specific structural element, for example:-

se = strel ('disk',1) represents a disk structural element with radius = 1, it looks as a 3×3 matrix

$$se = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

We used three morphological functions in our program which are: *"imerode", "imdilat"* and *"bwareopen"* functions. First we apply erosion to the binary segmented image, it erode the vessels using a specific structural element, Fig. 3.10 shows a segmented image after applying erosion with different structural elements, in this step small vessels disappear or they will be very thin.

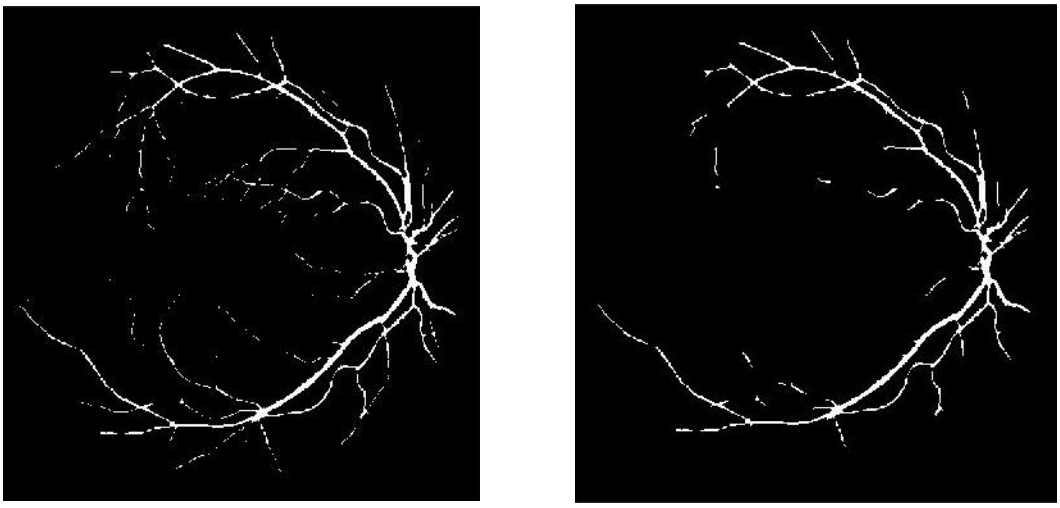
As we see in Fig. 3.10 there are some regions with very small areas which can be considered as noise, so that we need a function to remove those small areas which is *"bwareaopen",* so we remove the areas which is less than 30 pixels, and Fig. 3.11 shows the final results.



(a) (b)

(c) (d)

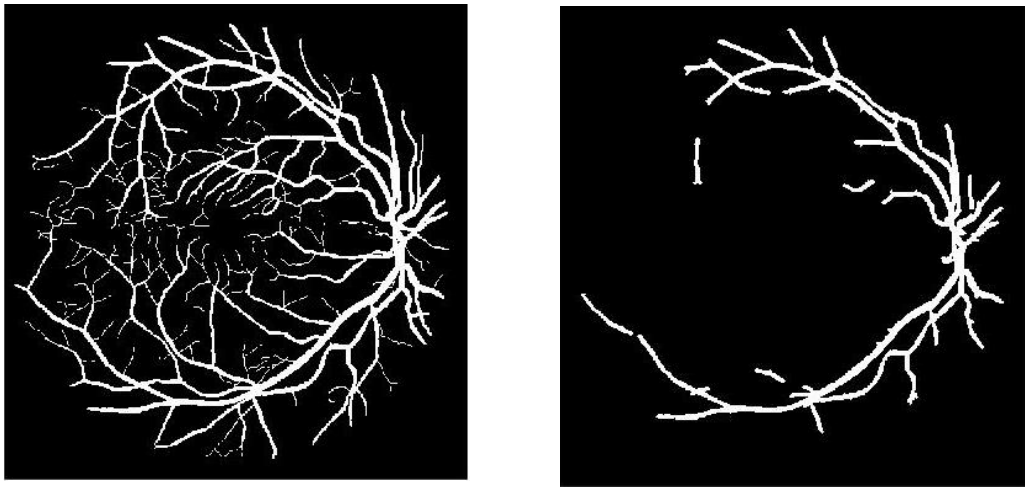
Fig.3.10: (a) Original image, images after erosion with: (b) *strel('disk',1),* (c) *strel('disk',2),* (d) *strel('disk',3).*



(a) (b)

Fig. 3.11: (a) Image after erosion with *se= strel('disk',2)* (b) then *bwareaopen* for areas less than 30.

The resulted image of the previous preprocessing step shows that the vessels become very thin so to return the width as the original image we need to dilate the resulted image with the same structural element. As we see in Fig.3.12 only the four main vessels are kept.



(a) (b)

Fig. 3.12: (a) The original segment image, (b) the image after the morphological processes.

A new technique is implemented here by dividing the input images into three slides, in which the first one includes the upper main vessels, the second one contains the OD region with some small vessels and the third one includes the lower main vessels. The second slide was excluded because it contains the OD region where the vessels are not required and it doesn’t contain fundamental information for diagnostic process.

Histogram matching is implemented as a preprocessing step using different reference images, but this step disturbs the classification; since the retinal images color change with human color, which means no perfect reference.

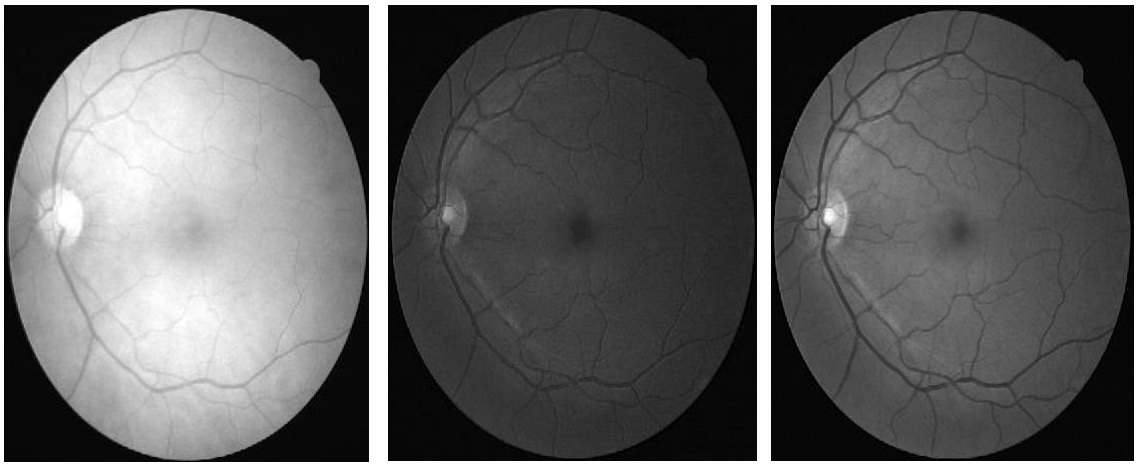
3.3.2 Features

In the next step we need to define the features for each pixel and decide the label for each one in the training stage. The work here was done in two tracks, first read features for main vessels and the other one is to read features for only the center line pixels. Many features can be used; 30 features which are shown in the following table (Table 3.1) are implemented in this work:

Table 3.1: All features tested in the algorithm

f1-f3	red, green and blue
f4	Intensity value
f5-f8	red, green, blue and intensity with histogram equalization
f9	The ratio between red and green channel
f10-f12	Hue, saturation and value without histogram equalization
f13-f15	Hue, saturation and value with histogram equalization
f16-f19	Average mean for red, green, grey and hue
f20-23	Gaussian mean for red, green, grey and hue
f24-f27	variance for red, green, grey and hue
f28-f29	The x, y position
f30	The distance between the pixel and OD

As one can see from Fig. 3.6 each pixel in retinal image represented using three channels, we get information from each channel separately, and form vectors f1, f2 and f3, Fig. 3.13 shows retinal image using red, green and blue channels.



(a) (b) (c)

Fig. 3.13: Retinal image (a) red, (b) blue, (c) green, channels

The true colored retinal images are converted into a gray scale one (intensity image) for f4, as shown in Fig. 3.14. The intensity image shows how dark or bright the pixel should appear in colored image, MATLAB deal with two types of images: *"double"* in which zero shows black pixels, while 1 shows white ones and hence the intensity levels are between 0 and 1. The second type is *"uint8"* in which zero still appear as black pixels while 255 is white, so the intensity levels are between 0 and 255.

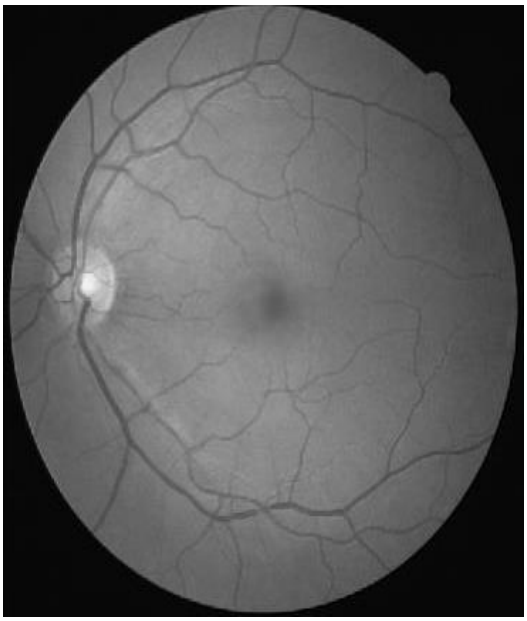
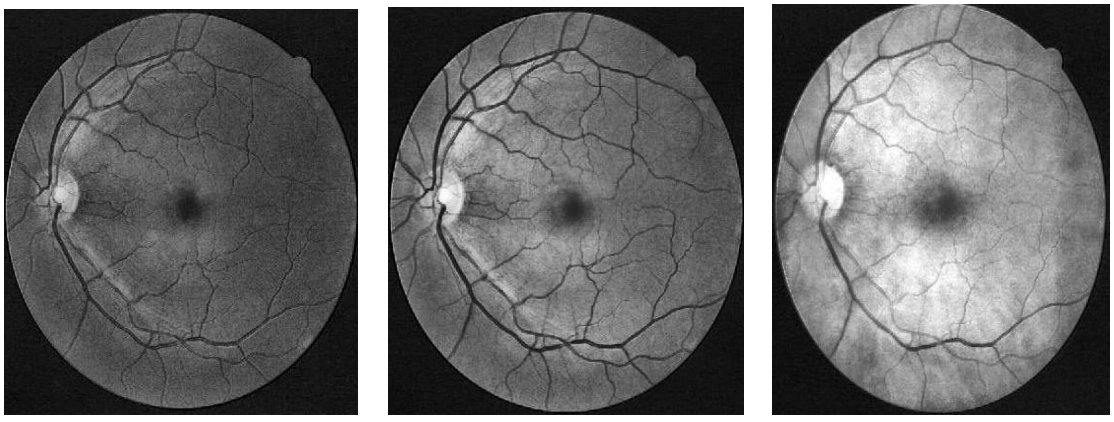


Fig. 3.14: Gray scale (intensity) retinal image

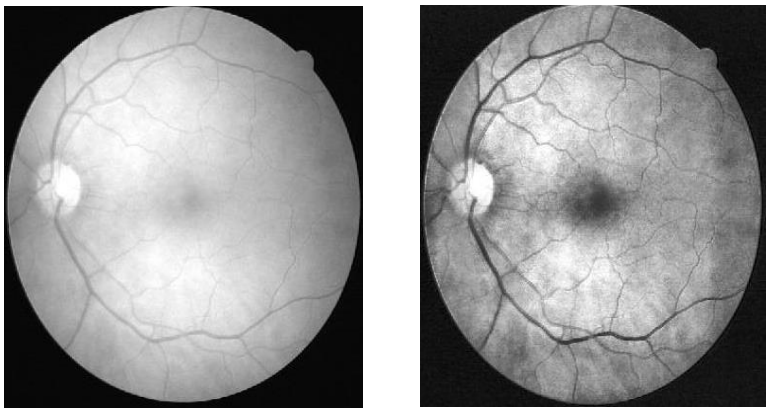
For f5, f6, f7 and f8 adaptive histogram equalization function *"adapthisteq"* was applied which works in small regions called tiles of the image by enhancing the contrast for each region it transforms each value using contrast-limited adaptive histogram equalization (CLAHE), then it uses bilinear interpolation to combine neighboring tiles [61]. Fig. 3.15 shows three channels after applying *"adapthisteq".*



(a) (b) (c)

Fig. 3.15: Apply *"adapthisteq"* for (a) blue, (b) green and (c) red channels

While Fig. 3.16 shows the difference between red color feature and red color after using adaptive histogram equalization, it's clear that second one shows more contrast between vessels.



(a) (b)

Fig. 3.16: Red channel (a) without, (b) with, adaptive histogram equalization.

RGB is consider as a main color space but it isn't the only one, HSV is also a color space which defined by Hue, Saturation and Value, see Fig. 3.17. For each channel in this color space, f10, f11 and f12 are calculated, see Fig. 3.16, then we again used adaptive histogram equalization for each channel.

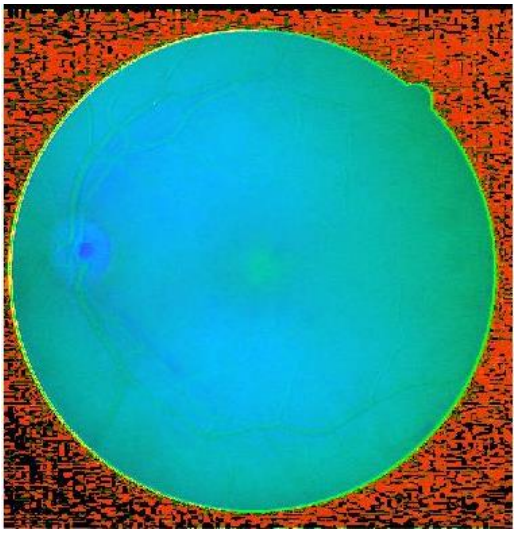
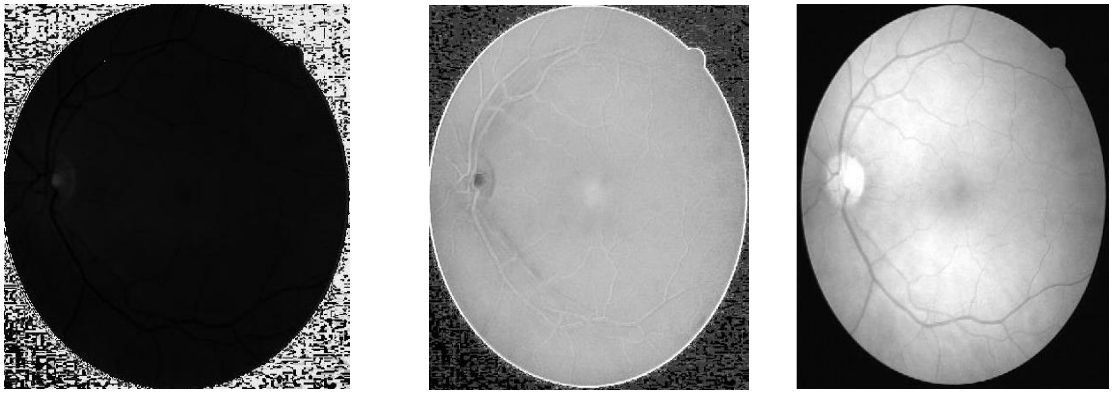


Fig. 3.17: Retinal image in HSV color space.



(a) (b) (c)

Fig. 3.18: Retinal image: (a) Hue (b) saturation and (c) value channels.

We note that value channel in HSV color space is exactly the same as red channel in RGB space, so value channel are excluded in Final feature space matrix.

For the features: mean and variance, we define the parameter n making the window size $2n+1$, then we add $2n$ zeros vector for both rows and columns from each surrounding side, finally we find the mean and variance for the center pixel in window using all members in that window. To find the mean, the following equation is used:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

while to calculate the variance we use the following equation:

$$var = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2$$

where μ is the mean of the window and x_i is i^{th} element.

For the mean features a 3×3 widow is used while for variance the size 9×9 is used.

3.3.3 Feature Extraction (Dimensionality Reduction)

Dimensionality reduction is the heart of our program. In this section we implement two methods for dimensionality reduction which are LDA and PCA. P. Belhumeur et al. used a two-stage linear discriminant method for face recognition [66], we suggest to apply the same method in retinal vessels classification then comparing results using only PCA or LDA, in the two stage method we first reduce the dimension from D to d_1 using PCA, in the next stage LDA is used to reduce the dimension again from d_1 to 1 .

The LDA and PCA algorithms are discussed in chapter 2, these algorithms are implemented in MATLAB to accommodate and use them properly for our method. We prepared a particular program for this task. Optimal $\boldsymbol{w}$ is calculated as its discussed in 2.2 and 2.4, then we multiply both the training and also testing data by it, in order to prepare the data for the classifiers.

3.3.4 Classification

Many classifiers are used in retinal vessels classification; three popular classifiers are chosen to be used in this work, then major voting for these classifiers is consider as final label.

SVM, K-NN and naïve Bayes classifier are chosen to be implemented; chapter 2 gives an introduction to each one, while appendix(C) gives more details.

MATLAB defines these classifiers: for SVM, it prepares or train the classifier by *"fitcliner"* and then it predicts the labels for a testing data

using *"predict"*. K-NN classifier easily uses the command "*knnclassify"*, we enter the training data, predefined labels and testing data. Then we get the labels for the testing data. Finally naïve Bayes are implemented using *"fitcnb"* to train the classifier and *"predict"* commands exactly as in SVM.

K-means which is unsupervised classifier, it doesn't need any training data, we enter the testing data in addition to the number of classes and it gives us the predicted labels for the testing data. K-means fails in retinal vessels classification, since it works randomly, hence it is excluded.

As we said, the output for each classifier is just a vector with predicted outputs (labels), but the image is represented by a matrix, so to display the result we need to build a function in order to convert the vector of labels into matrix by using the testing vessels segmentation to display artery in red and vein in blue color, see Fig. 3.19.

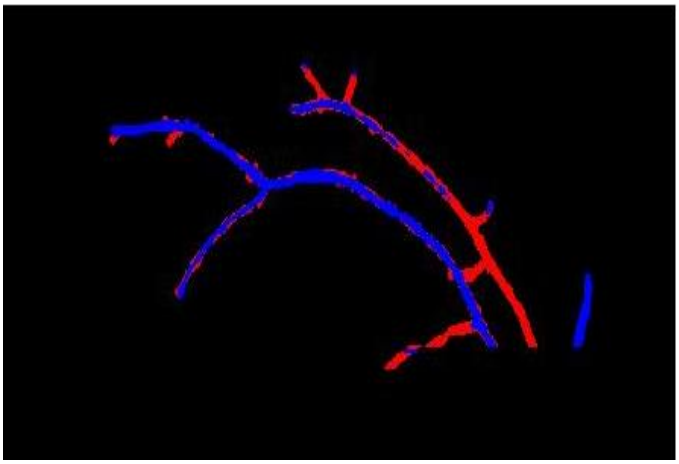


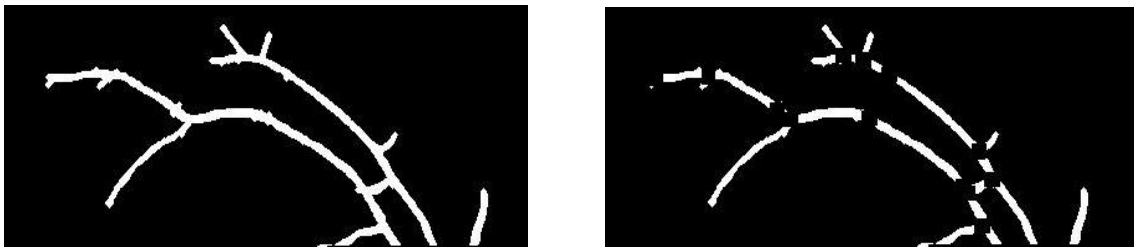
Fig. 3.19: Upper main vessels after classification.

3.3.5 Post-processing

As we see in Fig. 3.19 upper vessels contain some small regions which are classified incorrectly, that affects the performance, but we can avoid this problem by some post-processing steps.

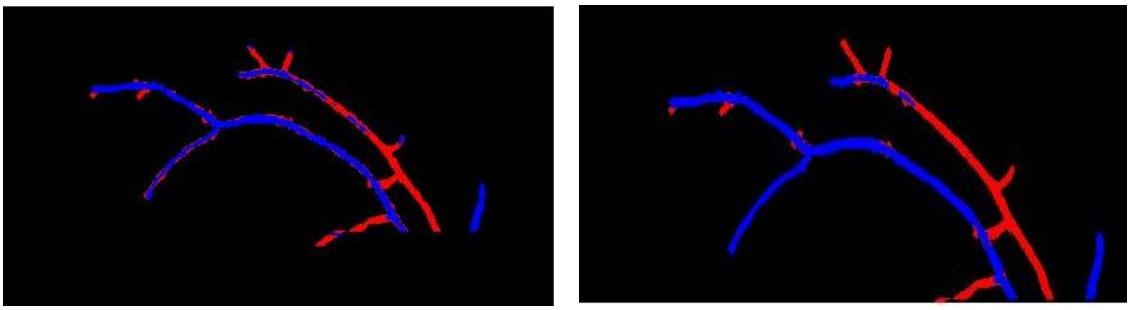
First we try to reclassify the incorrect small regions; this enhancement didn't improve the result as expected. So we decided to apply majority rule in each sub-segment. To implement this technique, we first remove the nodes (bifurcation and crossing points) by defining the skeleton of the graph then the points with four neighborhood are removed, see Fig. 3.20.

Consequently, we calculate how many pixels are classified as artery/vein for each sub-segment. Then the whole sub-segment is considered to be artery if artery's pixels are more than vein's pixels, and vice versa to be vein, Fig. 3.21 shows the result of applying the majority rule.



(a) (b)

Fig. 3.20: (a) Upper vessels, (b) after removing the nodes.



(a) (b)

Fig. 3.21: Classification result (a) before, (b) after applying majority rule.

As we can realize from Fig. 3.21 there are some regions still incorrectly classified because we removed the nods with four neighborhoods in the previous stage.

In the final step, we evaluate the accuracy for each algorithm to decide which one gives the best classification result. In the next chapter we will discuss performance and compare the results.

Chapter Four
Results and Discussion

Chapter Four

Results and Discussion

4.1 Introduction

In this work different methods are presented for dimensionality reduction in addition to classification methods for classifying retinal vessels into artery or vein. This kind of classification seems to be harder to define features that can distinguish between vessels rather than between vessels and non-vessels as in the case of segmentation.

For instance, many properties of vein\artery are similar, one of these properties: the color of one vein in a fundus image may be identical to some artery in different image or even to different region in the same image. Moreover, diseases and age of patient affect vessels' color and shape [55].

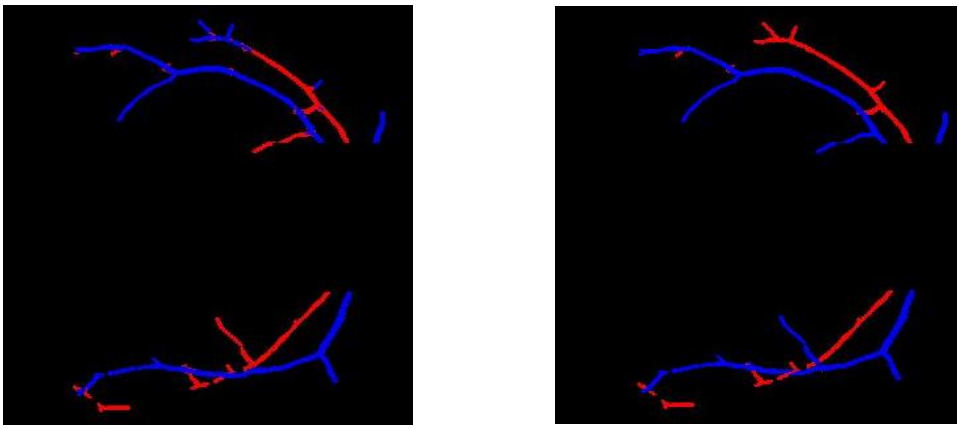
Accuracy calculation is an important step to compare algorithms and to decide which one gives the best way to distinguish between vessels. There are many methods to measure the performance of classification results. M. Divya et al. [10] used a confusion matrix to find the performance for using neural network in retinal vessels classification; they achieved 93% accuracy. K.Rothaus et al. [55] used area under the receiver operating characteristic (ROC). ROC curve is considered as graphical representation for the relation between specificity and sensitivity. In this work, accuracy, specificity and sensitivity are calculated for the different algorithms in

order to compare them and decide which one gives best discriminant between vessels.

4.2 Evaluation Metrics

To calculate the performance of implemented algorithms we compare the result with corresponding ground truth (GT) image, see Fig. 4.1. There are three types of pixels in the classified image: artery, vein and background. We excluded the background since it's not relevant to the performance of the used methodology and only those vessels appearing in the segmented images are considered. The remaining pixels are either artery or vein [22].

To calculate the performance we should first decide which class will be the positive pixel to calculate the accuracy for it. Artery pixels are considered as positive while vein pixels as negative. Sensitivity measures how well the method classifies artery, while specificity measures how well the method classifies vain, where accuracy gives the overall performance for a given method [12].



(a) (b)

Fig. 4.1: (a) The classified image, (b) ground truth image.

Pixels which are classified into artery as in GT are called true positive (TP). However, false positive (FP) represents those pixels which are classified as artery but actually they aren't. True negative (TN) counts those pixels which are not artery in GT and the classifier defines those pixels as in GT. Finally the pixels which are arteries in GT but are classified incorrectly are called false negative (FN) see Table 4.1

Table 4.1: Ground truth table

	Ground truth		
		Artery	Not-Artery
Method result (output)	Artery	TP	FP
	not Artery	FN	TN

Accuracy, sensitivity and specificity are described in the following equations:

$$sensitivity = \frac{\mathrm{TP}}{TP + FN}$$

$$specificity = \frac{TN}{TN + FP}$$

$$accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FN} + \mathrm{FP}}$$

4.3 Comparing Results of the Different Approaches

Through this section different approaches where implemented, actually by changing features and the dimension of the data. This process led us to choose the best performance of the algorithm under the used data base.

The first step, started by studying the result for the classified vessels using all suggested features, we notes the following:

- The SVM, Bayes and KNN classifier's results are vary widely, Figs. (4.2-3-4) illustrates sensitivity, specificity and accuracy for all testing images respectively. The x-axis represents the testing image number where the rest images are the training ones.
- The average sensitivity for implemented classifiers (SVM, Bayes and K-NN) are 78.9 %, 67.8 % and 49.3 %, while the average specificity are 90.3 %, 70.3 % and 80.3 %, and the average accuracy are 86.5 %, 71.1 % and 69.7 %, respectively.
- By comparing results of classification for the different classifiers on the used images, we can notes that inconsistent results influenced by images properties and the classifier abilities to work properly. This kind of results didn't prefer a particular classifier to be used.
- The sensitivity, specificity and accuracy after using major voting for three classifiers are 69.6 %, 85.8 % and 81.1 % respectively, which shows better results rather than the average of the three classifiers.
- The program took roughly 700 second to classify vessels using all features, our used machine is an hp i3 laptop with 4GB RAM.

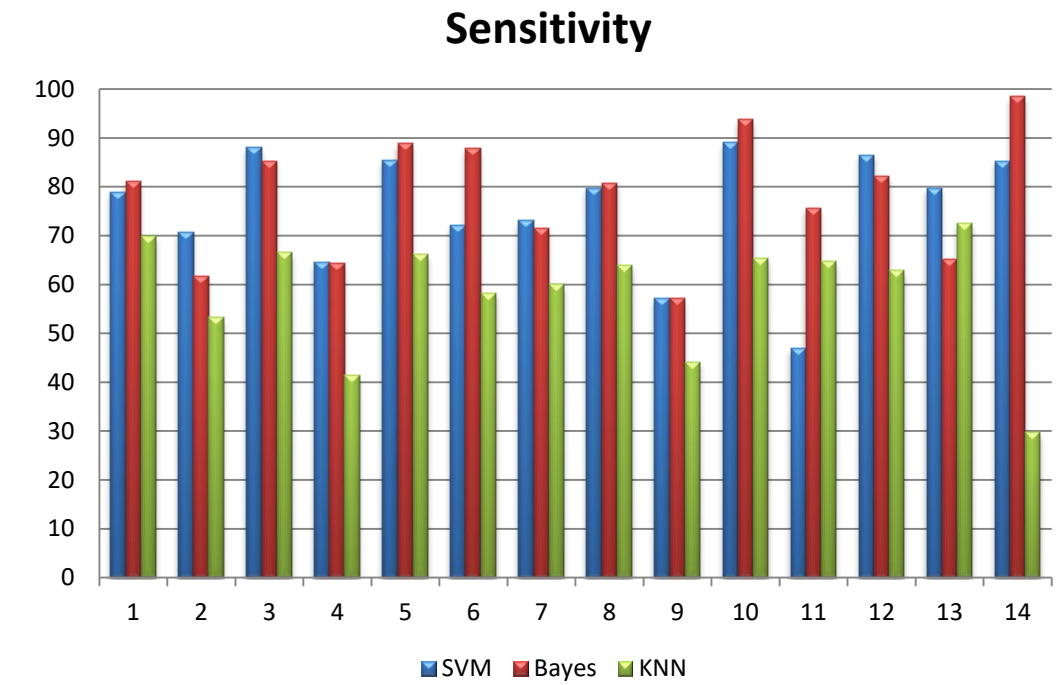


Fig. 4.2: Sensitivity rate for all testing images using the three classifiers before reduction.

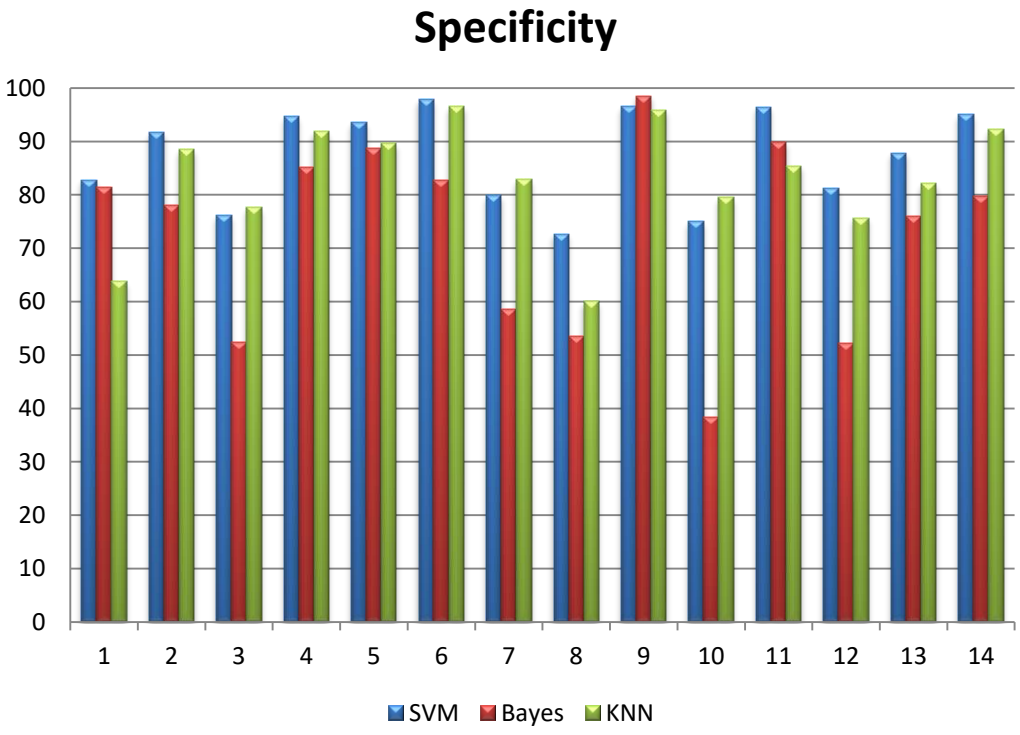


Fig. 4.3: Specificity rate for all testing images using the three classifiers before reduction.

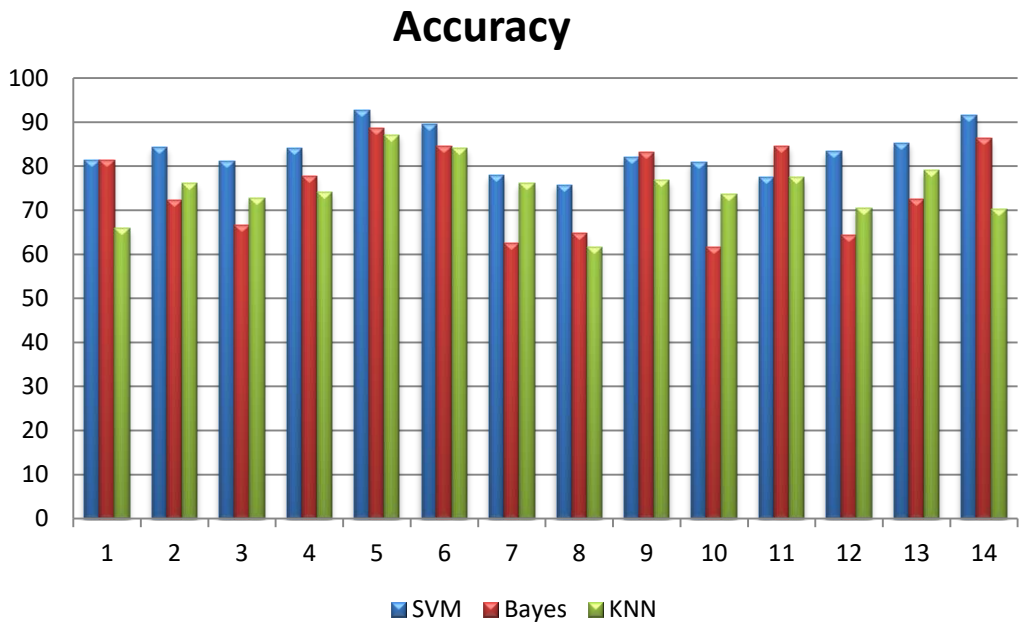


Fig. 4.4: Accuracy rate for all testing images using the three classifiers before reduction.

In the next step, LDA is implemented to reduce the dimension from 27 to one-dimensional space. As it's known LDA reduce time of processing, and for our program the time is reduced from 700 sec to 335 sec.

After applying the LDA, the performance of the three classifiers (SVM, Bayes and LDA) came to be similar. This is a good advantage that makes any of these classifiers suitable to be used by itself, which is a kind of saving resources as shown in Figs. (4.5-6-7). The classifiers after applying LDA improved their performance approximately to the best classifier. This reduction is very important as the LDA manages to estimate the effect of the best features and disregarding those with less influence or confusing ones.

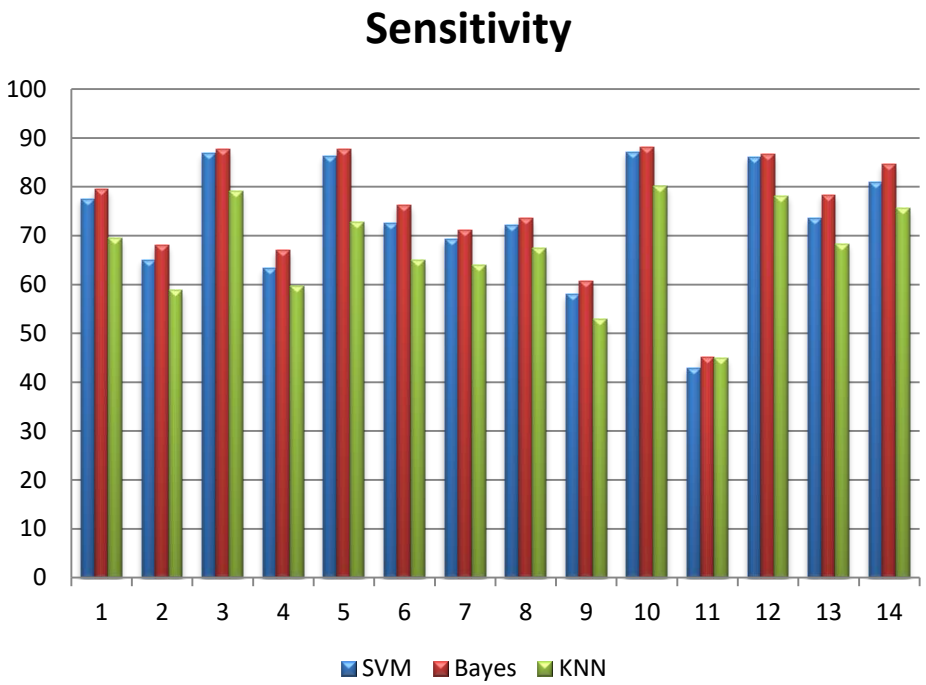


Fig. 4.5: Sensitivity rate for all testing images using the three classifiers after implement LDA.

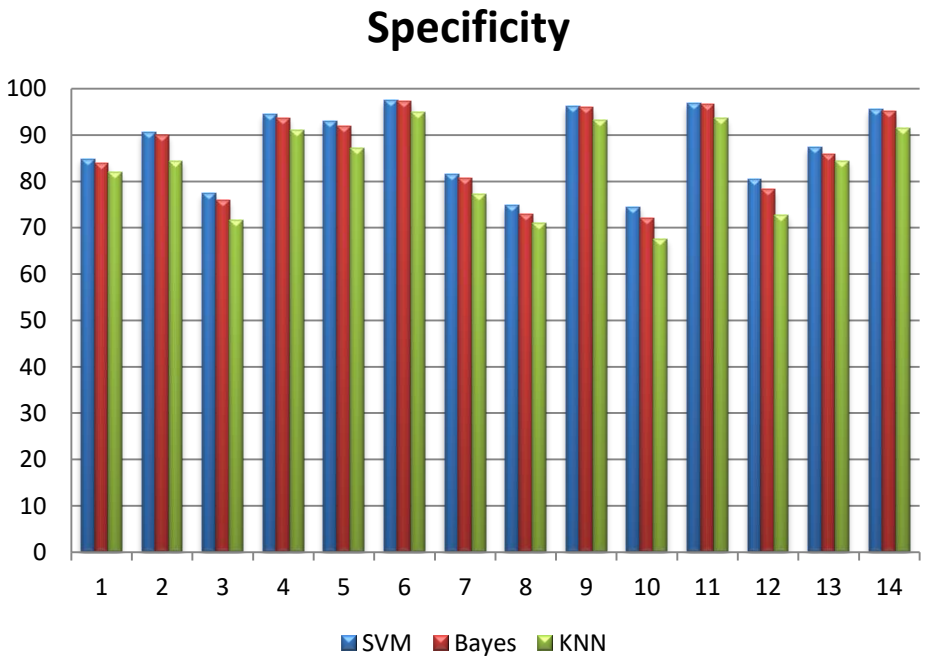


Fig. 4.6: Specificity rate for all testing images using the three classifiers after implement LDA.

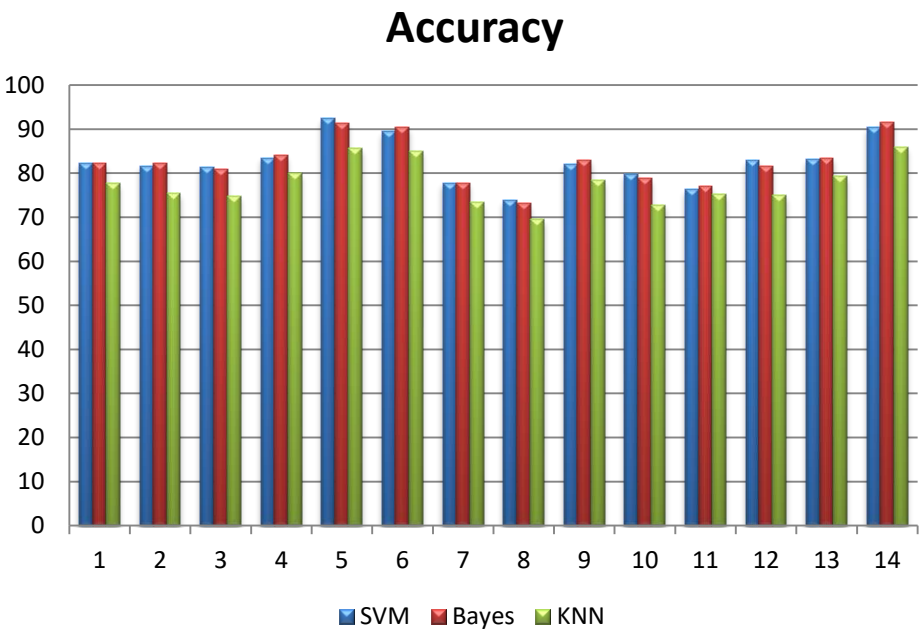


Fig. 4.7: Accuracy rate for all testing images using the three classifiers after implement LDA.

The sensitivity, specificity and accuracy after applying LDA and using the three classifiers are represented in the following table (Table 4.2)

Table 4.2: The average performance for the classifiers after implementing LDA.

	sensitivity	specificity	accuracy
SVM	78.2 %	89.6 %	85.5 %
Bayes	77.4 %	90 %	85.4 %
KNN	71.3 %	85.5 %	80.4 %

After applying the majority voting technique for the three classifiers, the average metrics came to be: sensitivity 77.9 %, specificity 89.8 % and accuracy 85.4 %. Comparing averages with and without LDA, all of sensitivity, specificity and accuracy show improvement in the classification result, noticing that accuracy is more important for evaluating the method.

LDA keeps and improves the performance while using only one dimension, moreover this reduces the process time to the half.

On the other hand, PCA is used to reduce the number of features; we first tested to reduce the features into one dimensional space. We found that PCA failed to reduce the dimension since we get the average sensitivity equal zero. Hence, we try to reduce the dimension ranging from two up to 20 to find the optimal reduced dimension for using PCA, see Table 4.3.

Table 4.3: Performance after applying PCA for different dimensions with majority voting technique.

Dimension	Sensitivity	Specificity	Accuracy
1	0 %	100 %	62 %
2	25.8 %	90.6 %	68.6 %
3	31.8 %	89.2 %	69.4 %
4	46.1 %	84.9 %	72 %
5	54.9 %	86.8 %	75.7 %
6	56.3 %	86.3 %	76 %
7	58.6 %	87.8 %	77.4 %
8	62.9 %	89.1 %	80 %
9	67.4 %	89 %	81.4%
10	68.4 %	90.1 %	82.6%
11	69.6 %	90.3 %	83.2 %
12	70.3 %	90.3 %	83.4 %
13	71.8 %	90.0 %	83.8 %
14	72.2 %	89.7 %	83.7 %
15	71.9 %	89.6 %	83.6 %
16	73.5%	89.6 %	84 %
17	73.7%	89.6 %	84 %
18	72.6 %	89.6 %	83.6 %
19	73.4 %	88.9 %	83.4%
20	73.6 %	88.6 %	83.3 %

It’s clear from the table that as the number of dimension approaches 17 the accuracy gets higher, and as we search for the best performance, it is preferable to reduce the dimension to 17. Also as in the case of the LDA, the PCA save processing time, if we reduce the dimension from 27 to 17, the time needed to classify all the testing images is 584.8 sec, which is less than using all features.

It’s clear that using LDA to reduce the dimension into one dimensional space is better and takes less time than using PCA to reduce the dimension to 17 dimensional space.

We suggested an approach for improving the A/V classification result by reducing the dimension into one dimensional space using two-stage method. PCA is implemented in the first stage to reduce the number of features from 27 to 17; this will avoid those features which contain in-consistent data in addition to keep those which are more relevant to A/V classification. In the next stage, LDA is applied to reduce the PCA outputs into one-dimensional space. Finally, we again apply the classifiers (SVM, Bayes and KNN) and the results are summarized in Table 4.4. The Majority voting technique is applied, giving average sensitivity, specificity and accuracy as 77.3 %, 89.3 % and 84.9 % respectively.

Table 4.4: The average performance of the classifiers after implementing two-stage method.

	sensitivity	specificity	Accuracy
SVM	77.5 %	89.2 %	84.9 %
Bayes	77.1%	89.4 %	84.9 %
KNN	70.4%	85 %	79.8 %

From data shown in Figs. (4.8-9-10) the output of the suggested approach (Two-stage) came to be slightly better than all previous results in this work.

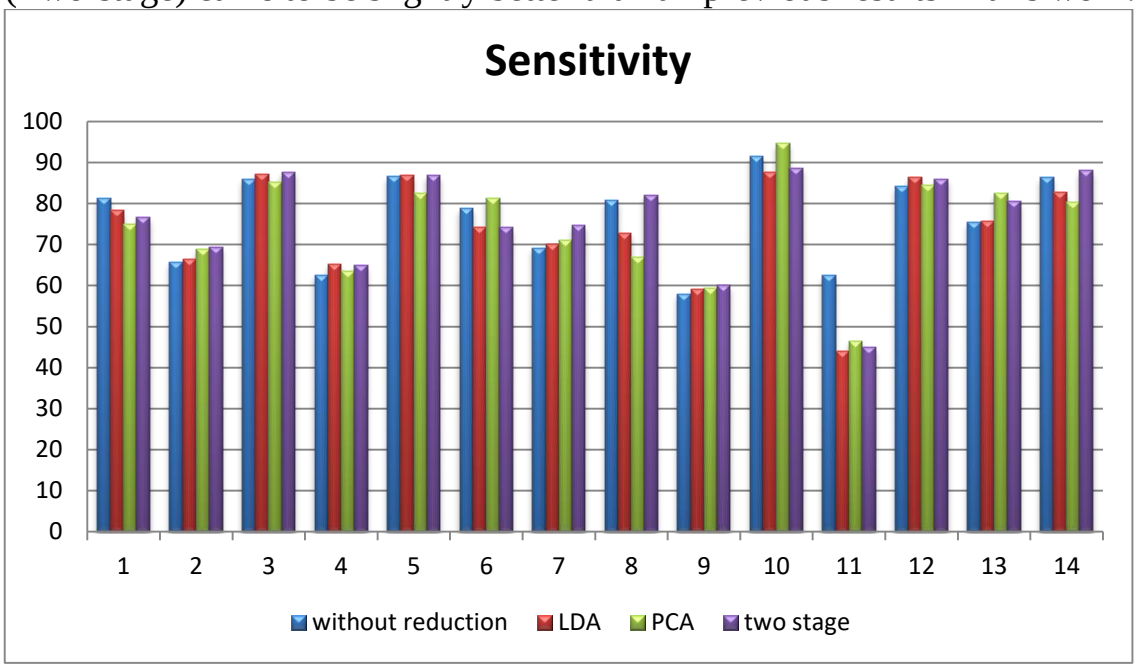


Fig. 4.8: A comparison between sensitivity rates for the reduction techniques under the majority voting.

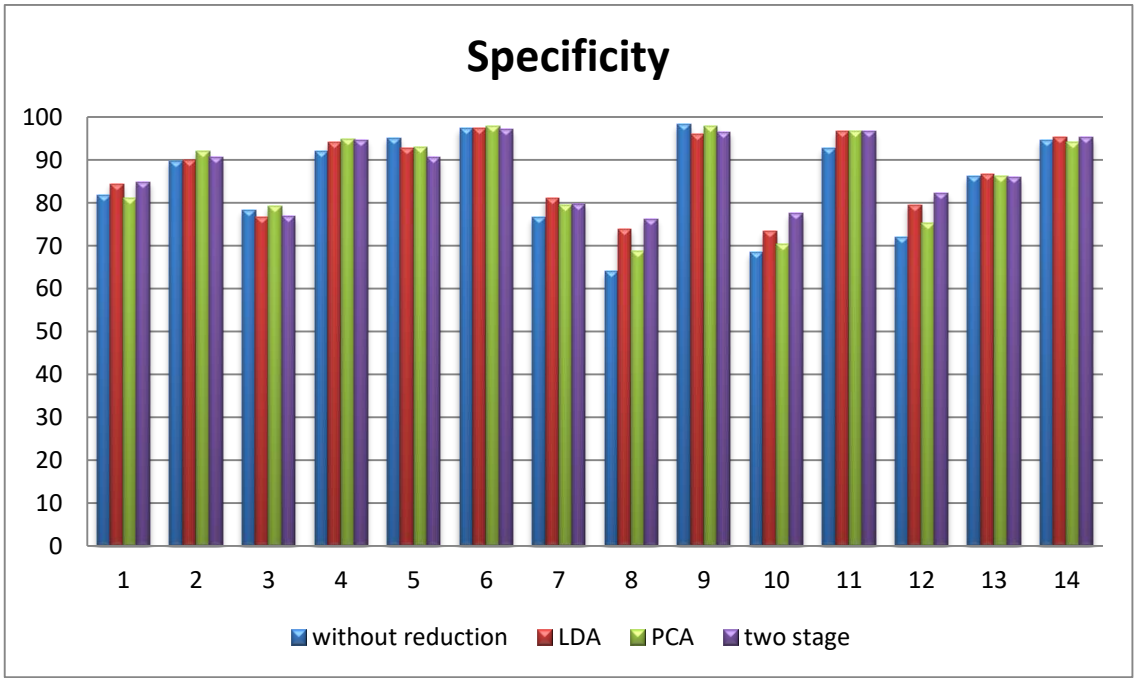


Fig. 4.9: A comparison between specificity rates for the reduction techniques under the majority voting.

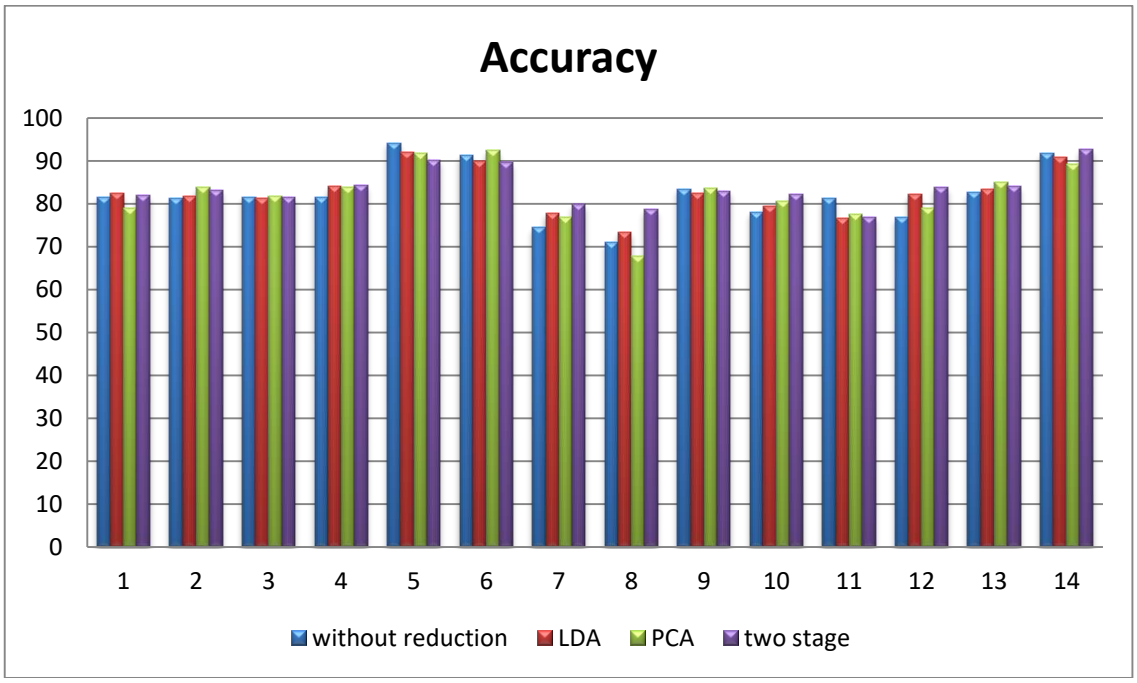


Fig. 4.10: A comparison between accuracy rates for the reduction techniques under the majority voting.

As we mention previously, the other approach using the center line pixels to train and classify the vessels was also implemented. The three dimensionality reduction techniques (LDA, PCA and Two-stage) are implemented to reduce the dimension as we explained previously, then we apply majority voting technique for the three classifiers (SVM, Bayes and KNN), Table 4.5 represent the result from this process.

Table 4.5: Performance of applying the reduction techniques to classify center line pixels.

	sensitivity	specificity	Accuracy
Without reduction	71%	84.6%	80.6%
LDA	75.7%	88.3%	83.7%
PCA	70%	85.5%	80.3%
Two-stage (LDA+PCA)	72.8%	87.1%	82%

We note that, if we use the data of whole vessels it gives better performance rather than using the center line pixels only. Since the data used is very small with respect to the first case (whole vessels pixels) so the required time for process is reduced remarkably, it takes approximatily 250 sec.

A structural enhancement method (majority rule) is implemented after reducing the dimension and applying the classifiers. We divide the main vessel into sub-vessels and apply the majority rule to enhance our classification result. Fig. 4.11 shows the final classification result after we train the classifiers using 13 training image then applying the major voting for suggested classifiers, and finally implementing majority rule technique for each sub-segment. As we expect, sensitivity, specificity and accuracy are improved. The following Table 4.6 shows the result after apply majority rule on the average classifiers results. While Figs. (4.12-13-14) show how results are enhanced after applying the majority rule.

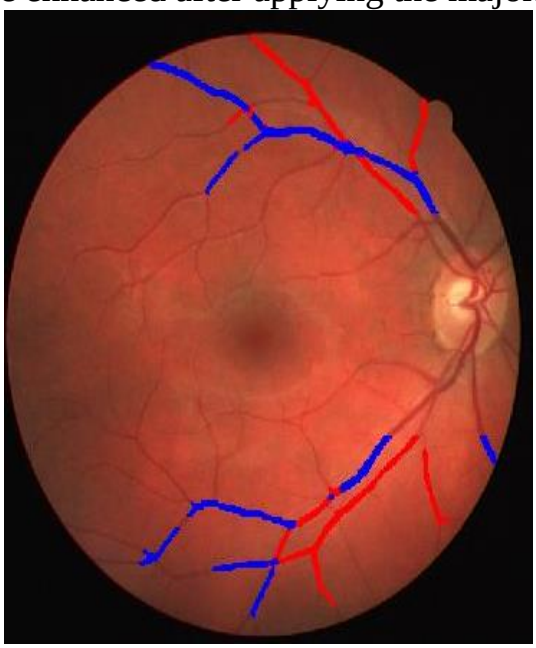


Fig. 4.11: Result of our algorithm.

Table 4.6: Performance of the reduction techniques after apply majority rule.

	sensitivity	specificity	accuracy
Without reduction	72.2 %	89.0 %	83.6 %
LDA	81.6 %	92.2 %	88.5 %
PCA	78.7%	92.2%	87.4%
LDA+PCA	80.8 %	91.9 %	87.4 %

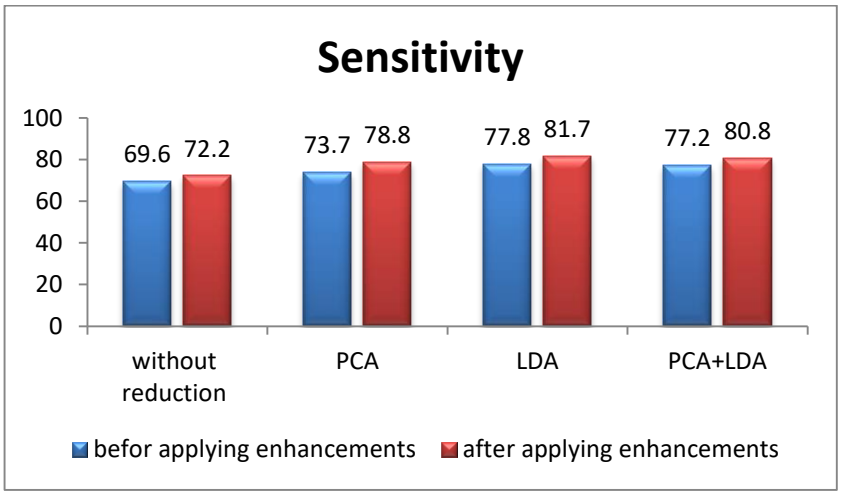


Fig. 4.12: Sensitivity rate for the reduction technique before and after applying enhancement.

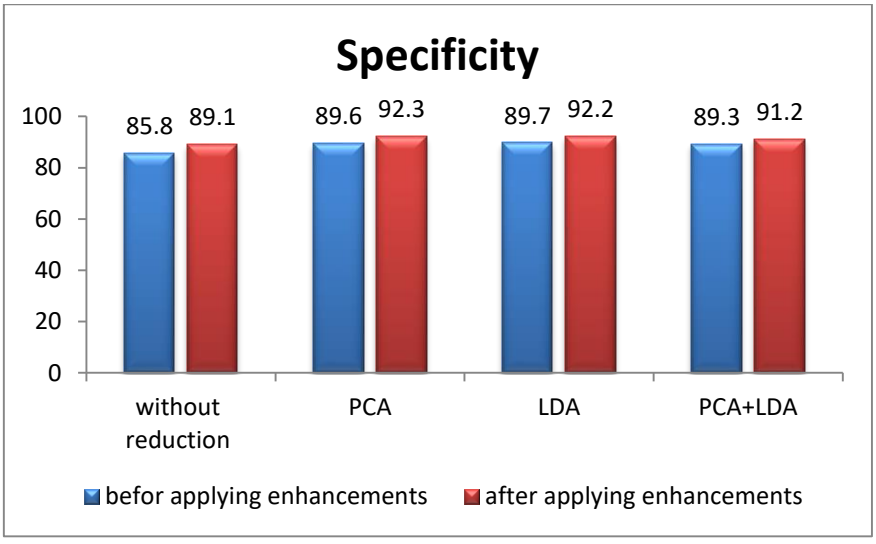


Fig. 4.13: Specificity rate for the reduction technique before and after applying enhancement.

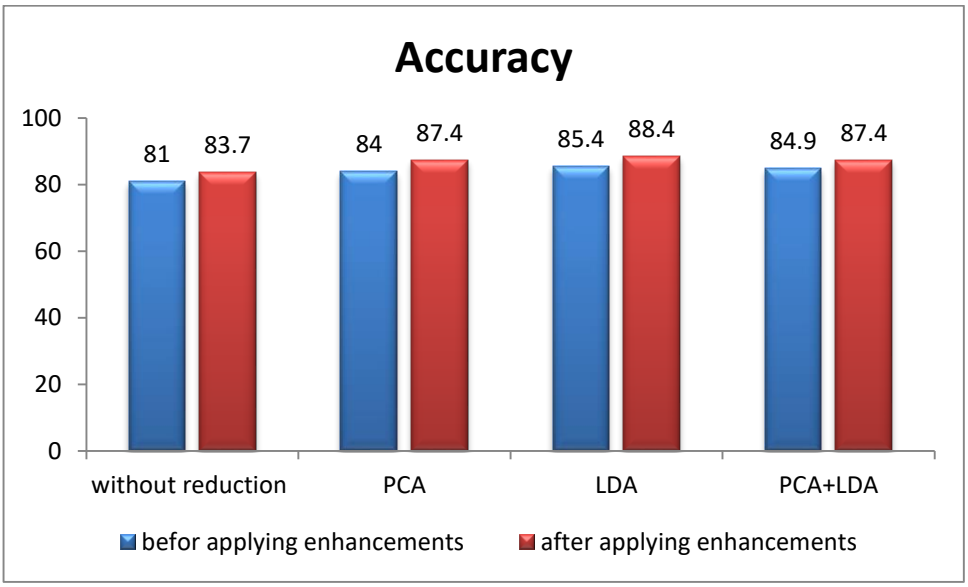


Fig. 4.14: Accuracy rate for the reduction technique before and after applying enhancement.

4.4 Difficulties Facing the Methodology.

Many difficulties facing our work; the following points are some of these difficulties:

1- The available GT for the used database are 20 images, 7 of them are infected and can't be used in training or testing steps.

2- The nature of fundus images are not very sharp and clear which affect the ability to get accurate features and hence affect the classification.

3- Vessels themselves are very similar with high variability in different images. Also Small vessels are very similar; so it's not feasible to classify them especially without structural knowledge. So that we are interested in classifying only the main vessels.

4- We tried to use the structure of the vessels but unfortunately the available segmentation is not good enough to distinguish between meeting points, which lead to high misclassification error.

5- Fundus images of different people are with different colors, which means no standard features.

4.5 Conclusion and Future Work.

In this work we propose a method for classifying retinal vessels into artery\vein, 13 retinal images are used to train the classifiers and the last image is used for testing the algorithm. Features are chosen carefully then three dimensionality reduction techniques are implemented. After we reduce the number of features, three classifiers are executed and major voting for these classifiers is considered as the final result.

As we know dimensionality reduction techniques keep the more effective features, expecting to improve the performance of the classifiers to separate the data in a good way. We note that the proposed two-stage method reduces the dimension of data from 27 to one-dimension consequently reduces the process time. Where in one dimensional space all three classifier are working very similar to each other. The accuracy results for the three classifiers (SVM, Bayes and K-NN) after applying the Two-stage method are 87.3 %, 87.5 % and 87.1 respectively. Also the performance is enhanced after reducing the dimension as Table 4.7 shows.

Table 4.7: Classification performance comparison: using all features and after reducing the dimension by two-stage method.

	sensitivity	specificity	accuracy
Without reduction	72.2 %	89.1 %	83.7 %
Two-stage method	80.8%	91.2 %	87.4 %

Also we notice that PCA failed to reduce the dimension into one-dimensional space, while LDA managed to do so.

LDA shows the best classification accuracy result 88.5 %, where the Two-stage method shows better stability in classifiers results as Table 4.8 shows.

Table 4.8: Performance comparison between using LDA and Two-stage method

	Using LDA method			Using Two-stage method		
	sensitivity	specificity	accuracy	sensitivity	specificity	accuracy
SVM	82.4 %	91.7 %	88.4 %	80.8 %	91.2 %	87.4 %
Bayes	69.1 %	71.8 %	72.6 %	80.7 %	91.4 %	87.5 %
K-NN	81.1 %	91.4 %	87.7 %	80.1 %	91.1 %	87.1 %

We plan to improve our suggested methodology to classify vessels in fundus images, through writing a new version of our program in MATLAB. This can be done first, by getting additional data with GT since the available database is limited. Secondly, we might improve other classifiers such as Neural Network (NN) and Fuzzy clustering to improve our work.

The preprocessing steps are also included to be improved in our future work, especially removing small vessels technique since in some images a main part of vessel are excluded, on the other side, non-important information are kept.

Finally, we plan to add more features for our algorithm, especially structural features such as the graph analysis and vessels width. Also new color space can be added for instance (LAB).

We are looking forward to use our program in ophthalmic section in the Al-Najah National university hospital in the near future.

References

[1] M. Abramoff and M. Garvin, **Retinal Imaging and Image Analysis,** IEEE Trans Med Imaging, Vol. 3, pp. 169-208, 2010.

[2] B. Aliahmad, **Novel Methods in Retinal Vessel Caliber Feature Extraction for Systemic Disease Assessment**, School of Electrical and Computer Engineering, College of Science, Engineering and Health, RMIT University, 2014.

[3] M. Al-Jazzar, **A Comparative Study Between Linear Discriminant Analysis and Multinomial Logistic Regression in Classification and Predictive Modeling**, Master thesis, AL Azhar University Gaza, 2012.

[4] E. Alpaydm, **Introduction to Machine Learning,** The MIT Press, Cambridge, Massachusetts, London, 2004.

[5] B. Amit, V. Prakash and B. Muna **Techniques of Fundus Imaging**, Sci J Med & Vis Res Foun, Vol. 2, pp.100–107, 2015.

[6] S. Balakrishnama and A. Ganapathiraju, **Linear Discriminant Analysis - A brief Tutorial**, Institute for Signal and Information Processing, Mississippi State University, MS State, MS, USA, March 2, 1998.

[7] S. Balakrishnama, A. Ganapathiraju and J. Picone, **Linear Discriminant Analysis for Signal Processing Problems**, DOI: 10.1109/SECON.1999.766096, IEEE, February, 1999.

[8] P. Belhumeur, J. Hespanha and D. Kriegman, **Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection**,

IEEE Trans, Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 711-720, July, 1997.

[9] K. Divya and D. Anita Sofia, **Classification of Artery/Vein in Retinal Images using Graph-based Approach**, International Journal of Advanced Research Trends in Engineering and Technology (IJARTET), Vol. 1, Issue 2, October, 2014.

[10] K. Divya, C. Senthamarai, D. Chitra, **An Approach to Classification of Retinal Vessels Using Neural Network Pattern Recoginition,** International Journal of Modern Trends in Engineering and Research (IJMTER), Vol. 02, Issue. 07,July, 2015

[11] R. Duda, P. Hart and D. Stork, **Pattern Classification**, John Wiley & Sons Inc, 2nd ed., 2001.

[12] R. Estrada, M. Allingham and P. Mettu, **Retinal Artery-Vein Classification via Topology Estimation**, IEEE Transactions on Medical Imaging, Vol. 34, Issue. 12, pp. 2518 – 2534, December, 2015.

[13] M. Favali, S. Abbasi-Sureshjani, B. ter Haar Romenij, and A. Sarti, **Analysis of Vessel Connectivities in Retinal Images by Cortically Inspired Spectral Clustering,** Journal of Mathematical Imaging and Vision, Vol. 56, issue. 1, pp. 158-172. September, 2016.

[14] W. Ford, **Numerical Linear Algebra with Applications Using MATLAB**, 1st ed, Academic Press is an imprint of Elsevier, USA, 2015.

[15] R. Goel, V. Kumar, S. Srivastava and A. Sinha, **A Review of Feature Extraction Techniques for Image Analysis,** International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), Vol. 6, Issue. 2, pp. 153-155, February, 2017.

[16] G. Golub and C. VanLoan, **Matrix Computations**, 4th ed. Baltimore, The Johns Hopkins Univ. Press, 2013.

[19] E. Grisan, A. Ruggeri, **A Divide et Impera Strategy for Automatic Classification of Retinal Vessels into Arteries and Veins**, Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol. 1, p. 890, September 17, 2003.

[20] H. Hamid, **A New Approach for Classifying Large Number of Mixed Variables**, World Academy of Science, Engineering and Technology, International Journal of Mathematical and Computational Sciences, Vol. 4, No. 10, 2010.

[21] N. Hatami and M. Goldbaum, **Automatic Identification of Retinal Arteries and Veins in Fundus Images Using Local Binary Patterns**, nvestigative Ophthalmology and Visual Science, Vol. 55, Issue. 5, p. 232, 2014.

[22] Y. Hou, **Automatic Segmentation of Retinal Blood Vessels Based on Improved Multiscale Line Detection**, Journal of Computing Science and Engineering (JCSE), Vol. 8, No. 2, pp. 119-128, June 2014.

[23] P. Howland and H. Park, **Two-stage Methods for Linear Discriminant Analysis: Equivalent Results at a Lower Cost**, Technical report, Georgia Institute of Technology, 2009.

[24] M. Jabbar, B. Deekshatulu and P. Chandra**, Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm**, International Conference on Computational Intelligence: Modeling Techniques and applications, doi: 10.1016/j.protcy.2013.12.340, Vol. 10, pp. 85-94, 2013.

[25] H. Jelinek, C. Depardieu, C. Lucas, D. Cornforth, W. Huang and M. Cree, **Towards Vessel Characterization in the Vicinity of the Optic Disc in Digital Retinal Images,** Image Vis Comput Conf., pp. 2-7, November 28, 2005.

[26] Y. Jieping, J. Shuiwang, **Discriminant Analysis for Dimensionality Reduction: An Overview of Recent Developments**, DOI: 10.1002/9780470522356.ch1, Biometrics: Theory, Methods, and Applications, pp. 1-19, 2009.

[27] V. Joshi, J. Reinhardt, M. Garvin and M. Abramoff, **Automated Method for Identification and Artery-Venous Classification of Vessel Trees in Retinal Vessel Networks**, PLoS One, Vol. 9, Issue 2, February, 2014.

[28] A. Joussen, T. Gardner, B. Kirchhof and S. Ryan, **Retinal Vascular Disease**, Springer-Verlag Berlin Heidelberg, Germany, 2007.

[29] S. Khan and A. Ahmad, **Cluster Center Initialization Algorithm for K-Means Clustering**, Pattern Recognition Letters, Vol. 25, DOI:10.1016/j.patrec.2004.04.007, pp. 1293-1302, March, 2004.

[30] N. Khosla, **Dimensionality Reduction Using Factor Analysis**, dissertation, Griffith University, Australia, December, 2004.

[31] J. Kim, B. Kim and S. Savarese, **Comparing Image Classification Methods: K-Nearest-Neighbor and Support-Vector-Machines**, World Scientific and Engineering Academy and Society (WSEAS) Stevens Point, Wisconsin, USA, pp. 133-138, 2012.

[32] H. Kolb, E. Fernandez and R. Nelson, **Simple Anatomy of the Retina Webvision: The Organization of the Retina and Visual System,** Salt Lake City (UT): University of Utah Health Sciences Center, 1995.

[33] C. Kondermann, D. Kondermann and M. Yan, **Blood Vessel Classification into Arteries and Veins in Retinal Images**, Medical Imaging, International Society for Optics and Photonics, Vol. 6512, p. 651247, March, 2007.

[34] G. Kumar and P. Kumar Bhatia, **A Detailed Review of Feature Extraction in Image Processing System**, International conference on advanced computing and communication Technologies, IEEE, pp. 5-12, 2014.

[35] K. Lee, **A Multi-class Classification using Ensembles of Multinomial Logistic Regression Models**, dissertation, The Graduate School, Stony Brook University: Stony Brook, NY, 2012.

[36] H. Li, W. Hsu, ML. Lee and H. Wang , **A piecewise Gaussian Model for Profiling and Differentiating Retinal Vessels**, International Conference on Image Processing (ICIP), IEEE, Vol. 1, p. 1-1069, September 14, 2003.

[37] J. Lowell, **Automated Retinal Analysis**, theses, Durham University, 2006. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/2887/.

[38] R. MacAusland, **The Moore-Penrose Inverse and Least Squares**, University of Puget Sound, Washington, April, 2014.

[39] A. Maharjan, **Blood Vessel Segmentation from Retinal Images**, Master's Thesis, University of Eastern Finland, 2016.

[40] S. Maheswari and S. Anandhi, **Artery Vein Classification of Blood Vessels in Retinal Image: An Automated Approach**, International Journal of Advance Research in Science and Engineering (IJARSE), Vol. 4, Issue. 02, pp. 71-77, March, 2015.

[41] S. Marsland, **Machine Learning an Algorithmic Perspective**, Campman and hall, CRC, Taylor and Francis group, New York, 2009.

[42] M. Miri, Z. Amini, H. Rabbani and R. Kafieh, **A Comprehensive Study of Retinal Vessel Classification Methods in Fundus Images**, Journal of Medical Signals & Sensors, Vol. 7, Issue. 2, Pp. 59-70, April-June 2017.

[43] Q. Mirsharif, F. Tajeripour and H. Pourreza, **Automated Characterization of Blood Vessels as Arteries and Veins in Retinal**

Images, Comput Med Imaging Graph, Vol. 37, Issue. 7-8, pp. 607-617, October-December, 2013.

[44] J. Moini, **Phlebotomy: Principles and Practice**, Jones and Bartlett learning, LCC, an Ascend Learning Company, Chapter 5, pp: 35-51, 2013.

[45] C. Muramatsu, Y. Hatanaka, T. Iwase, T. Hara and H. Fujita, **Automated Selection of Major Arteries and Veins for Measurement of Arteriolarto-Venular Diameter Ratio on Retinal Fundus Images**, Computerized Medical Imaging and Graphics, Vol. 35, Issue. 6, September 2011.

[46] C. Muramatsu, Y. Hatanaka, T. Iwase, T. Hara and H. Fujita, **Automated Detection and Classification of Major Retinal Vessels for Determination of Diameter Ratio of Arteries and Veins**, SPIE Medical Imaging, International Society for Optics and Photonics, p. 76240J, March 4, 2010.

[47] H. Narasimha-Iyer, J. Beach, B. Khoobehi and B. Roysam, **Automatic Identification of Retinal Arteries and Veins from Dual-Wavelength Images Using Structural and Functional Features.** IEEE Trans Biomed Eng, Vol. 54, Issue. 8, pp. 1427 – 1435, August, 2007.

[48] P. Nicoli, **Modeling the Intensity Profiles of Retinal Vessels: An Extended Hermite Model and its Validation**, dissertation, University of Padua, Padua, 2010.

[49] M. Niemeijer, B. van Ginneken and M. Abramoff, **Automatic Classification of Retinal Vessels into Arteries and Veins**, SPIE Medical Imaging, International Society for Optics and Photonics, Vol. 7260, p. 72601F, February, 2009.

[50] M. Niemeijer, X. Xu, A. Dumitrescu, P. Gupta, B. van Ginneken and J. Folk, **Automated Measurement of the Arteriolar-to-Venular Width Ratio in Digital Color Fundus Photographs**. IEEE Trans Med Imaging, Vol. 30, Issue. 11, pp. : 1941 – 1950, November, 2011.

[51] J. Patrick and E. Marshall, **Ophthalmic Photography: Retinal Photography, Angiography, and Electronic Imaging**, 2nd ed., Butterworth-Heinemann Medical; ISBN: 0750673729, August, 2002.

[52] K. Petersen and M. Pedersen, **The Matrix Cookbook**, Technical University of Denmark, January, 2005.

[53] Z. Qiao, L. Zhou and J. Huang, **Effective Linear Discriminant Analysis for High Dimensional, Low Sample Size Data**, Proceedings of the World Congress on Engineering (WCE), Vol. 2, pp. 1070-1075, 2008.

[54] D. Relan, T. MacGillivray, L. Ballerini and E. Trucco, **Retinal Vessel Classification: Sorting Arteries and Veins**, 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 7396-7399, July 3, 2013.

[55] K. Rothaus and X. Jiang, **Classification of Arteries and Veins in Retinal Images Using Vessel Profile Features**, International

Symposium on Computational Models for Life Sciences (CMLS-11), Vol. 1371, No. 1, AIP Conf. Proc, pp. 9-18, 2011 .

[56] A. Ruggeri, E. Grisan and M. De Luca, **An Automatic System for the Estimation of Generalized Arteriolar Narrowing in Retinal Images**. 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, p. 6463, August 22, 2007.

[57] S. Sanjani, J. Boin and K. Bergen, **Blood Vessel Segmentation in Retinal Fundus Images**, [online] Available: https://stacks.stanford.edu/file/druid:yt916dh6570/Bergen_Boin_Sanjani_Blood_Vessel_Segmentation.pdf, 2013.

[58] A. Sharma and K. Paliwal, **Linear Discriminant Analysis for the Small Sample Size Problem: An Overview**, Springer-Verlag, Berlin Heidelberg, 2014.

[59] J. Shivaram, R. Patil and H. Aravind, **Classification of Fundus Photographs using Full Width Half Maximum Algorithm**, International Journal of Computer Applications (0975 – 8887), Vol. 32, No.4, pp. 19-24, October, 2011.

[60] D. Silverthorn, C. Garrison, B. Johnson and W. Ober, **Human Physiology: An Integrated Approach**, 4th ed. Pearson Education, Inc., San Francisco, 2007.

[61] R. Szeliski, **Computer Vision: Algorithms and Applications**, Springer, London, 2011.

[62] L. Tramontan, E. Grisan and A. Ruggeri, **An Improved System for the Automatic Estimation of the Arteriolar-to-Venular Diameter Ratio (AVR) in Retinal Images,** 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, August 20, 2008.

[63] R. Vegaa, G. Sanchez-Antea, L. Falcon-Moralesa, H. Sossab and E. Guevara, **Retinal Vessel Extraction Using Lattice Neural Networks with Dendritic Processing**, Computers in Biology and Medicine archive, Vol. 58, Issue. C, pp. 20-30, March, 2015.

[64] J. Vinayak, **Analysis of Retinal Vessel Networks Using Quantitative Descriptors of Vascular Morphology**, dissertation, University of Iowa, 2012.

[65] X. Xiayu, **Automated Delineation and Quantitative Analysis of Blood Vessels in Retinal Fundus Image**, dissertation, University of Iowa, 2012.

[66] P. Xu, G. Brock and R. Parrish, **Modified Linear Discriminant Analysis Approaches for Classification of High-Dimensional Microarray Data**, Computational Statistics and Data Analysis, Vol. 53, No. 5, pp. 1674-1687, 2009.

[67] J. Ye, T. Xiong, Q. Li, R. Janardan, J. Bi, V. Cherkassky and C. Kambhamettu, **Efficient Model Selection for Regularized Linear Discriminant Analysis,** Proceedings of the 15th ACM International Conference on Information and Knowledge Management, Arlington, Virginia, USA, pp.532–539, 2006.

[68] H. Yu and J. Yang, **A Direct LDA Algorithm for High-Dimensional Data with Application to Face Recognition**, Pattern Recognition Letters, Vol. 34, pp. 2067-2070, 2001.

[69] A. Zamperini, A. Giachetti, E. Trucco and K. Chin, **Effective Features for Artery-Vein Classification in Digital Fundus Images,** 25th International Symposium on Computer-Based Medical Systems (CBMS), IEEE, pp. 1-6, June 20, 2012.

Appendices

Appendix (A): Algebraic Background

A-1: Matrix Transpose.

Let $A \in \mathbb{R}^{m\times n}$. A transpose matrix of A , which is denoted by A^T , can be defined as interchanging i^{th} row by i^{th} column in the matrix A . In other words, if $A = [a_{ij}]_{m\times n}$, then $A^T = [a_{ji}]_{n\times m}$.

The matrix transpose has some properties [14]:

1- $(A^T)^T = A$

2- $(A \pm B)^T = A^T \pm B^T$

3- $(cA)^T = c\,A^T$, where c is a scaler.

4- $(AB)^T = B^T A^T$

5- If A is a non-singular matrix, then A^T is also non-singular and $(A^T)^{-1} = (A^{-1})^T$

If the entry of A are complex numbers ,which can be written as $\alpha = \alpha_1 + \alpha_2 i, where\ i = \sqrt{-1}$, transpose can be obtained by taking the transpose of the matrix then take the conjugate for each element, it is called conjugate transpose and written as A^*.

Matrix transpose is a very important definition which has been used to define other matrix properties, for example: symmetric and skew-symmetric matrix.

Definition [16]: A matrix $A \in \mathbb{R}^{n\times n}$ is said to be symmetric, if $A^T = A$, and it is said to be skew-symmetric if $A^T = -A$. For a given matrix $A \in \mathbb{R}^{m\times n}$, $A^T A$ is symmetric matrix and for $A \in \mathbb{R}^{n\times n}$, $A + A^T$ is also symmetric.

Definition [14]: a matrix $A \in \mathbb{R}^{n\times n}$ is said to be diagonalizable if it can be diagonalized, which means that there exist a non-singular matrix S such that $D = S^{-1}A\,S$ where D is a diagonal matrix.

Spectral theorem [14]**:** if A is a real symmetric matrix, then A is diagonalizable and has only real eigenvalues also it always has n linearly independent eigen-vectors. In other words, there exist a matrix V , which is orthogonal($V^T = V^{-1}$) and consists of eigen-vectors as columns, such that $D = V^{-1}AV$, D is a diagonal matrix with the eigen-values of A in its diagonal.

A-2: Vector and Matrix Derivatives.

Vector derivative with respect to vector [52]

Let $\vec{\boldsymbol{x}} = \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_p \end{pmatrix}$, $\vec{\boldsymbol{a}} = \begin{pmatrix} a_1 \\ a_2 \\ . \\ . \\ . \\ a_p \end{pmatrix}$ be $p \times 1$ vectors, and define $y = \vec{\boldsymbol{x}}^T.\vec{\boldsymbol{a}}$

$$y = x_1a_1 + x_2a_2 + \cdots + x_pa_p$$

$$\frac{\partial y}{\partial \vec{\boldsymbol{x}}} = \left(\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \ldots, \frac{\partial y}{\partial x_p}\right)^T = \begin{pmatrix} a_1 \\ a_2 \\ . \\ . \\ . \\ a_p \end{pmatrix} = \vec{\boldsymbol{a}}$$

where

$$\frac{\partial y}{\partial\, \vec{\boldsymbol{x}}^T} = \left(\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \ldots, \frac{\partial y}{\partial x_p}\right) = (a_1, a_2, \ldots, a_p) = \vec{\boldsymbol{a}}^T$$

The derivative of a 2×2 matrix with respect to a vector [52]

Let $\vec{\boldsymbol{x}} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$ be a symmetric matrix and define $Q = \vec{\boldsymbol{x}}^T A\, \vec{\boldsymbol{x}}$

$$Q = (x_1 \quad x_2)\begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$Q = a_{11}x_1^2 + 2a_{12}x_2x_1 + a_{22}x_2^2$$

$$\frac{\partial Q}{\partial \vec{\boldsymbol{x}}} = \begin{pmatrix} \frac{\partial Q}{\partial x_1} \\ \frac{\partial Q}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 2a_{11}x_1 + 2a_{12}x_2 \\ 2a_{22}x_2 + 2a_{12}x_1 \end{pmatrix}$$

$$\frac{\partial Q}{\partial \vec{\boldsymbol{x}}} = 2\, A\vec{\boldsymbol{x}}$$

Similarly

$$\frac{\partial Q}{\partial\, \vec{\boldsymbol{x}}^{\boldsymbol{T}}} = 2\, \vec{\boldsymbol{x}}^T\, A$$

A-3: Eigen-Values and Eigen-Vectors.

Some problems in applied mathematics require to find a vector $\vec{\boldsymbol{v}}$ such that $A\vec{\boldsymbol{v}} = \lambda\vec{\boldsymbol{v}}$, where λ is a scaler and $\vec{\boldsymbol{v}}$ is a non-zero vector, which means that $A\vec{\boldsymbol{v}}$ parallel to $\vec{\boldsymbol{v}}$. The value of λ is called eigen-value, where the eigen-vector $\vec{\boldsymbol{v}}$ is computed to the associated λ.

Finding the value of $\vec{\boldsymbol{v}}$ and λ is considered as one of the most important problem in numerical linear algebra, as they are used in many different applications.

The eigen-value problem is

$$A\vec{\boldsymbol{v}} = \lambda\vec{\boldsymbol{v}},$$

which is equivalent to

$$(A - \lambda I)\vec{\boldsymbol{v}} = \vec{\boldsymbol{0}}$$

gives

$$\det(A - \lambda I) = \vec{\mathbf{0}}$$

For a given matrix $A \in \mathbb{R}^{n \times n}$, the $\det(A) = \prod_{i=1}^{n} \lambda_i$, where λ_i's are the eigenvalues of A.

Computing the eigen-values and corresponding eigen-vectors is a complex task which is discussed in many numerical analysis books [16, 14].

Appendix (B): Statistical Background

Statistics is a very important subject which is widely used in many applications; one of them is in the machine learning world which is based on statistics. Some basic statistical concepts will be defined in this section.

B-1: Variable

The variable can be defined as a feature which differs from one object to another. For example, height, width, eye color, skin color and marital status, are considered as variables. The first two of these give numerical data, and this type is defined as quantitative variables, while the other ones are non-numerical data hence defined as qualitative variables.

The data of a studied object are the values of the variables' collected in a matrix. The matrix of all individuals called the data matrix or the dataset. The data matrix organized such as a particular column contains all observation for a particular feature or variable, and a particular row consists of all variables for one observation. So the data matrix looks like

$$A = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{k1} & x_{k2} & \cdots & x_{kn} \end{pmatrix}$$

where x_{ij} is the value of i^{th} observation with j^{th} variable, $i = 1,2,\dots,k$ (number of the data elements) , $j = 1,2,\dots,n$ (number of features) [41].

B-2: Averages

Different types of averages are present in statistic, the **mean** is considered as the commonly used average which can be calculated by dividing the sum of all observations in specific dataset by the number of these observations;

it also describes the central tendency for the data. The mean can be calculated as:-

$$\bar{x} = \frac{\sum x_i}{n} \ ,$$

where x_i is the i^{th} observation, $\bar{x}$ is the mean and n is the number of the observations. Another average, **median**, can be calculated by sorting the given data points in a descending or ascending order and then selecting the point which lies in the middle. The **mode** is another average which is the simplest average, it requires to count how many times each point appears and choose the most repeated one as a mode [41].

B-3: Variance and Covariance.

Calculating the data center is not sufficient to understand the behavior of the data, we need to calculate how the data variates around the center. To do so, there are several methods such as calculating the range, variance and standard deviation.

The range can be easily computed by subtracting the minimum value from the maximum one. On the other hand, the standard deviation which is the most accurate and commonly used method to calculate how the values are spread out, it is calculated by using the following formula:

$$S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

$$= \sqrt{\frac{\sum_{i=1}^{n}(x_i^2 - n\bar{x}^2)}{n-1}}$$

$$= \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2/n}{n-1}}$$

where $\bar{x}$ is the sample mean and n sample size. We called the value of $\sum_{i=1}^{n}(x_i - \bar{x})^2$ as sum of squared. As the value of the standard deviation decreases, the data becomes closer to the mean and more tight. While increasing the value means the data is spreading apart from the mean. Standard deviation is always a positive number.

The square of S_x is considered as sample variance $(S_x)^2$.The variance calculates variation in one variable from the mean, but if we need to study more than one variable how they vary together, we need to define what we call covariance matrix.

The covariance calculates if the two variables are dependent or not; if the covariance equals zero then the variables are uncorrelated and hence independent. In the case where the covariance is positive the two variables are directly proportional i.e. one variable increases the other one will also increase. But if the variables are inversely proportional then the covariance will be negative.

If we compute the covariance for each pair of variables and then put these together in a matrix, we will have what we call covariance matrix.

The covariance of two variables x and y can be calculated using this formula [41]

$$cov(x, y) = E\,[(x - \mu_x)\,]\;E\big[\big(\,y - \mu_y\big)\big]$$

recalling that the mean (expectation) of x can be written as $E\,[x]$.

The covariance matrix can be used to look at the correlation between all pairs of variables within a set of data, this covariance matrix Σ representes as [41]

$$\Sigma = \begin{bmatrix} E\,[(x_1-\mu_1)(x_1-\mu_1)\,] & E\,[(x_1-\mu_1)(x_2-\mu_2)\,] & \dots & E\,[(x_1-\mu_1)(x_n-\mu_n)\,] \\ E\,[(x_2-\mu_2)(x_1-\mu_1)\,] & E\,[(x_2-\mu_2)(x_2-\mu_2)\,] & \dots & E\,[(x_2-\mu_2)(x_n-\mu_n)\,] \\ \vdots & \vdots & \ddots & \vdots \\ E\,[(x_n-\mu_n)(x_1-\mu_1)\,] & E\,[(x_n-\mu_n)(x_2-\mu_2)\,] & \dots & E\,[(x_n-\mu_n)(x_n-\mu_n)\,] \end{bmatrix}$$

where x_i is the element of i^{th} variable and μ_i is it's mean, the covariance matrix Σ can also be written in a matrix form as

$$\Sigma = E[(X - E[X])(X - E[X])^T\,]$$

Σ is a square matrix with trace equals one since the variable must be dependent and it is a symmetric matrix since $cov(x, y) = cov(y, x)$ [41].

B-4: Gaussian Distribution.

Gaussian or normal distribution is the most well know distribution with one or more dimension. In the case of one dimension, it has a bell shaped curve as shown in Fig. B.1, and it can be expressed in the following formula

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

While in higher dimensions it looks like

$$f(x) = \frac{1}{(2\pi)^{d/2}.|\Sigma\,|^{1/2}} \exp\left(-\frac{1}{2}\,(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

where μ is the mean , σ is the standard deviation, d is the dimension and Σ is $n \times n$ covariance matrix [41].

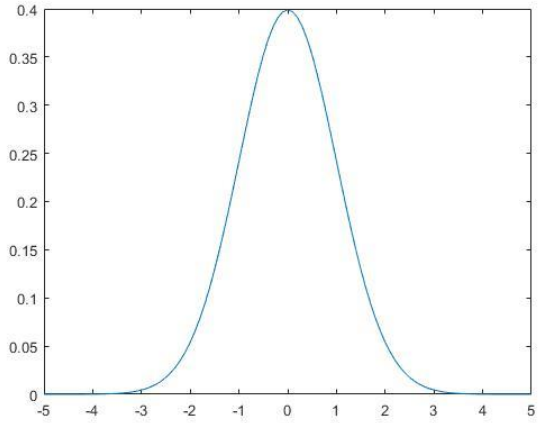


Fig. B.1: Normal distribution in one dimension with $\mu = 0$, $\sigma = 1$.

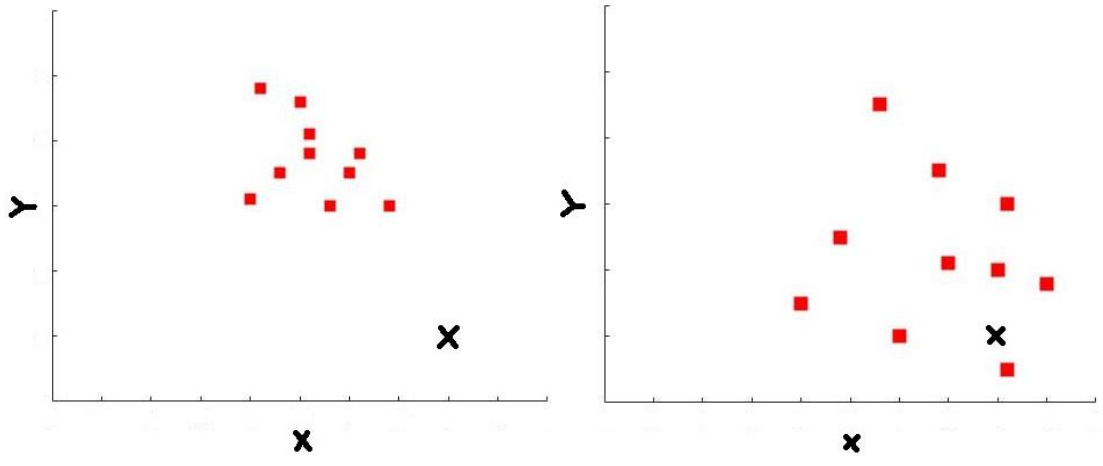
B-5: Mahalanobis Distance.

Assume we have two datasets with one test point labelled by large ‘X’ as shown in Fig. B.2, the first one is tightly correlated while in the other one the data are spread out, the test point lies on the same distance from the data mean but if we ask: is the test point considered as a part of the data? we might say no for dataset in Fig. B.2-a and yes to the other one (Fig. B.2-b), that’s because we can’t only consider the distance from the data mean, but we should also consider how the data is spread out.

In 1936, the distance which depends on the variance in addition to the data mean is called Mahalanobis distance and it is defined by:

$$D_M(x) = \sqrt{(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

where μ is the data mean and Σ is the covariance matrix. If the covariance matrix equals the identity matrix then it is reduced to Euclidean distance [41].



(a) (b)

Fig. B.2: Two different datasets with a test point labelled by ‘X’.

Appendix (C): Classifiers

C-1: Support Vector Machine (SVM)

There are many methods for linear classification; each one makes different assumptions. Moreover different methods have different bias and different objective function, this leads to find a different linear discernment.

Consider the case of two classes C_1 , C_2 , and use the labels $+1$ / -1 for each class respectively. Given the sample $\mathcal{X} = \{\boldsymbol{x}^t , r^t\}$, where $r^t = +1$ if $\boldsymbol{x}^t \in C_1$, and $r^t = -1$ if $\boldsymbol{x}^t \in C_2$, $t = 1,2, \dots , n$ (number of observations). We want to find w_0 (bias) and $\boldsymbol{w}$ (weight vector), such that

$$\boldsymbol{w}^T \boldsymbol{x}^t + w_0 \geq +1 \quad , \quad r^t = +1$$

$$\boldsymbol{w}^T \boldsymbol{x}^t + w_0 \leq -1 \quad , \quad r^t = -1$$

This can be rewritten as

$$r^t \ (\boldsymbol{w}^T \boldsymbol{x}^t + w_0) \geq +1 \qquad (C.1)$$

Note that we don't require simply $r^t \ (\boldsymbol{w}^T \boldsymbol{x}^t + w_0) \geq 0$, which means that we don't only want the observations to be on the right side of hyperplane, but also we want them to be away from this hyperplane in order to get a better generalization. The distance from this hyperplane to the closest observation is called the margin. The optimal separation hyperplane is the one that maximizes the margin [4].

The distance from $\boldsymbol{x}^t$ to the liner discriminant is represented by

$$\frac{|\boldsymbol{w}^T \boldsymbol{x}^t + w_0|}{\|\boldsymbol{w}\|} = \frac{r^t \ (\boldsymbol{w}^T \boldsymbol{x}^t + w_0)}{\|\boldsymbol{w}\|}$$

We want this distance to be larger than or at least equal to some value ρ which should be maximized to get optimal separation. So our problem becomes

$$\frac{r^t\ (\boldsymbol{w}^T\ \boldsymbol{x}^t + w_0)}{\|\boldsymbol{w}\|} \geq \rho \qquad (C.2)$$

This inequality has many solutions, so for unique solution we fixed $\rho\ \|\boldsymbol{w}\| = 1$, thus to maximize the margin we need to minimize $\|\boldsymbol{w}\|$ which also means minimize $\frac{1}{2}\|\boldsymbol{w}\|^2$, so our minimization problem [4]:

$$\min \frac{1}{2}\ \|\boldsymbol{w}\|^2\ , \qquad r^t(\boldsymbol{w}^T\ \boldsymbol{x}^t + w_0) \geq +1 \quad \forall t \qquad (C.3)$$

This is a standard quadratic optimization problem which depends on the number of dimension, but in some cases we might need to map the problem to a new space with dimension higher than the original one, as shown in Fig. C.1, so we need a method whose complexity doesn't depend on the dimension of the data. The optimization problem can be rewritten using the Lagrange multipliers α^t as follows

$$L_p = \frac{1}{2}\ \|\boldsymbol{w}\|^2 - \sum_{t=1}^{N} \alpha^t\ [r^t(\boldsymbol{w}^T\ \boldsymbol{x}^t + w_0) - 1]$$

$$= \frac{1}{2}\ \|\boldsymbol{w}\|^2 - \sum_{t=1}^{N} \alpha^t\ r^t(\boldsymbol{w}^T\ \boldsymbol{x}^t + w_0) + \sum_{t=1}^{N} \alpha^t \qquad (C.4)$$

This should be minimized with respect to $\boldsymbol{w}$ and w_0, while maximized with respect to $\alpha^t \geq 0$. Since the main term of the problem is convex and the linear constrain is also convex, then this problem is called convex quadratic optimization problem [4].

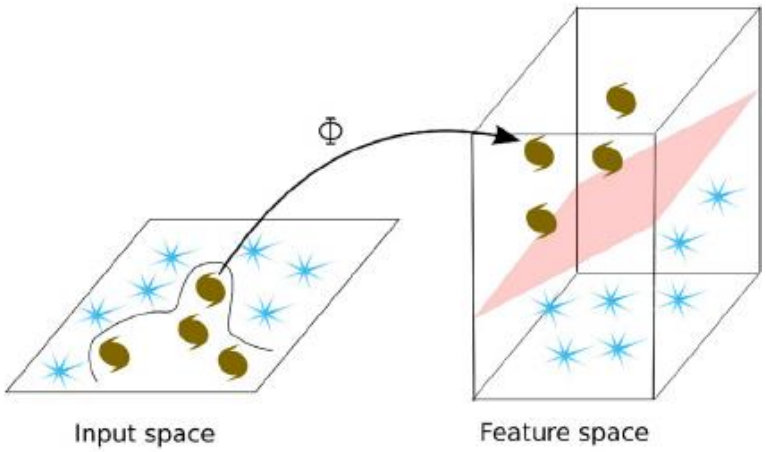


Fig. C.1: Separating data by converting it into higher dimension.

Therefor this is equivalent to solving a dual problem using Karush-Kuhn-Tucker conditions. The dual is to maximize L_p with respect to α^t, subject to the constrains that gradient of L_p with respect to $\boldsymbol{w}$ and w_0 are 0 and also $\alpha^t \geq 0$:

$$\frac{\partial L_p}{\partial \boldsymbol{w}} = 0 \;\rightarrow\; \boldsymbol{w} = \sum_t \alpha^t \, r^t \, \boldsymbol{x}^t \qquad (C.5)$$

$$\frac{\partial L_p}{\partial w_0} = 0 \;\rightarrow\; \sum_t \alpha^t \, r^t = 0 \qquad (C.6)$$

Plugging these two equations in equation (C.4) we get [64]

$$L_p = \frac{1}{2}\,(\boldsymbol{w}^T \boldsymbol{w}) - \boldsymbol{w}^T \sum_t \alpha^t \, r^t \, \boldsymbol{x}^t - w_0 \sum_t \alpha^t \, r^t + \sum_t \alpha^t$$

$$= -\frac{1}{2} \sum_t \sum_s \alpha^t \alpha^s \, r^t \, r^s \, (\boldsymbol{x}^t)^T \, \boldsymbol{x}^s + \sum_t \alpha^t \qquad (C.7)$$

This can be solved using quadratic optimization methods. The set of $\boldsymbol{x}^t$ whose $\alpha^t > 0$ are called supported vectors, these vectors lie on the margin. As we see from equation$(C.5)$ $\boldsymbol{w}$ is written as sum of training instances that are selected as supported vectors $\boldsymbol{x}^t$ satisfying:

$$r^t\ (\boldsymbol{w}^t\ \boldsymbol{x}^t + w_0) = 1$$

As we said the supported vectors lie on the margin, so using this fact, w_0 can be calculated as follows:

$$w_0 = r^t - \boldsymbol{w}^t\ \boldsymbol{x}^t$$

For testing we calculate $g(x) = \boldsymbol{w}^t\ \boldsymbol{x} + w_0$, then [4]

we choose $\begin{cases} C_1 & if\ g(x) > 0 \\ C_2 & otherwise \end{cases}$

C-2: Naïve Bayes Classifier

Naïve Bayes assumes that the elements of features vector are conditionally independent, so for a given class C_i , the values of different features don't affect each other. It's called naïve because of the previous assumption, since it tells us that the features are independent, which means if our aim is to classify vessels, it assumes that intensity and red color are independent, which clearly isn't true.

The classifier tries to estimate $P(\boldsymbol{x}_j|C_i) = P(\boldsymbol{x}_j^1\ , \boldsymbol{x}_j^2, \dots, \boldsymbol{x}_j^n |\ C_i\ )$, where the superscripts index is the element of the vector, which is equal to the product of multiplying together all individual probabilities, which is much easier to compute.

$$P(\boldsymbol{x}_j|C_i) = P(\boldsymbol{x}_j^1 = \boldsymbol{a}_1|C_i) \times P(\boldsymbol{x}_j^2 = \boldsymbol{a}_2|C_i) \times \dots \times P(\boldsymbol{x}_j^n = \boldsymbol{a}_n|C_i)$$

$$P(\boldsymbol{x}_j|C_i) = \prod_k P(\boldsymbol{x}_j^k = \boldsymbol{a}_k|C_i)$$

So the classifier rule for the Naïve Bayes' classifier is to select the class C_i which maximizes the following:

$$P(C_i) \prod_k P(\boldsymbol{x}_j^k = \boldsymbol{a}_k|C_i)$$

This is clearly a great simplification over evaluating the full probability, so it might came as surprise that the Naïve Bayes' classifier has been shown to have comparable result to other classification methods [4].

C-3: K-Nearest Neighborhood Classifier.

K-nearest neighborhood (K-NN) is one of the simplest classifiers which has been used since 1970 in many domains like pattern recognition and statistical estimation. It classifies unlabeled observations by assigning them to the class of the most similar labeled examples. The distance between test point and training points are calculated, then the point with lowest distance is called nearest neighbor.

The method calculates the distance between new observations and the k nearest labeled observations. There are many methods to calculate this distance and for any of them there are three conditions should be satisfied which are:-

1) $dist\ (A,B) \geq 0$ and $dist(A,B) = 0 \ \leftrightarrow A = B$
2) $dist\ (A,B) = dist\ (B,A)$
3) $dist\ (A,C) \leq\ dist\ (A,B) + dist\ (B,C)$

where $dist\ (A,B)$ defines the distance between the two points A and B.

By default, the k-NN function employs Euclidean distance; where some methods used another function for calculating the distance, for example Mahalanobis and Manhattan distance. The nearest neighbor algorithms can yield excellent performance when used with a proper distance measure.

Choosing appropriate parameter k, which defines how many neighbors will be chosen for K-NN, is a very sensitive issue since large k reduces the

impact of variance caused by random error, but it runs the risk of ignoring small and important patterns. Appropriate k should balance between overfitting and underfitting. Some authors suggest choosing k equal to the square root of the number of observations in the training dataset [4, 31, 24].

جامعة النجاح الوطنية

كلية الدراسات العليا

استخدام النمذجة الرياضية والتمايز بالتحليل الخطي لتصنيف الأوعية الدموية إلى شرايين وأوردة

إعداد

الاء محمد بواقنة

إشراف

د. هادي حمد

قدمت هذه الاطروحة استكمالا لمتطلبات الحصول على درجة الماجستير في الرياضيات المحوسبة بكلية الدراسات العليا في جامعة النجاح الوطنية في نابلس، فلسطين

2018م

استخدام النمذجة الرياضية والتمايز بالتحليل الخطي
لتصنيف الأوعية الدموية إلى شرايين وأوردة

أعداد
الاء محمد بواقنة
باشراف
د. هادي حمد

الملخص

تصنيف الاوعية الدموية في العين الى شريان ووريد تعتبر مسألة مهمة، وذلك لأنها تساعد في الكشف المبكر عن الامراض المزمنة التي قد تصيب العين والذي بدوره يحمي من العمى. هناك العديد من الخصائص التي يمكن استخراجها من صورة شبكية العين وذلك للمساعدة في تصنيف الأوعية الدموية، لكن بعض هذه الخصائص يمكن أن يكون عديم الفائدة ويعيق عملية التصنيف وهادراً للوقت، وهذا يؤدي الى تصنيف غير دقيق للأوعية الدموية.

في هذا العمل تم مناقشة بعض الطرق لتقليص عدد الابعاد (الخصائص) مثل تحليل التمايز الخطي (Linear discriminant analysis (LDA)) وكذلك تحليل المكونات الرئيسية (Principal component analysis (PCA)) .وتم اقتراح تطبيق طريقة اضافية لتقليص عدد الابعاد من خلال مرحلتين بحيث أُسقطت الخصائص من بعد الى بعد أصغر باستعمال تحليل المكونات الرئيسية ومن ثم تطبيق تحليل التمايز الخطي لتقليص عدد الابعاد الناتج مرة اخرى الى بعد واحد فقط . ولتحقيق هذا الهدف تم اعداد تطبيق من خلال برنامج الماتلاب من أجل تنفيذ هذه الطرق بالاضافة الى عمليات اخرى على الصور. تم استعمال صور من قاعدة البيانات (DRIVE)، حيث أُجريت بعض التحسينات على هذه الصور ومن ثم تمت عملية استخراج الخصائص من الأوعية الدموية، بعد ذلك طُبقت الطرق المذكورة سابقا من أجل تقليل عدد الأبعاد. في النهاية تم اعتماد النتائج الخارجة من تطبيق قاعدة الاغلبية (الأكثرية) لثلاثة من طرق التصنيف وهي خوارزمية أقرب جار (K-NN)، متجهات الآت دعم التمييز أو شبكات دعم التمييز (SVM) ومصنف بايز (Naïve Bayes).

في الجزء الاخير من هذا العمل، تم قياس وتقييم النتائج من خلال الحساسية، النوعية والدقة. اظهرت النتائج أن كل من تحليل التمايز الخطي و تقليص عدد الابعاد من خلال المرحلتين كان قادراً على تقليص عدد الابعاد (الخصائص) إلى بعد واحد بالاضافة إلى تقليص الوقت الازم لإجراء عملية التصنيف. علاوة على ذلك كلتا الطريقتين حسنت النتائج النهائية و لكن تقليص عدد الابعاد من خلال المرحلتين أظهرت نتائج أفضل قليلاً. إن تقليص عدد الابعاد من خلال مرحلتين تمكنت من تصنيف الأوعية الدموية بشكل صحيح في 14 صورة بمعدل دقة مقدراه 86.4% .