An-Najah National University

Faculty of Graduate Studies

Mathematical Optimization Approach to Enhance the Retinal Segmentation Through the Support Vector Machine Classifier

By

Islam Ibrahim Sweileh

Supervisor

Dr. Hadi Hamad

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

2021

Mathematical Optimization Approach to Enhance the Retinal Segmentation Through the Support Vector Machine Classifier

By

Islam Ibrahim Sweileh

This Thesis was Defended Successfully on 1 / 2 /2021 and approved by:

Defense Committee Members		Signature
1. Dr. Hadi Hamad	/ Supervisor	.....
2. Pro. Saed Mallak	/ External Examiner	.....
3. Dr. Amjad Hawash	/ Internal Examiner	.....
4. Dr. Mohammed Yasin	/ Internal Examiner	.....

Dedication

I dedicate my work to my family members; my parents, my brothers Thaer, Thabit, Zain, and to my sisters Isra’ and Abrar. I also dedicate my work to my dear husband Anas and to my father and mother in law who all encouraged me to finish this work successfully. Thanks for all of them who worked as a source of inspiration to me.

Acknowledgments

At the beginning thanks God, I am grateful to have completed this thesis. This thesis would not have been possible without the support of many people. Many thanks to my supervisor, Dr. Hadi Hamad, who read my work numerous revisions and helped to make some sense of the confusion. Thanks to An-Najah National University for awarding me a Dissertation Completion Fellowship. And finally, thanks to my parents, husband “Anas” and my beloved son “Kinan” who endured this long process with me, always offering support, love and passion.

الإقرار

أنا الموقعة أدناه، مقدمة الرسالة التي تحمل العنوان:

Mathematical Optimization Approach to Enhance the Retinal Segmentation Through the Support Vector Machine Classifier

أقر بأن ما اشتملت عليه هذه الرسالة إنما هو من نتاج جهدي الخاص باستثناء ما تمت الإشارة إليه حيثما ورد، وإن هذه الرسالة ككل، أو أي جزء منها لم يقدم لنيل أي 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: إسلام إبراهيم أحمد حميلح

Signature: التوقيع: إسلام حميلح

Date: التاريخ: 1/2/2021

List of Contents

List of Tables

List of Figures

Mathematical Optimization Approach to Enhance the Retinal Segmentation Through the Support Vector Machine Classifier

By
Islam Ibrahim Sweileh
Supervisor
Dr. Hadi Hamad

Abstract

The eye of the human is considered as a sensitive and delicate part which is imposed to many diseases. Since it is an important part of the body and needs intensive care, many algorithms are suggested to diagnose and detect ophthalmic diseases.The retina plays an important role in seeing things; it is composed of many parts such as optic disc, arteries, veins, macula … etc. Artery/vein vessels are affected in different ways as an influence of diseases, hence detecting these vessels will make a good contribution in diagnosing the diseases, and image processing is a powerful tool to aid doctors to achieve this goal.

Segmentation is defined as detecting the required object by defining the location in the image. It is the first step in a long and complicated process to diagnose the diseases; difficulty raises due to the vary change in vessels color in the same image and also in different images, also the spherical shape of retina add more difficulties, in addition to noise and out layers. They have different shapes and sizes that make this work difficult.

The methodology implemented in this work was constructed depending on the Support Vector Machine (SVM) which is used for classification task , the idea is to map the training data into a higher dimensional feature space where a separating hyperplane with the weight

vector and the bias can be found to maximize the margin from the closest data points. The features that help us for this classification are the Gray-level based feature and Moment invariants-based features. This algorithm was trained and implemented on DRIVE dataset which is a public database on the internet. The code was built under the python environment.

To evaluate and compare the performance of the algorithm, the method compared its results with ground truth which is available in the dataset, using sensitivity, specificity and accuracy giving the results 66.5 %, 85.1%, 80.6% respectively.

Introduction

The segmentation of blood vessels can be done through different methods that can be classified into Supervised and Unsupervised approaches. Supervised techniques require the images which are manually segmented for the purpose of training the system and afterwards the system is tested using the defined parameters.

Unsupervised approaches use inherent patterns to do segmentation. All segmentation techniques end in the classification of each pixel into either a vessel or non-vessel. The classified vessels form blood vessel patterns, the abnormal ones can help in the detection of various diseases.

SVM is considered a successful method when used for pattern classification problems. Utilizing the Support Vector approach to a special practical problem involves resolving many questions based on the problem definition and the design involved within it. One of the major challenges here is that of choosing an appropriate kernel for the given application. There are standard choices such as a Gaussian or polynomial kernel that are the default options, but if these proved to be ineffective, or if the inputs are discrete structures, more elaborate kernels will be needed. Once the choice of kernel and optimization criterion has been made the key components of the system are in place.

Traditional classification methods perform poorly when working directly with high dimensionality data, but Support Vector Machines can avoid the pitfalls of very high dimensional representations. A very similar approach to the techniques described for text categorization can also be used for the task of image classification, and as in that case linear hard margin machines are frequently able to generalize well. The first real-world task on which SVM were tested was the problem of hand-written character recognition. Furthermore, multi-class SVMs have been tested on these data, putting in mind that prior knowledge is certainly remarkable [12].

Chapter One

Preliminaries

1-1: Introduction

1-2: The Retina

1-2: Common Retinal Diseases and Conditions

1-3: Image Processing

1-4: Image Segmentation

1-5: Clustering Methods

Chapter One

Preliminaries

1-1 Introduction

The human eye has a spherical shape with approximately 20 mm in diameter. It is enclosed by four membranes: the cornea and sclera construct the outer cover, the choroid, and the retina, see Fig.1.1. The pigment of the iris differs according to its position; the front has visible pigment while the back iris has a black one. The pupil of the eye (iris center) diameter varies from approximately .

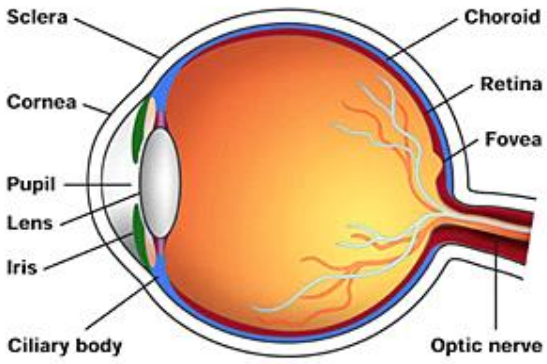


Fig. 1.1: The retina of the eye

The cornea, the transparent part of the eye, primarily protects and helps reflect light. Light travels through an opening in our eye called the pupil. It is like a camera shutter; it expands or contracts to let light through. The amount of light entering through the pupils is regulated by the iris, the colored part of our eye that increases or decreases the diameter of the pupil, thereby controlling the amount of light entering the retina. After that, the

light that has passed through the pupil reaches the lens; It acts like a magnifying glass that reflects light onto the retina below.

The shape of the lens constantly changes depending on whether you are looking at a near or far object. The process of changing shape is called accommodation. With age, the lens becomes harder and cannot reflect light properly. So, we need reading glasses. Typically, the lens reflects light upside down and rotates towards the back of the eye (retina). Our brain flips the image so that we see the correct image.

1-2 The Retina

The retina is an amazing structure; it converts light into nerve signals, which allows us to see in different conditions and distinguish many different wavelengths and colors. The retina is considered part of the brain and is directly connected to it by bundles of nerve fibers. [1]

The retina, the nerve part of the eye, is the inner lining of the eye that receives light from a focused object outside the eye, hence the receptors (cones and rods) where they are located. Concentrated in the center of the retina - very sensitive to light [2], see Fig.1.1. By converting light into nerve signals, the useful is extracted, and the redundancy of the environment is ignored.

The retina receives blood through the central arteries / veins of the retina and the choroid. However, retinal vessels enter and exit the eye through the optic nerve and enter the layer of nerve fibers, important

arterial and venous branches, which together form an arch and supply each of the quadrants of the retina [4].

The retina is the layer that lines the inner surface of the eye and contains photoreceptors (cells that are sensitive to light).Photoreceptors transfer the light into neural signals, which are transmitted to human brains via the optic nerves. In order to register the conditions of retina, an image of the retina can be obtained (fundus image, Fig. 1-2). The retinal microscope is a fundus camera system commonly used to take images of the retina. The retinal image contains important diagnostic information that can be used to determine if the retina is healthy or not.

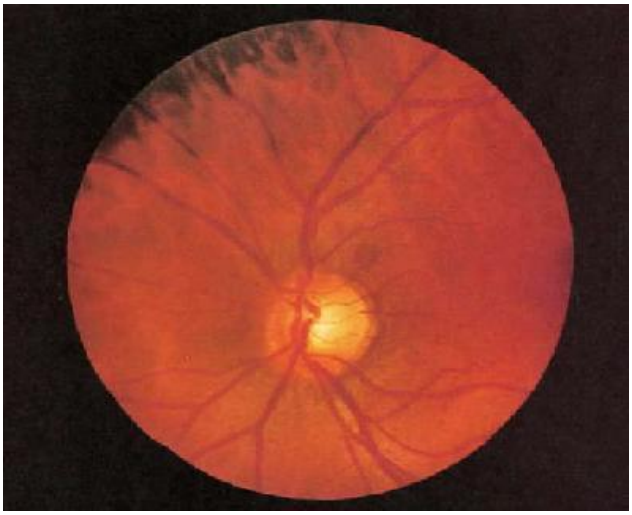


Fig. 1.2: A Fundus image of the Retina

1-3 Common retinal diseases and conditions

Retinal diseases pose many risks for person's vision and can lead to blindness. A list of some retinal diseases are presented in the following :

Epiretinal membrane:

Epiretinal membrane is a sensitive tissue that looks like crinkled cellophane lying on top of the retina. This condition can cause upset balance, blurred vision [5], see Fig.1.3.

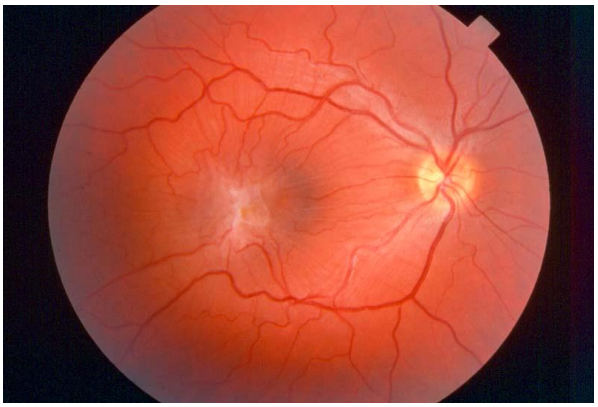


Fig. 1.3: The Fundus Image of Epiretinal membrane

Macular hole:

A macular hole is a small hole in the center of the retina at the back of the eye. In most cases, this is due to an abnormality involving the interface between the retina and the overlying glassy [4], as illustrated in Fig. 1.4.

Fig. 1.4: A Fundus image with Macular Hole

Retinitis pigmentosa:

Retinitis pigmentosa is a degenerative disease that affects the retina. This condition can lead to severe vision loss or blindness [1], its effect is shown in Fig. 1.5.

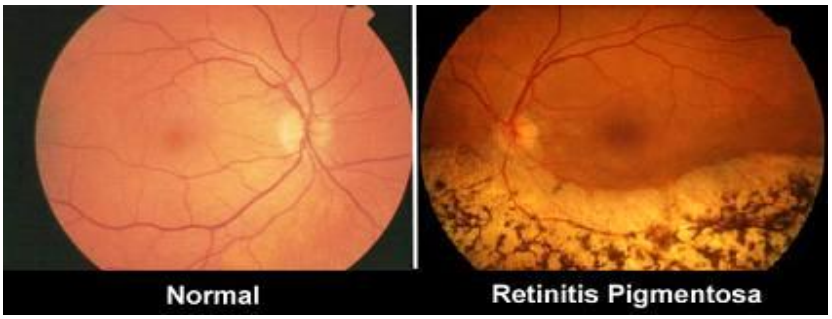


Fig. 1. 5: A Comparison between Normal Retina and Retinitis Pigmentosa

Cystoid macular edema:

Cystoid macular edema is a painless disorder that effects the retina. As the retina is the light-sensitive layer of tissue that lines the back wall of the eye and is responsible for vision, when this condition occurs, swelling develops in the macula, which is the central portion of the retina critical for seeing fine details clearly. These swollen areas fill with fluid, causing the macula to thicken, see Fig.1.6.

Fig. 1. 6: A Fundus image of Cystoid Macular Edema

Eye diseases are widespread in Palestine. For example, Dr. Riyad G. Banayot conducted a study on children at Al Ain Hospital in Hebron, which shows the following results for children aged (0-5) years, associated with a list of diseases such as those shown in Table 1.1 [4].

Table 1.1: Eye diseases at Al-Ein Hospital

No.	Name of disease	Percentage of catching
1	Ocular disorders	31.6 %
2	Conjunctival diseases	23.7 %
3	Strabismus and amblyopia	13.8 %

1427 people studied (2854 eyes) in the West Bank and in the Gaza Strips results showed a high incidence of eye diseases, especially conjunctivitis, cataracts (including nuclear), clouds, pigmentary dystrophies of the cornea, retina and diabetic retinopathies, congenital diseases. Blindness was mainly the result of corneal opacities, which most often developed after trachoma, and congenital eye diseases. Parental marriage appears to play an important role in the development of congenital and abiotrophic retinal pigment disorders.

1-4 Image Processing

The importance of digital image processing stems from two main areas of application: improving image information for human interpretation and analysis, and digital scene processing for automatic machine perception [2].

Artificial Intelligence (AI) is defined as the exploration of intelligent agents: any device that detects its environment and takes actions that maximize its chances of successfully achieving its goals. In common parlance, artificial intelligence is used when a machine imitates human cognitive functions such as learning and problem solving [2].

Artificial intelligence aims to provide ophthalmologists with new automated tools for the diagnosis and treatment of eye diseases. This change is partly due to the fact that big players in the digital world like Google and IBM have been paying more attention to the medical potential of AI in recent years. However, in AI ophthalmology work, computational analysis is seen as a pathway to more efficient and objective methods of understanding the flow of images that create new eye care techniques, according to ophthalmologists involved in the effort. [2].

Many promising computer algorithms are used in the field of retinal diseases. For example, in 2016, researchers at the Google Brain Initiative published that their artificial intelligence system Deep Learning accurately detected diabetic retinopathy (DR) and diabetic macular edema in fundus images. In addition, it is used for other retinal diseases including age-related macular degeneration (AMD), pediatric premature retinopathy (ROP), and pseudandrous reticular is to better identify or evaluate other ocular conditions, including pediatric cataracts, glaucoma, corneal ecstasies, and oculoplastic reconstruction. [2]

There is a holistic approach, from validation to complex management, where these algorithms can do things better and make things more objective. There are many cases where doctors disagree, but the artificial intelligence system produces the same results every time, said Dr. Michael D. Abramoff, a prominent figure in the discovery of AI in the world. Iowa Ophthalmic University in Iowa City [2].

A digital image is a discrete representation of an image, which can be defined as a spatial matrix, where the row (x) and column (y) are spatial (plane) coordinates in the displayed image so that the entity value, that is, f (x, y) represents the gray level in gray image (0-255) (see Fig. 1.7). Each element of the matrix image is a pixel that measures the image resolution [2].

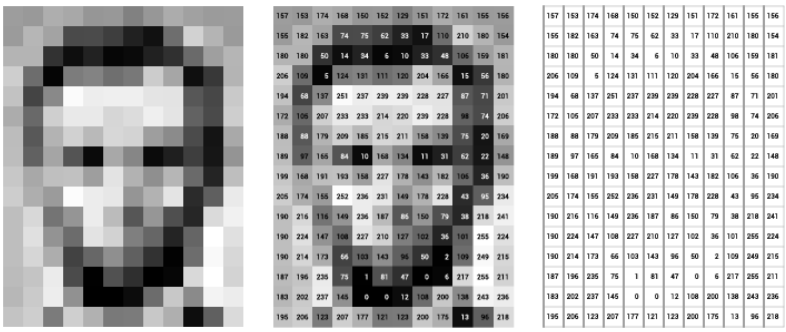


Fig. 1.7: Image - Matrix correspondants**.**

Digital images can be classified by type based on their color characteristics. For example, RGB (red, green, and blue) used in color monitors, CMYK (cyan, magenta, yellow and black) used in printing, and HSI (hue, saturation, intensity) when manipulating an image. [2].

Image processing is widely used in various areas of our life; Medical image processing plays an important role in this regard. Using image enhancement techniques, doctors can also explain X-rays and other biomedical images. Today there are indicators for the automatic detection of medical errors, which depend on digital images obtained from the background, X-rays, resonance and many other types [2].

1-5 Image Segmentation

Image segmentation is the process of dividing an image into regions or objects. This is the first step in image analysis. Image processing presents images and edits them to improve them, while image analysis tries to determine what is in the image [6].

The main idea behind image segmentation is to split an image into parts or objects. There are different image segmentation techniques that differ in complexity, performance and application [6].

Segmentation is one of the most important steps in image processing; dividing the inside of the image into many parts which is something more important and easier for further process. It may also depend on different features that are contained in the image; they might be color, intensity… etc. The principal purpose of segmentation is to decrease the information for easy analysis. It is also useful in image analysis and image compression [6].

Segmentation techniques can be classified as follows (see Fig. 1.9) [12] :

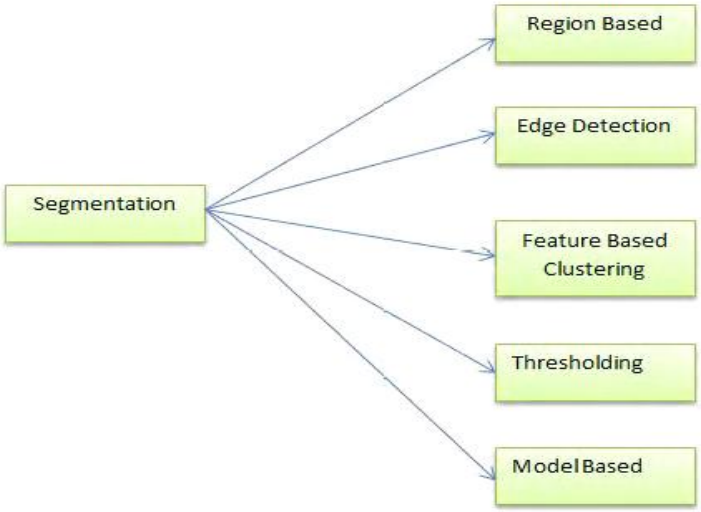


Fig. 1.8 : Classification types of segmentation

1. **Region Based:**

This method assumes that the pixels associated with the object are placed together for segmentation. Segmentation by region is also called similarity-based segmentation [8], in which at least one pixel is associated with an area and counted at each step.

There are two basic operations of region-based segmentation, splitting and merging. Merging segmentation is accomplished through the following steps:

1. Obtain an initial segmentation of the image.

2. Merge those adjacent segments that are similar to form single segments.

3. Go to step 2 until no segments that should be merged remain.

The initial segmentation may simply be all pixels, i.e., each pixel is a segment by itself. The meaning of merging is to decide if two segments should be merged or not. The merging criterion is based on:

- Grey value similarity.
- The edge strength of the boundary between the segments.
- The texture of the segments.

The basic form of image segmentation using splitting is:

1. Obtain an initial segmentation of the image.

2. Split each segment that is inhomogeneous.

3. Go to step 2 until all segments are homogeneous.

2. **Edge Detection:**

Segmentation can also be done by using edge detection techniques. In this technique the boundary is identified to segment the image, where edges are detected to identify the discontinuities in the image. Edges on the region are traced by identifying a pixel value and it is compared with the neighboring pixels. For this kind of classification both fixed and adaptive feature of Support Vector Machine (SVM) are used [9].

Also, there is no need for a detected edge to be closed; the canny edge detector is an example of edge detector method which has the following steps (as described in Fig. 1.10) [12] :

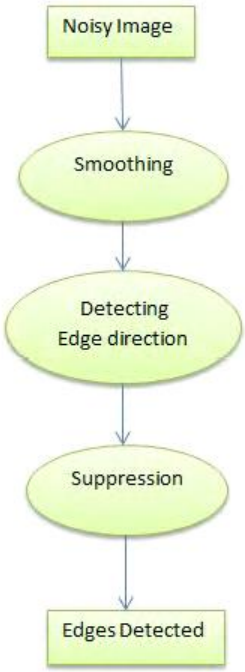


Fig. 1.9 : Canny edge detection

- To reduce the effect of noise, the surface of the image is smoothed using Gaussian convolution.

- Sobel operator is applied to the image to detect the edge strength and edge directions.

- The direction of the edges is taken into account for non-maximum suppression, for example. Pixels not associated with edges are captured and then minimized.

- The last step, for example, is to remove the broken edges. The threshold of the image is computed, and then the pixel value is compared to the resulting threshold. If the pixel value is greater than the threshold value, it is considered a border or rejected (no border) [7].

The separated edges give the layout of an object. The objects are traced when the separated edges are calculated accurately [7]. Adding more about this type of detection:

* Edges are detected by calculating the minimum and maximum of the first derivative in the gradient edge detector.

* Sobel Edge Detector uses convolution kernel to detect edges.

* The spatial gradient size is calculated for edges in the Roberts Edge Detector.

* The Canny Edge detector also uses a high spatial gradient, but requires more computational effort than the Sobe lend Robert Edge detector. [11] [7].

a b c

Fig. 1. 10 : a) Sobel, b) Roberts, C) Canny edge detectors

3. **Feature Based Clustering:**

Grouping is done by moving the image in the form of a histogram. The grouping takes place accordingly. For example, colored pixels can be grouped together using fuzzy technology [11]. For structured images, a basic clustering algorithm is used, such as the K-Means clustering method. When one segment using feature based clustering, changes are made and color components based on image characteristics as the difference in the values of intensity and color [7].

The fuzzy clustering method is used to segment color images by defining the color space in the image space, this method is used to identify the colored area of the image [12].

The clustering technique is performed on the cylindrical elements of the color space solution, and the area is obtained by a histogram and is recognized as a cluster using a threshold [12]. Sorting pixels for layout has a drawback [7] (see Figure 1.12).

Fig. 1.11 : Clustering

4. Thresholding:

Thresholding is the easiest way of segmentation; it can be done by threshold values which are obtained from the histogram of the edges of the original image. If edges detection is accurate, then the thresholding will be accurate too [8]. Comparing to other methods, thresholding is the simplest and computationally the fastest.

Segmentation is done through a modified way called adaptive thresholding. The gray level points where the gradient is high, are added to thresholding surface for segmentation. The drawback of this segmentation technique being not suitable for complex images (see Fig. 1. 13).

Fig. 1. 12 : Thresholding segmentation.

5. Model Based*:*

The model, also known as Markov Random Field (MRF) [14], is used in color segmentation. Portions of color pixel tuples are recognized as independent random variables for further processing. This method is

combined with edge detection for precise edge detection [6]. In addition, MRF has a limitation in the smoothness of the spatial domain and there are relationships between color components [14].

The initial segmentation is done at a coarse resolution, then the process continues iteratively at a higher resolution. Resolution segmentation is performed only for part of the image, so it is fast [15]. Segmentation can be performed using a Gaussian Markov Random Field (GMRF) that takes into account spatial relationships between pixels. Gaussian Markov Model (GMM) segmentation is also used for region growth. A GMM extension that recognizes the region as well as signals of advantage within the GMM. The space of functionality is also recognized using this technique [16].

One of the uses of segmentation is in the recognition and measurement of medical images such as retinal segmentation. In this work, the retina is divided into two parts: blood vessels (in the foreground) and non-blood vessels (in the background).

1-6 Clustering Methods

Various techniques for the segmentation of the retinal image have been published in the literature. Based on machine learning methods, retinal blood vessel segmentation can be classified into two groups: supervised methods [1, 8-11] and unsupervised methods [12-15]. Supervised methods are based on the previous labeling data which identifies whether a pixel belongs to a vessel or non-vessel class. Whereas,

unsupervised methods do not use prior labeling information and have the ability to learn and organize information on its own to find the patterns or clusters that resemble, in this case, the blood vessels.

The detection of diseases of the autonomic nervous system is becoming a potential mission due to the extraordinary capabilities of new computers using classification and grouping techniques. Making computers able to argue and deduce facts is called artificial intelligence. This is partly machine learning,which splits into two types:

- **Supervised Learning**; where training data of correct examples is given to define some combining parameters to generalize the respond correctly.

Supervised algorithms work as classifiers; such that input feature vectors are determined to which class they belong based on exemplars used for training the algorithm to distinguish between the classes, each vector belongs to one out of N defined classes [17].

Supervised methods learn the rule for vessel extraction by means of an algorithm which extracts knowledge based on a training set of images which are manually processed and segmented. A very significant approach is based on the application of line operators and support vector machine (SVM) for pixel classification.

In addition, the learning algorithm receives a sequence of pairs of inputs and outputs of following form $\langle x_i, y_i \rangle$, where determined the possible input, and where the is the output assigned to . Pairs of inputs and

outputs are called examples. The learning algorithm should be able to find a function f that calculates as the examples where $f\ (x_i) = y_i$ for all i, which provides a good expectation of the outputs of inputs that were not identified.

The input can be descriptions of the objects, while the output classes are the objects. Equivalently, inputs can be predictions of situations and the output represent actions, Now, if the function is discreet valued, then the outputs are called classes and the learning task is called classification [17].

• **Unsupervised Learning;** the data is classified without examples depending on similar features between the input data. That means, data is not presented with an appropriately packed sequence of inputs and outputs.

Unsupervised learning approaches for detecting (classifying) blood vessels attempt to find inherent patterns of blood vessels in retinal images that can be used to determine whether a particular pixel belongs to the blood or not. In this approach, the manual data are used only for measuring performance metrics.

Also, Clustering is a form of unsupervised learning employed in pattern recognition tasks. The algorithm divides the inputs into a fixed number of clusters (subsets); those inputs in the same cluster are close to each other regarding some metric on the space of input.

The diversity of information and data made it impossible to define one specific way (algorithm) to enable machines to understand and analyze correctly things; hence, many algorithms from different terminologies and perspectives are proposed, for instance, support vector machine, evolutionary learning, reinforcement learning, Markov models, decision trees, multi-layer perceptron, k-means... etc.[17].

Clustering and classification are fundamental tasks in Data Mining. Classification is used mostly as a supervised learning method, and it's a predictive method. On the other hand, clustering is used for unsupervised learning, and the goal of this is descriptive, in the following a number of these algorithms are described:

- **Neural Network:**

One of the popular neural clustering algorithms is the self-organizing map (SOM). This algorithm creates a single layer network. SOM algorithm has been used in vector quantization and speech recognition. This is useful for viewing large data in 2D or 3D spaces. But it is very sensitive to the initial choice of the weight vector; also to it's various parameters, like the learning rate as well as the proximity range [10].

This type of algorithm represents each cluster by a neuron or "prototype". The input data is also represented by neurons, which are connected to the prototype neurons. Each such connection has a weight, which is learned adaptively during learning. The prototype neurons compete for the current instance, and the winner is the neuron whose

weight vector is closest to the instance currently presented. By addition, the winner and its neighbors learn by having their weights adjusted.

- **Fuzzy Clustering:**

The familiar fuzzy clustering algorithm that the FCM (Fuzzy C-Means) algorithm. Although the FCM algorithm is more effective than the hard mean K algorithm for avoiding local minima, it can still converge towards the local minima of the squared error criterion. The most important question in fuzzy grouping is the construction of the membership function. The various options include those based on the cluster similarity and center of gravity expansion. A generalization of the FCM algorithm using a family of objective functions was proposed [17].

- **Decision Trees:**

As for the decision trees, the information is represented by a hierarchical tree, every single leaf points contains a unique kind of probabilistic description of the mentioned concept. The most famous algorithms:

COBWEB: This algorithm figures that all kind of the attributes are independent. The goal is to achieve a high level of predictability in the values of the dummy variables for a particular cluster. This algorithm might be not the suitable choice for clustering large databases.

CLASSIT: Is the COBWEB extension for continuous information, unfortunately it has problems similar to COBWEB algorithm [10].

- **Partitioning Methods:**

Partitioning methods is to move the instances by moving them from one cluster to another one, according to the initial partition. Such methods usually require the number of clusters to be predefined by the user. A complete process of listing all possible partitions is required to achieve overall optimization in partitioned clustering. Since this is not possible, some greedy heuristics can be used in the form of iterative optimization. The displacement method iteratively moves points between k clusters [10].

5. Support Vector Machine Methods:

Currently, various classification algorithms are widely used in various applications, especially in machine learning. While each has its own merits, each has its drawbacks from different perspectives. The SVM (Support Vector Machine) classifier was chosen because of its features.

When working in an SVM environment, you should optimize the size used and suggest the correct kernels. This should improve the performance and reduce time complexity when ranking. Then different classifier models are explored by optimizing and proposing different parameters and kernels. The results are then analyzed. Further explanation and details can be found in the second chapter.

Chapter Two

The Support Vector Machine Classifier

2-1: Introduction

2-2: Statistical Learning Theory

2-3: Support Vector Machine

2-3-1 History

2-3-2 Principle of SVM

2-3-3 Properties of SVM

2-3-4: Lagrange

2-3-5: Kernel Trick

2-4: Application and Related Work

Chapter Two

The Support Vector Machine (SVM) Classifier

2-1 Introduction

Machine learning is implemented by computer programs to optimize performance criteria and leverage previous examples and experiences. The criterion is expressed by a parameter driven model; these parameters are coordinated and optimized by examining the algorithm using training data or previous experience. Models used in machine learning can be predictive or descriptive for extracting knowledge from data, or both. These models use statistical theory to construct mathematical models [1].

Support Vector Machines (SVMs) are mainly used for binary classification. The main idea of this method is to find a hyperplane that divides the data into d-dimensions into different classes. However, the data often cannot be divided linearly. Therefore, SVM introduces the concept of kernel-induced functional space, which maps data into a higher dimensional space where the data is converted into shared data. Transforming data into a larger space generally poses computational problems in addition to overfitting [21].

SVM is an integral part of the global machine learning community. Pixel maps have made SVM a well-known algorithm because it offers higher accuracy than complex neural networks with extended functions in many applications, such as handwriting recognition [18], it is also used

effectively in facial recognition, pattern classification, and regression applications.

2-2 Statistical Learning Theory

The statistical learning theory mainly provides a framework for understanding the problem of inference, hence gaining knowledge, making predictions and decisions or building models from a dataset [22].

The supervised learning problem in statistical theory is formulated as follows: suppose we have a set of training data $\{(X_1, Y_1) \ldots (X_n, Y_n)\}$ in $R_n \times R$ which sampled according to unknown probability distribution $P(x, y)$, and have $V\big(y, f(x)\big)$ as loss function to measure the error, for x as an input data the function $f(x)$ well represent the "predicted" in place of the actual value y. The problem which the theory tries to solve is to find a function f that minimizes the expectation of the error on new data; (this function f is to minimize the expected error):

$\int\int V(y, f(\boldsymbol{x}))\, P(\boldsymbol{x}, y)\, d\boldsymbol{x}\, dy$ [23].

In statistical modeling we would choose a model from the hypothesis space, which is closest to the underlying function in the target space [22].

The main purpose of statistical learning is to get a description of an unknown dependency between measurements of objects and certain properties of these objects. The measurements, also known as" input variables", are expected to be observable in all objects of interest. On the

contrary, the properties of the objects, or "output variables", are in general available only for a small subset of objects.

The purpose of estimating the dependency between the input and output variables is to be able to determine the values of output variables for any object of interest.

Learning and inference in a statistical learning theory is a process which can be roughly summarized in the following steps:

1. Observe a phenomenon.

2. Construct a model of that phenomenon.

3. Make predictions using this model.

2-3 Support Vector Machine (SVM)

SVM, which is considered to be one of the most popular algorithms in machine learning today, SVM is a supervised classification method implemented in machine learning. The basic version of SVM accepts data records as inputs and determines for each of them to which of the two possible classes each input belongs, making it a non-probabilistic linear binary classifier [19].

Discriminatory SVMs are becoming an increasingly popular method for machine learning issues involving classification and regression.

For a certain set of training models, which all belong to one of two possible classes, SVM represents the models as points in space, which are mapped in such a way that samples of different classes with a distance between them are separated from each other ; as much as possible [19].

Adding to linear classification, carrier vectors can successfully perform nonlinear classification using kernel methods by implicitly mapping the input data into multidimensional spaces.

2-3-1 History

SVM was introduced by Vapnik in 1992. SVM often offers significantly better classification performance than other classification methods if the data set is of appropriate size [18]. It was originally developed for binary classification. Classifying multiple classes with SVM is still a research challenge.

The SVM algorithm is a non-linear circular generalization of the portrait algorithm developed in Russia in the 1960s (Vapnik and Lerner 1963, Vapnik and Chervonenkis (VC) 1964). It is firmly anchored in the statistical learning theory developed over the past three decades by Vapnik and Chervonenkis (1974) and Vapnik (1982, 1995). In short, VC theory characterizes the properties of learning machines that allow them to generalize well to new data [18].

In its current form, SVM has been extensively developed by Vapnik and his staff at AT&T Bell Laboratories. Due to this industrial context, SVM research has so far had a strong trend towards real applications. The first works focused on OCR (optical character recognition). In a short time, SVM collectors became competitive with the best systems available for object recognition and OCR activities [11].

2-3-2 Principle of SVM

SVM's are one of the most popular classification algorithms and well known for their strong theoretical foundations, generalization performance and ability to handle high dimensional data. Starting with linear SVMs, followed by their extension to the nonlinear case [19].

Now let us take the following example; suppose the dataset is of two dimensions (x_1 , x_2) and each element of the data belongs to one of the two classes c_1 or c_2, where the element of c_1 means (+1) and the element of c_2 means (-1). By plotting the given data, it can be noticed that c_1 and c_2 are located in different regions and one can distinguish between them by adding a line to classify the dataset. There are many lines which can correctly classify the data into the two classes, but which one is better? see (Fig. 2.1).

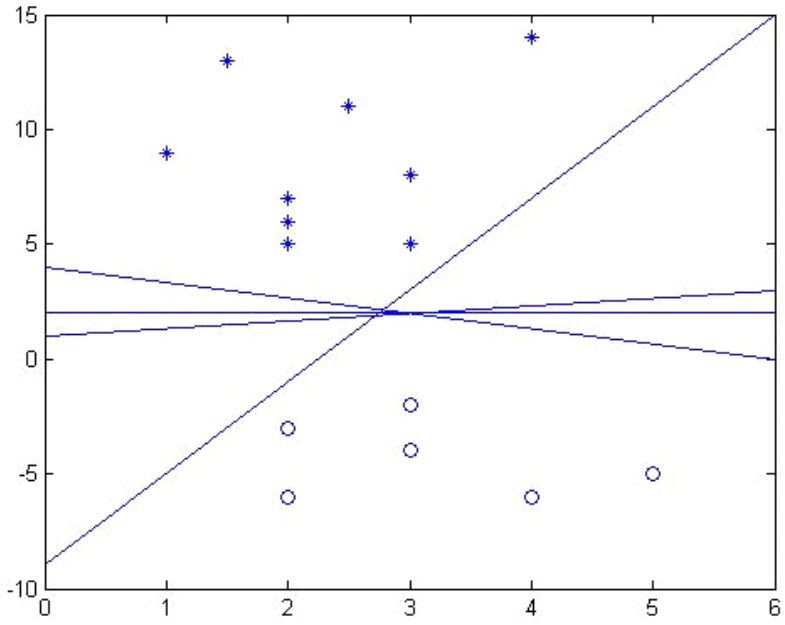


Fig. 2.1: Possible Classification Lines

In the figure above, it can be seen that there are many rows (hyperplanes) which can classify the data correctly. The question, however, is which of these solutions the best is? The idea here is to know which hyperplane reaches the maximum separation [11].

Define the best separation hyperplane for the dataset, which is used to quantify the optimal row, i.e. the best row, according to our criteria. To measure distance, it is necessary to move away from the line (in a direction perpendicular to the line) before reaching a data point [18].

It is assumed that the training data indicates an underlying process that the algorithm is trying to learn and that the test data on which the algorithm is evaluated after training comes from the same underlying process [19].

This major radius is called the edge and is labeled M (see Fig.2.2). This is the ultimate edge classifier. The data points of each class closest to the edges of the classification line have support vectors. The points represented as squares are called support vector points (see Fig. 2.2).

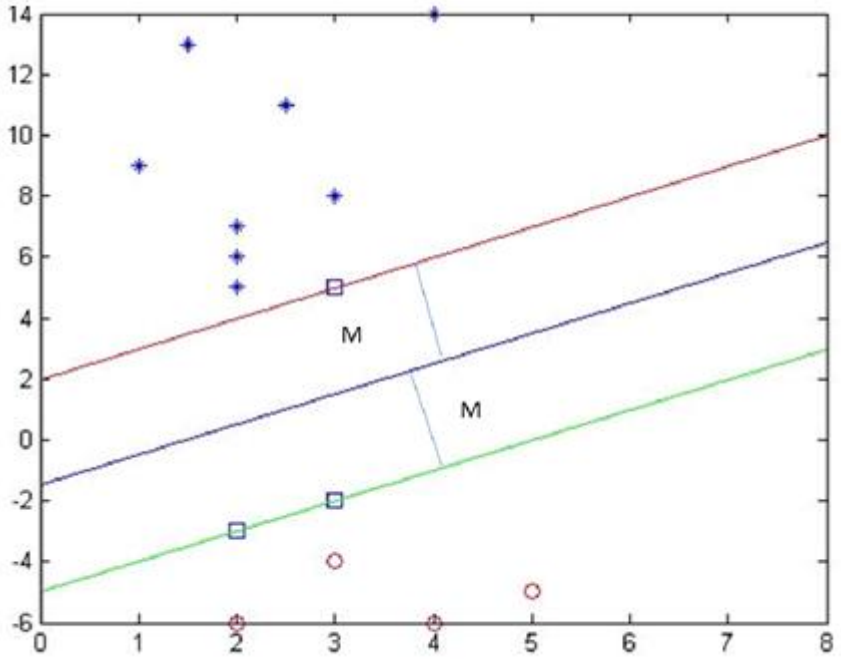


Fig. 2. 2: The margin region (M) separating the classes

To describe the hyperplane by using a standard equation of a straight line is defined as:

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

Where

$\boldsymbol{w}$: is the weight vector.

$\boldsymbol{x}$: is the particular input vector.

b: is the contribution from the bias weight.

Fig (2.2) has two classes; values of $\boldsymbol{x}$ which gives a positive value, considered as class1, and vectors of negative values are considered as class 2.

The region between the hyperplanes that separates the two classes is called no man's land. it usually has no data. If a point is in this region, these points do not belong to any of the classes.

To check if a value belongs to class 1or 2, the method evaluates $\boldsymbol{w}.\boldsymbol{x} + b$, where $\boldsymbol{w}.\boldsymbol{x}$ is defined as the inner product

$$\boldsymbol{x} = \sum_i \boldsymbol{w}_i \ .\boldsymbol{x}_i$$

if it is positive then the point belongs to class1, oppositely, if the value is negative it belong to class 2, in case the absolute value of $\boldsymbol{w}.\boldsymbol{x} + b$ is less than the margin M, then the method does not define the type of the points satisfying in the inequalities :

$$-M \leq \boldsymbol{w}.\boldsymbol{x} + b \leq M$$

Then the point is not classified [31].

The points of the support vector are therefore at the boundaries of the hyperplanes

$\boldsymbol{w}.\boldsymbol{x} = M \quad (b = 0)$.

If $\boldsymbol{x_1}$ is a support vector at the boundary of class 1, then $\boldsymbol{w}.\boldsymbol{x}^{+} = 0$, similarly, if we have as the support vector at the boundary of class 2, then $\boldsymbol{w}.\boldsymbol{x}^{-} = 0$.

To find the edge length, find the closest point to $\boldsymbol{x}^{+}$ located on the Class 1 boundary line, then move perpendicular to Class 2 boundary line. The distance between the two dividing lines is $2M$ [31].

Notice that $\boldsymbol{x}^{+}$ and $\boldsymbol{x}^{-}$ are support vectors, so

$$\mathbf{w}.\boldsymbol{x}^{+} = 0$$

$$\mathbf{w}.\boldsymbol{x}^{-} = 0$$

The two equations are equal

$\boldsymbol{w}.\boldsymbol{x}^{+} = \boldsymbol{w}.\boldsymbol{x}^{-}$

$\rightarrow \boldsymbol{w}.(\boldsymbol{x}^{+} - \boldsymbol{x}^{-}) = 0$, by inner product

$$\boldsymbol{a}.\boldsymbol{b} = \|\boldsymbol{a}\| \, \|\boldsymbol{b}\| \, cos\,\theta$$

where,

$\|\boldsymbol{a}\|$: is the length of vector $\boldsymbol{a}$.

$\|\boldsymbol{b}\|$: is the length of vector $\boldsymbol{b}$.

θ: is the angle between $\boldsymbol{a}$ and $\boldsymbol{b}$.

Since. $(\boldsymbol{x}^{+} - \boldsymbol{x}^{-}) = \mathbf{0}$, either $\|\boldsymbol{w}\|$ or $\|\boldsymbol{x}^{+} - \boldsymbol{x}^{-}\|$ needs to be zero, $\|\mathbf{w}\|$ is not equal to zero, and $\|\boldsymbol{x}^{+} - \boldsymbol{x}^{-}\|$ is not equal to zero. So, only cosθ should be equal to zero, this means $\theta = \frac{\pi}{2}\ or - \frac{\pi}{2}$.

This includes $(\boldsymbol{x}^{+} - \boldsymbol{x}^{-})$ is a straight line between two points lying on boundary and $\boldsymbol{w}$ must be perpendicular to ($\boldsymbol{x}^{+} - \boldsymbol{x}^{-}$).

If it is perpendicular to the classifier line, then it is clearly perpendicular to the boundary lines of both class 1 and 2, so the direction travelled in between two boundary lines, written as an equation

$$\boldsymbol{x}^{-} = \boldsymbol{x}^{+} + v\,\mathbf{w}$$

It is known that $|\boldsymbol{x}^{+} - \boldsymbol{x}^{-}| = \mathbf{2}\boldsymbol{M}$, and so one can use the equation above to compute that:

$$\boldsymbol{M} = \frac{\mathbf{1}}{\mathbf{2}|\boldsymbol{w}|} = \frac{\mathbf{1}}{\mathbf{2}\sqrt{\boldsymbol{w}.\boldsymbol{w}}}$$

Using the above equation helps to compute the margin **M** [31].

We want to find $\boldsymbol{w}$ and $\boldsymbol{b}$ that give us the greatest possible value $\boldsymbol{M}$. We need to make $\boldsymbol{M}$ as large as possible in the same way by making $\boldsymbol{w}.\boldsymbol{w}$ as small as possible.

If this were the only constraint, then set $\boldsymbol{w} = \mathbf{0}$, so the problem is solved, but it is required to move the classification line to separate class1 from class2. Therefore find a decision limit that classifies it well by making it as small as possible [31].

To avoid the mistakes, write down a set of constraints that the classifier should get the answer right. One can write down $t_i \times y_i$, that is, the target multiplied by the output, and this will be positive if the two are the same and negative otherwise. Then use the equation

$t_i(\mathbf{w}.\boldsymbol{x} + b) \geq 1.$

It is worth considering the case where one classifier makes a mistake by placing a point on the wrong side of the line while another places the same point on the wrong side of the line. It can be argued that the first classifier is better than the second because the errors were not fatal. Therefore, one must include this information in our minimization criteria. To solve the problem, add a term to the minimization criterion so as to minimize $\boldsymbol{w} + \lambda \boldsymbol{x}$, where λ is an offset parameter that determines the weight placed on each of the two criteria (small λ means large margin) about some errors, while a large λ means the opposite). This actually turns the problem into a soft margin grader that allows for some errors. The function to be minimized has the following formation:

$$L(\boldsymbol{w}, \in) = \boldsymbol{w}.\boldsymbol{w} + \boldsymbol{\lambda} \sum_{i=1}^{\boldsymbol{R}} \in_{\boldsymbol{i}}$$

where R is the number of misclassified data points, and each $\in_i$ is the distance to the correct boundary line for the missing point, it is sometimes known as a slack variable [19,20].

Example: Linear case

Suppose given the following positively labeled data points in R^2 :

$$Class\ 1\ = \left\{\begin{pmatrix}3\\1\end{pmatrix},\begin{pmatrix}3\\-1\end{pmatrix},\begin{pmatrix}6\\1\end{pmatrix},\begin{pmatrix}6\\-1\end{pmatrix}\right.$$

and the following negatively labeled data points in R^2 : (see Fig. 2.3)

$$Class\ 2 = \left\{\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix},\begin{pmatrix}0\\-1\end{pmatrix},\begin{pmatrix}-1\\0\end{pmatrix}\right.$$

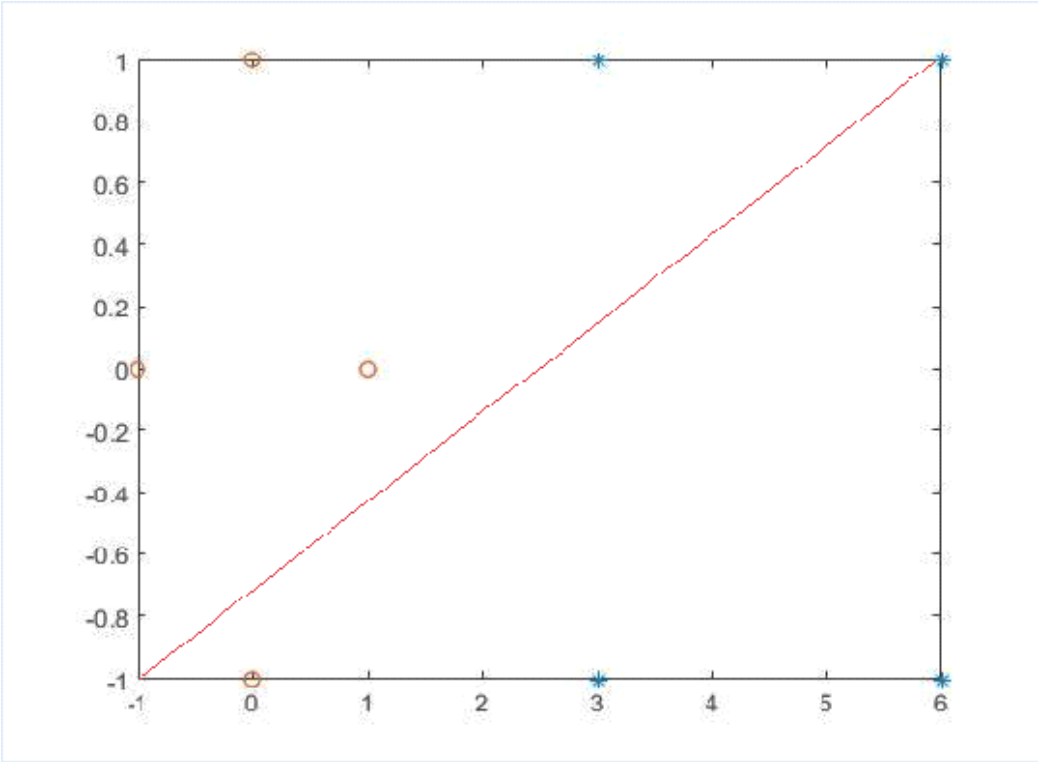


Fig. 2. 3: Separating the data in a feature space

To discover a simple SVM that accurately discriminates the two classes since the data is linearly separable, use a linear SVM by inspection, it should be obvious that there are three support vectors. (see Fig. 2.4)

$$s_1 = \begin{pmatrix}1\\0\end{pmatrix}, s_2 = \begin{pmatrix}3\\1\end{pmatrix}, s_3 = \begin{pmatrix}3\\-1\end{pmatrix}$$

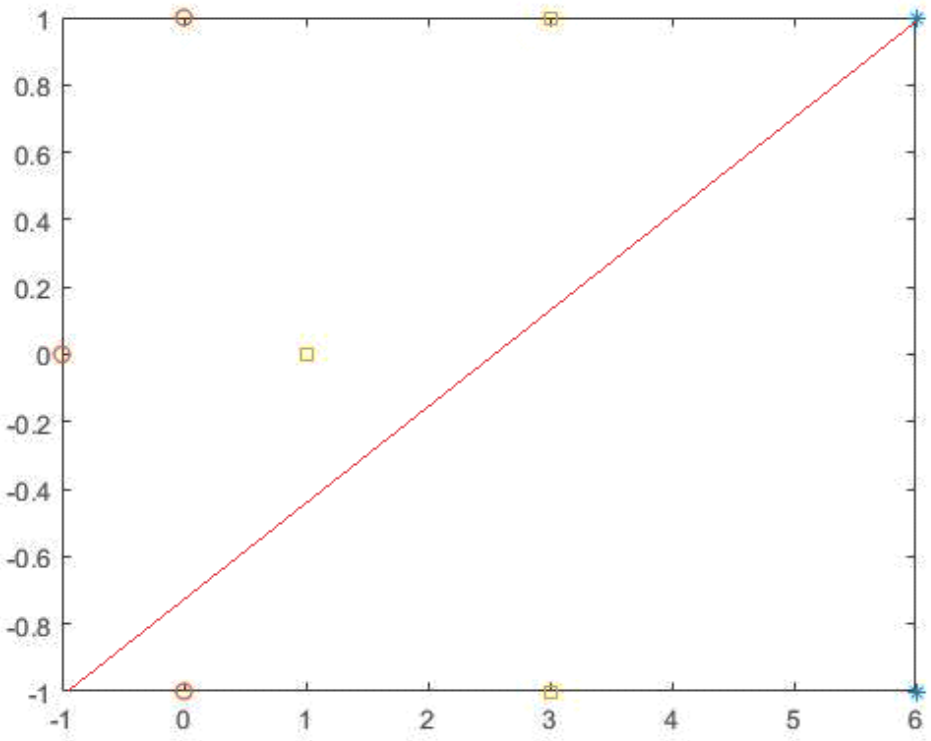


Fig. 2: 4: The squares are the support vector points

Augmented vectors with a 1 are considered as a bias input, for clarity to differentiate these an over-tilde is used. So, if $s_1 = (10)$, then $\sim s_1 = (101)$. our task is to find values for the α_i such that

$$\alpha_1 \phi(s_1).\phi(s_1) + \alpha_2 \phi(s_2).\phi(s_1) + \alpha_3 \phi(s_3).\phi(s_1) = -1$$

$$\alpha_1 \phi(s_1).\phi(s_2) + \alpha_2 \phi(s_2).\phi(s_2) + \alpha_3 \phi(s_3).\phi(s_2) = +1$$

$$\alpha_1 \phi(s_1).\phi(s_3) + \alpha_2 \phi(s_2).\phi(s_3) + \alpha_3 \phi(s_3).\phi(s_3) = +1$$

Since for now we have let $\phi(\,) = I$, this reduces to:

$$\alpha_1 (\sim s_1).(\sim s_1) + \alpha_2 (\sim s_2).(\sim s_1) + \alpha_3 (\sim s_3).(\sim s_1) = -1$$

$$\alpha_1 (\sim s_1).(\sim s_2) + \alpha_2 (\sim s_2).(\sim s_2) + \alpha_3 (\sim s_3).(\sim s_2) = +1$$

$$\alpha_1(\sim s_1).(\sim s_3) + \alpha_2(\sim s_2).(\sim s_3) + \alpha_3(\sim s_3).(\sim s_3) = +1$$

So far, the discussion of SVM has taken place in the current context of learning about linear decision limits. However, the power of SVMs also can be fully harnessed while extended by linear SVMs to allow for more general decision surfaces.

These were linear examples and how SVM works. If there is a nonlinear dataset, SVM cannot simply draw a linear hyper-plane. So SVM uses the kernel trick. When having nonlinear data, the kernel method helps to find patterns and relationships to achieve high accuracy in the final machine learning model [20].

2-3-3 Properties of SVM:

SVMs belong to a family of generalized linear classifiers and can be interpreted as an extension of the perceptron. A special property is that they simultaneously minimize the empirical classification error and maximize the geometric margin; hence they are also known as maximum margin classifiers [41].

SVM possess the following properties:

- Flexibility of choosing a similar function.

- Simplicity of solution when dealing with large data sets support vectors are specifically used to specify the separation.

• It has the ability to handle large feature spaces-complexity, it is not dependent on the dimensionality of the feature space.

• Overfiting can be controlled by soft margin.

• Nice math property: a simple convex optimization problem is guaranteed to converge to one global solution.

2-3-4 Lagrange

When we have a function f, constrained by another function g, we have an extremum (max or min) when the normals to these functions are parallel, that is, the functions are tangent to each other [29].

This gives the fundamental equation for Lagrange Multipliers [45]:

$$\nabla f(x) = \lambda \nabla g(x)$$

Where λ is the Lagrange multiplier, the equation shows that this type of gradients are parallel, however can have other dimensions or directions, i.e. λ is a scale factor [45].

Lagrange Multipliers are a mathematical method used to solve constrained optimization problems of differentiable functions. Basically you have some function $f(x_1 \dots x_n) : R^n \to R$ that you want to optimize (i.e. find the min or max extremes), in addition to this function, you also have a constraint $g(x_1 \dots x_n) = 0$. So, we are trying to optimize f, while constraining f with g, the constraint acts here as a boundary [45].

In the core of Lagrange Multipliers is found the below equation:

$$\nabla f(x) = \lambda \nabla g(x)$$

So this space shows us that the gradient of f is almost the same value to some multiplier (Lagrange multiplier) times the gradient of g. So, having [2].

$$g(x) = 0$$

In the context of SVMs, the last two equations are mixed together into single equation we called it the Lagrangian [45]:

$$L(x) = f(x) - \lambda\, g(x)$$

Using this equation, we look for points where [2]:

$$\nabla L(x, \lambda) = 0$$

Example: One Equality Constraint

Given the function:

$$f(x, y) = 2 - x^2 - 2y^2$$

With the constrain:

$$g(x, y) = x + y - 1 = 0$$

Find the extreme value.

Solution:

At the begining, to put the equations in a form of a Lagrangian:

$$L(x, y) = f(x, y) - \lambda\, g(x, y)$$

$$L(x, y) = 2 - x^2 - 2y^2 - \lambda(x + y - 1)$$

Solve for the gradient of the Lagrangian:

$$\nabla L(x, y, \lambda) = \nabla f(x, y) - \lambda \nabla g(x, y)$$

which gives:

$$\frac{\partial}{\partial x} L(x, y, \lambda) = -2x - \lambda = 0$$

$$\frac{\partial}{\partial y} L(x, y, \lambda) = -4y - \lambda = 0$$

$$\frac{\partial}{\partial \lambda} L(x, y, \lambda) = x + y - 1 = 0$$

From the first two equation, we have $x = 2y$. Substituting this into the last equation gives $x = \frac{2}{3}$ and $\frac{1}{3}$. These values gives $\lambda = -\frac{4}{3}$ and $\frac{4}{3}$.

Multiple Constraints

Lagrange multipliers also work well with different restrictions, simply adding one more limit to the problem. Note that with equality constraints, the problem does not stay within a limit, but actually touches the limit. A simple Lagrange reformulation takes into account several restrictions [45].

$$L(x,\lambda) = f(x) - \sum_i \lambda_i \ g_i(x)$$

Here $g_i(x)$ and λ_i are the multiple constraints. Keep in mind that each one of this constraint has its own multiplier.

Again, here we week for the points where [45]:

$$\nabla L(x,\lambda) = 0$$

Example: Given the multivariable function

$$f(x) = x^2 + \ y^2$$

With the following two constrains

$$g_1(x,y) = x + 1 = 0$$

$$g_2(x,y) = y + 1 = 0.$$

Find the extreme value.

Solution: First, we put the equations into the form of a Lagrangian:

$$L(x,y,\lambda) = f(x,y) - \lambda_1 g_1(x,y) - \lambda_2 g_2(x,y)$$

$=x^2 + y^2 - \lambda_1(x+1) - \lambda_2(y+1)$

solve for the gradient of the Lagrangian :

$$\nabla L(x,y,\lambda) = \nabla f(x,y) - \lambda_1 \nabla g_1(x,y) - \lambda_2 \nabla g_2(x,y) = 0$$

which gives:

$$\frac{\partial}{\partial x} L(x, y, \lambda) = 2x - \lambda_1 = 0$$

$$\frac{\partial}{\partial y} L(x, y, \lambda) = 2y - \lambda_2 = 0$$

$$\frac{\partial}{\partial \lambda_1} L(x, y, \lambda) = x + 1 = 0$$

$$\frac{\partial}{\partial \lambda_2} L(x, y, \lambda) = y + 1 = 0$$

The last two equations give $x = -1$, $y = -1$. Substituting this in the first two equations give $\lambda_1 = -2$, $\lambda_2 = -2$ and $f = 2$.(extreme value of the function)

Applying Lagrange Multipliers Directly to SVMs

Let's assume that two classes of two-dimensional data to be separated. Also assume that each class consists of only one point. These points are:

$$\boldsymbol{x}_1 = A_1 = (1,1)$$

$$\boldsymbol{x}_2 = A_2 = (2,2)$$

To find the hyperplane that separates these two classes. First from SVM theory, we know that the equations are:

$$f(\boldsymbol{w}) = \frac{1}{2} \|\boldsymbol{w}\|^2$$

$$g_i(\boldsymbol{w}, b) = y_i[\langle \boldsymbol{w}, \boldsymbol{x_i} \rangle + b] - 1 \geq 0$$

by expanding $g_i(\boldsymbol{w}, b)$ a bit further:

$$g_1(\boldsymbol{w}, b) = [\langle \boldsymbol{w}, \boldsymbol{x_1} \rangle + b] - 1 \geq 0$$

$$g_2(\boldsymbol{w}, b) = -[\langle \boldsymbol{w}, \boldsymbol{x_2} \rangle + b] - 1 \geq 0$$

Next, put the equations into the form of a Lagrangian:

$$L(\boldsymbol{w}, b, \lambda) = f(\boldsymbol{w}) - \lambda_1 g_1(\boldsymbol{w}, b) - \lambda_2 g_2(\boldsymbol{w}, b)$$

$$= \frac{1}{2} \|\boldsymbol{w}\|^2 - \lambda_1([\langle \boldsymbol{w}, \boldsymbol{x_1} \rangle + b] - 1) - \lambda_2(-[\langle \boldsymbol{w}, \boldsymbol{x_2} \rangle + b] - 1)$$

solve for the gradient of the Lagrangian:

$$\nabla L(\boldsymbol{w}, b, \lambda) = \nabla f(\boldsymbol{w}) - \lambda_1 \nabla g_1(\boldsymbol{w}, b) - \lambda_2 \nabla g_2(\boldsymbol{w}, b) = 0$$

which gives:

$$\frac{\partial}{\partial \boldsymbol{w}} L(\boldsymbol{w}, b, \lambda) = \boldsymbol{w} - \lambda_1 \boldsymbol{x}_1 - \lambda_2 \boldsymbol{x}_2 = 0$$

$$\frac{\partial}{\partial b} L(\boldsymbol{w}, b, \lambda) = -\lambda_1 + \lambda_2 = 0 \quad \rightarrow \quad \lambda_1 = \lambda_2$$

$$\frac{\partial}{\partial \lambda_1} L(\boldsymbol{w}, b, \lambda) = [\langle \boldsymbol{w}, \boldsymbol{x_1} \rangle + b] - 1 = 0$$

$$\frac{\partial}{\partial \lambda_2} L(\boldsymbol{w}, b, \lambda) = [\langle \boldsymbol{w}, \boldsymbol{x_2} \rangle + b] + 1 = 0$$

Now,

$$[\langle \boldsymbol{w}, \boldsymbol{x_1} \rangle + b] - 1 = [\langle \boldsymbol{w}, \boldsymbol{x_2} \rangle + b] + 1 = 0$$

$$\langle \boldsymbol{w}, \boldsymbol{x}_1 \rangle + b - 1 = \langle \boldsymbol{w}, \boldsymbol{x_2} \rangle + b + 1$$

$$\langle \boldsymbol{w}, \boldsymbol{x}_1 \rangle + b = \langle \boldsymbol{w}, \boldsymbol{x_2} \rangle + b + 2$$

$$\langle \boldsymbol{w}, \boldsymbol{x_1} \rangle - \langle \boldsymbol{w}, \boldsymbol{x_2} \rangle = 2$$

$$\langle \boldsymbol{w}, [\boldsymbol{x}_1 - \boldsymbol{x}_2] \rangle = 2$$

we can break $\boldsymbol{w}$ down into its components:

$$\boldsymbol{w} = (w_1 , w_2)$$

$$\langle \boldsymbol{w}, [\boldsymbol{x}_1 - \boldsymbol{x}_2] \rangle = 2$$

$$\langle (w_1 , w_2), [(1,1) - (2,2)] \rangle = 2$$

$$\langle (w_1 , w_2), [-1 , -1] \rangle = 2$$

$$-w_1 - w_2 = 2$$

$$w_1 = -(2 + w_2)$$

$$(w_1 , w_2) - \lambda_1[(1,1) + \lambda_2(2,2)] = 0$$

$$(w_1 , w_2) - \lambda_1[(1,1) + \lambda_1(2,2)] = 0$$

$$(w_1 , w_2) + \lambda_1[(1,1)] = 0$$

which yields

$$w_1 + \lambda_1 = 0$$

$$w_2 + \lambda_1 = 0$$

Equating these gives

$$w_1 = w_2$$

So, by substituting:

$$w_1 = w_2 = -1$$

$$\lambda_1 = \lambda_2 = 1$$

$$b = 1 - \langle \overline{\boldsymbol{w}}, \overline{\boldsymbol{x_1}} \rangle$$

$$= -1 - \langle \overline{\boldsymbol{w}}, \overline{\boldsymbol{x_2}} \rangle$$

$$= -1 - \langle (-1,-1),(1,1) \rangle$$

$$= -1 - \langle (-1,-1),(2,2) \rangle = 3$$

2-3-5 Kernel Trick

The key information used in SVM is that the largest dimension space does not have to be addressed directly (since it turns out that only the point product formula is needed in that space). In general, SVMs are intuitive, theoretically sound, and have proven effective in practice. SVMs also have been extended in order to solve regression problems (where the system is trained to generate a numerical value instead of a (yes or no) classification) [21].

If the data is linear, you can use a separating hyperplane to separate the data. Even though; it's often happens that the data is not linear and the data series are inextricably linked. To make it separable, kernels are used to convert the input data non-linearly by mapping it into a large separable

space where the new assignment is then linearly separable see Fig (2.5) [25] .

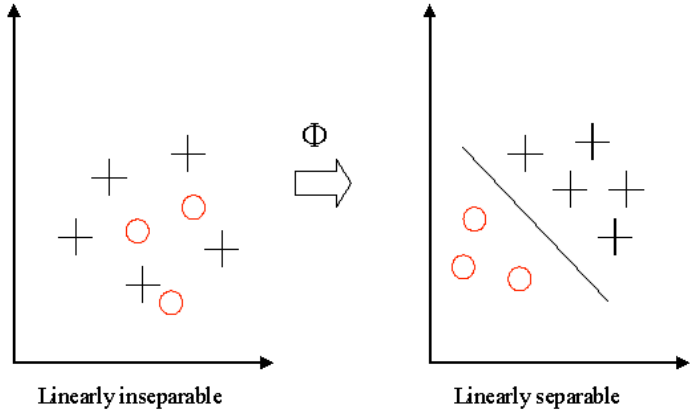


Fig. 2. 5: kernels convert the non- separable data into separable

The above mentioned map is to be defined by the Kernel:

$$K(x, y) = \emptyset(x).\emptyset(y)$$

Characteristic space: By converting the data into a characteristic space, it is very hard to define a similarity measure based on the scalar product. If the functional space is chosen correctly, model recognition can be easily calculated [25].

$$\langle x_1 . x_2 \rangle \rightarrow K(x_1, x_2) = \langle \emptyset(x_1) . \emptyset(x_2) \rangle$$

When are obtained, the problem is solved for a simple linear scenario in which data is separated by a hyperplane. The Kenral trick allows SVM's to convert the nonlinear data into linear where separated by linear boundaries [18].

The main idea consists of several parts:

- The algorithm is expressed using only the inner products of data sets. This is also called as dual problem. For a hyperplane test $\boldsymbol{w}'$ and $\boldsymbol{b}$ this can be done by recognizing that itself is always some linear combination of the data x_k,

$\boldsymbol{w} = \sum \lambda_k \, x_k$, so

$$\boldsymbol{w}' \cdot \boldsymbol{x} = \sum \lambda_k x_k \, x$$

The main data is passing through non-linear maps in order to form new data related to the new dimensions, so through adding to each data vector a corresponding product of some dimensions of the original data [18].

Instead of doing the inner product on these new, larger vectors, think of storing the inner product of two elements x_j', x_k in a table $k(x_j, x_k) = x_j' x_k$, so now the inner product of these large vectors is just a table lookup. But instead of doing this, just "invent" a function $k(x_j, x_k)$ that could represent dot product of the data after doing some nonlinear map on them. This function is the kernel [18].

Kernel Function:

The idea of the kernel function is to perform operations in the input space and not in the space of potentially large characteristics. Therefore, the internal products do not have to be evaluated in the functionality space.

So we are in need to the function to map attributes from the input area to the feature area. The kernel function plays a crucial role in SVM and its performance. It is based on the reproduction of the Hilbert Spaces nucleus [24, 28].

$$K\,(x, x') \;=\; \langle \emptyset(x)\,, \emptyset(x')\,\rangle$$

If K means a symmetric positive definite function, which can satisfy Mercer's Conditions, then:

$$K\,(x, x') \;=\; \sum_{m}^{\infty} \alpha_m\, \emptyset_m(x) \emptyset_m(x')\,, \quad \alpha_m \geq 0$$

$$\iint K\,(x\,, x')\; g(x) g(x') dx\; dx' \;> 0\,, \qquad g \in L_2$$

The kernel thus represents a legitimate internal product in the functional field. The training set cannot be linearly separated into an input space. The training set can be linearly separated in the functionality space. This is known as a kernel trick.

The different functions of the kernel are listed below:

• Polynomial: Polynomial mapping is a popular method of nonlinear modeling. The core is generally preferable because it avoids problems when the bags go to zero [36].

$$K\,(x, x') \;= \langle x\,, x'\,\rangle^{d}$$

$$K\,(x, x') \;= (\langle x, x'\,\rangle + 1)^{d}$$

• Multi-Layer Perceptron: The long established MLP, with a single hidden layer, also has a valid kernel representation [28].

$$K\ (x\quad ,x')\ =\ tanh(\rho\langle x, x'\rangle + e\ )$$

2-4: Applications and Related Work

Segmentation of blood vessels performs with different methods, which can be divided into supervised and unsupervised approaches. The other techniques are pattern recognition, adaptive filtering, and the vessel monitoring, mathematical morphology, parallel or hardware methods. Supervised techniques require picture to be manually segmented to train the system and then test the system [25].

Unsupervised approaches use intrinsic models to perform segmentation. The abnormal blood vessel configuration can help to identify various diseases [25].

SVM has a proven track record when it's in use for model classification issues. Once we apply the support vector approach to a given practical issues involves the solving of a series of questions according to the issues definition and associated object. It shows that the big challenges are choosing the right kernel for the particular application.

There are standard options like a Gaussian or polynomial kernel. However, if they are inefficient or if inputs are of discrete structure, the more complex kernels are required. The kernel provides the description language that the computer uses to display data. After selecting the kernel

and optimization criteria, the key components of the system are in place [25].

Traditional classification approaches work weakly when it is working directly due to the high dimensionality of the data. However, the supporting of vector machines can avoid errors in very large representations. Similar approach to these kind of techniques described for text categorization which can be also using for the task of classifying images and, as in this case, linear machines with hard edges can often generalize well.

The first real-world task on which Support Vector Machines were tested was the problem of hand-written character recognition. Furthermore, multi-class SVMs have been tested on these data. It is interesting not only to compare SVMs with other classifiers, but also to compare different SVMs amongst themselves. They turn out to have approximately the same performance, and furthermore to share most of their support vectors, independently of the chosen kernel. The fact that SVM can perform as well as these systems without including any detailed prior knowledge is certainly remarkable [25].

A quick review of some related works:

• S. Roychowdhury, et al. [27] proposed a computer-aided screening system (DREAM) that evaluate fundus images with varying illumination and range of view, and introduce a severity grade for diabetic retinopathy (DR) using machine learning . Classifiers like K-nearest neighbor (KNN),

Gaussian Mixture Model (GMM), support vector machine (SVM), and AdaBoost are analyzed for classifying retinopathy lesions from non-lesion. GMM and KNN classifier are considered to be the best classifiers for bright and red lesion allocation, respectively.

• B. Zohra, et al. [27] proposed a computer-based approach for detecting diabetic retinopathy stage using color fundus images. The features are eradicated from raw images using image processing method and then afterwards it is transferred to the support vector machine (SVM).

• Berrichi, et al. [3] proposed a computer based approach for detecting diabetic retinopathy stage using color fundus images. The features are eradicated from raw images using image processing method and then afterwards it is transferred to the support vector machine (SVM).

The major strengths of SVM are: training is relatively easy, no local optimal, unlike in neural networks. SVM scales relatively well to high dimensional data and the trade-off between classifier complexity and error can be controlled explicitly. The weakness includes the need for a good kernel function.

Chapter Three

Implementing SVM in Retinal Segmentation

3-1: Introduction

3-2: Implementing the SVM

3-3: Software description

3-4: Performance metrics

3-5: Convolutional neural networks (CNNs)

Chapter Three

Implementing SVM in Retinal Segmentation

3-1: Introduction

Retinal imaging has been widely used in medical society to diagnose vascular and non-vascular diseases [29]. They provide data in changes in the current vascular structure of the retina, which are familiar in diseases such as diabetes, occlusion, glaucoma, hypertension, cardiovascular disease, and stroke [30, 31]. These diseases generally alter the reflectivity, tortuosity, and configuration of blood vessels [32]. For instance, hypertension modifies the angle of branching or tortuosity of vessels [33] and diabetic retinopathy can cause neovascularization; develop new blood vessels. If left untreated, these conditions can lead to visual disturbances and even blindness [34]. Early detection of these changes is important to take preventive measures, and therefore severe vision loss also prevented [35].

Automatically segmentation for the retinal blood vessels from retinal picture could be extremely power tool for medical diagnosis. To do this, the segmentation method used must be as accurate and reliable as possible. The main purpose of segmentation is to distinguish an object of interest and the background of the picture.

Manual segmentation of the retinal blood vessels is exhausting and time-consuming, and making a detailed segmentation can be challenging if the complexity of the vascular network is too high [32]. Thus, automated segmentation is valuable, as it decreases the time and effort required, and in the best scenario, an automated algorithm can provide as good or better segmentation results as an expert by manual labeling [34]. For practical applications, it would be better to have algorithms that do not critically depend on configuring many parameters so that also non-experts may utilize this technology with ease [11].

Automated blood vessel segmentation has 3 faced challenges related to low contrast in images, wide range of vessel widths and variety of different structures in retinal images such as retinal image boundaries, optic disc and retinal lesions caused by diseases [11]. Hence, many authors have proposed several different blood vessel segmentation approaches based on different techniques. Even though, different methods are available for retinal segmentation, there is still space for improvement.

Retinal blood vessel segmentation algorithms are primarily focused on automatic detection linked to diabetic retinopathy that considered the leading cause of blindness in recent years. Vision loss associated with diabetic retinopathy can be avoided if the disease is detected early [12].

3-2: Implementing the SVM

Actually, automatic detection of the blood vessels in retinal images is a challenging task; the contrast of retinal image dimension as distance of a pixel from the center of the retinal image increases, presence of noise, variability of vessel width, presence of some pathological lesions, all make the task more difficult.

There are three basic approaches for automated segmentation of blood vessel: thresholding method, tracking method and machine trained classifiers. In the first method, many of different operators are used to enhance the contrast between vessel and background, such as Sobel operators, Laplacian operators, Gaussian filters which model the gray cross-section of a blood vessel [13]. Then the gray threshold is selected to determine the vessel. And this gray threshold is crucial, because small threshold induces more noises and great threshold causes loss of some fine vessels, so adaptive or local threshold is used to different sections of the image.

Vessel tracking is another technique for vessel segmentation, whereby vessel centre locations are automatically sought along the vessel longitudinal axis from a starting point to the ending point [14]. This method may be confused by vessel crossings and bifurcations.

3-3: software description

Computer and human vision

Computers and human vision seem to have the same function and purpose. Both systems interpret multidimensional geographic data to obtain information and visual identification. Due to the complexity of the human visual system and the lack of a full understanding of the neural aspects of human perception, the best computerized visual system cannot replicate the human eye.

Therefore, the development of image processing technology focuses on improving the human visual model rather than replicating it. Vision is the ability to selectively focus on interesting things. The human eye can see thousands of shades of color and about twenty shades of gray. Human vision is a complex system that absorbs and stimulates visual stimuli in the form of light. For example, as shown [13], the human brain compensates for the loss of visual acuity by emphasizing the boundaries between regions of different intensities.

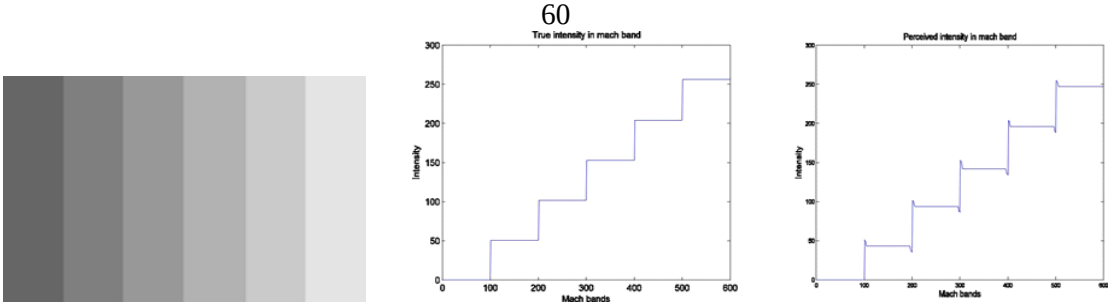
The guides shown in Figure 3: 1a, often referred to as the Mach band, are composed of several bands represented by a constant density in increasing order. Although each strip has a constant intensity, it seems to vary slightly from light to dark at the edges. In other words, a slightly lighter edge appears on one side and a slightly darker edge appears on the other side from each of the inner edges.

The slope within each strip does not exist but is perceived by the human eye. Figure 3: 1b shows the actual power levels, while Figure 3: 1c shows the perceived power values. The visual perception of Mach bands at boundaries between intensity levels is considered to help border recognition to compensate for decreased visual acuity. This phenomenon is especially important in the field of medical diagnosis, where an intensity gradient can occur in an image where the image data do not support this observation.

Human vision, enhanced by a brain that specializes in ensuring human survival, can have different perceptions and images than computer vision, which relies on an algorithm designed to process information from separate pixels into image data. .

The use of expertise makes the criteria autonomous by drawing conclusions that may contain information that does not really exist in the image data.

This can be an advantage or a disadvantage depending on the experts and skills. In all cases, combining the subjective information with the quantitative diagnostic imaging data provides the specialist with objective information on the outcome of the diagnosis [13].



(a) (b) (c)

Fig. 3. 1: Mach band example, (a) Mach band, (b) true intensity profile in Mach band, and (c) perceived intensity in Mach band

Image data structure

Digital images are made up of a series of dots or picture elements, commonly called pixels, that are stored as a series of numbers. This is spatial data indexed by two spatial coordinates. Typically, the variables x and y refer to the horizontal and vertical axes of an image, respectively.

The pixel value represents the color or intensity of each pixel, and the position of the pixels in the array corresponds to their position in the image. When more than one value is needed to encode the pixel information, a multidimensional array represents the image. When encoding an RGB image, the matrix has a three-dimensional intensity: red, green and blue intensities.

Color spaces:

A color space is a notation with which we can specify colors; Human perception of the visible electromagnetic spectrum. Color manipulation is possible by changing the properties. Thus, a wide color gamut can be generated from the three RGB primary colors. In fact, it is widely used in modern displays: monitors, digital cameras and televisions [2, 3]. Each primary color has a range of values that depends on the bit resolution.

A three-dimensional cube represents an RGB color space. The value of each point in the cube represents the intensity value of each primary color. Digital systems usually use 8-bit sizes, so a range between 0 and 255 indicates the intensity of a color [14].

Methodology

By implementing the current method, the complete process separated in the following four distinct steps:

1. Preprocessing

2. Feature Extraction

3. Classification

4. Post production

Initially, a retinal imaging kit is used to create a multi-layered, trained direct-acting neural network. The preprocessing step includes removing the edges of the lighter bands by applying a morphological aperture, followed by creating an image with shadow correction to create a smooth background image. The processed image is then sent to the module, which generates five grayscale features and two momentary invariant features.

Second, the images from the test system are used to test the trained neural network (simulation). Before the simulation, the characteristics of the test model are derived following the same steps as for the formation of the neural network. The test image properties then undergo an imaging process using a trained neural network, resulting in an output image of tool segmentation.

Finally, the segmented blood vessel output image is passed through a post-processing module to fill the pixel gaps in detected vessels and to remove falsely detected vessels. For iterative filling, the 8-neighborhood around each pixel is taken into account. If the neighbors are vessel pixels, then the pixel is classified as a vessel. After classification, misclassified pixels are removed depending on the size of connected area. If the pixel's connected area consists of less than 10 vessel-pixels, the pixel is marked as a non-vessel. After post-processing steps the final output image is produced, which is used to calculate the performance measures of the

algorithm being implemented. The calculated measures are sensitivity, specificity, and accuracy.

Preprocessing:

This is the first step in this retinal image analysis system. This step is necessary because the images have little contrast compared to the background, uneven lighting, uneven noise and homogeneity. The different stages of pretreatment include:

 Image Cropping: Performed to cut out all the necessary parts of the entire image to achieve the area of interest (ROI). The image is cropped to the desired sides so that the vessels to be segmented are clearly visible. See Figure 3-2.

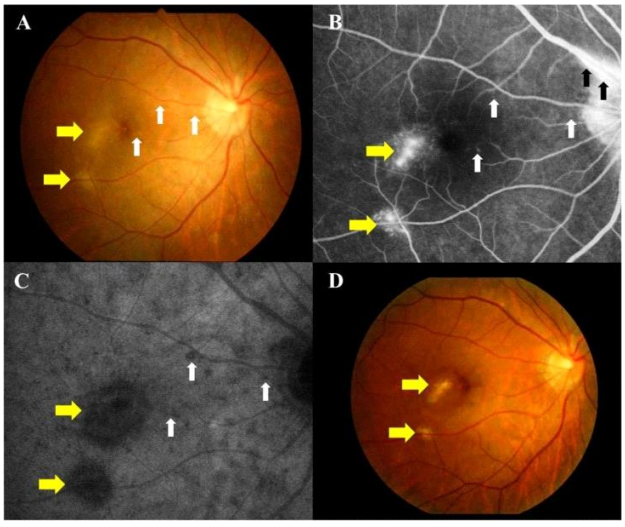


Fig. 3.2: Image cropping

- **Green Channel Extraction*:*** Green channel extraction: RGB input image is divided into three channels: red, green and blue. The green plane shows the blood vessels clearly visible, than the other two channels. (see

Figure 3: 3). The algorithm is then used to extract the characteristics of the green channel.

(a) (b) (c)

Fig. 3.3: RGB components of a retinal image a) Red b) Green c) Blue

 Contrast Enhancement*:* The contrast of the extracted image can be enhanced using different contrast enhancement methods which includes different histogram equalization techniques. In this algorithm CLAHE (Contrast Limited Adaptive Histogram Equalization) is performed. It is the final preprocessing step consists on generating a new vessel enhanced image, which proves more suitable for further extraction of moment invariants based features. While bright retinal structures are removed (i.e., optic disc, possible presence of exudates or reflection artifacts), the darker structures remaining after the opening operation become enhanced (i.e., blood vessels, fovea, possible presence of micro aneurysms or hemorrhages) [44].

Training the classifier

Support Vector Machine Classifier :- For the classification task, the idea is to map the training data into a higher dimensional feature space where a separating hyperplane with the weight vector (w) and the bias (b), can be found to maximize the margin from the closest data points (see chapter 2) .

Feature Extraction

Feature extraction is a special form of dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be significantly redundant (much data, but not much information) then the input data will be transformed into a reduced representation set of features.

 Gray-level based feature: - **Grayscale** is a range of monochromatic shades from black to white. Many image editing programs allow you to convert a color image to black and white, or grayscale. This process removes all color information, leaving only the luminance of each pixel.

It is calculated as the difference between the gray level value in the candidate pixel and the average statistical value representing its surroundings.

Since blood vessels are always darker than their surroundings, features based on describing gray-level variation in the surroundings of candidate pixels seem a good choice.

A set of gray-level-based descriptors taking this information into account were derived from homogenized images I_H considering only a small pixel region centered on the described pixel (x, y) . $S_{x,y}^{w}$ stands for the set of coordinates in a sized square window centered on point (x, y). Then, these descriptors can be expressed as

$$
\begin{aligned}
f_1(x,y) &= I_H(x,y) - \min_{(s,t)\in S_{x,y}^{9}} \{I_H(s,t)\} \\
f_2(x,y) &= \max_{(s,t)\in S_{x,y}^{9}} \{I_H(s,t)\} - I_H(x,y) \\
f_3(x,y) &= I_H(x,y) - \operatorname{mean}_{(s,t)\in S_{x,y}^{9}} \{I_H(s,t)\} \\
f_4(x,y) &= \operatorname{std}_{(s,t)\in S_{x,y}^{9}} \{I_H(s,t)\} \\
f_5(x,y) &= I_H(x,y).
\end{aligned}
$$

Advantages: To store a single colour pixel of an RGB colour image we will need 8*3 = 24 bits (8 bit for each colour component), but when we convert an RGB image to grayscale image, only 8 bit is required to store a single pixel of the image.

 Moment invariants-based features: In these various descriptors that are invariant to rotation, translation and scale are chosen for extracted features. Features based on moment invariants for describing small image regions formed by the gray-scale values of a window centered on the represented pixels [44].

The vasculature in retinal images is known to be piecewise linear and can be approximated by many connected line segments. For detecting these quasi-linear shapes, which are not all equally wide and may be oriented at any angle, shape descriptors invariant to translation, rotation and scale change may play an important role. Within this context, moment invariants proposed by Hu provide an attractive solution and are included in the feature vector.

To show the output of some examples by the algorithm, the segmentation of blood vessels for two images of DRIVE database are shown Fig. 3.4.

Fig. 3.4: Colored images of the DRIVE database with their output segmentation: a) RGB images #: 1, 15 b) corresponding segmentations

3-4: Performance metrics

Sensitivity, Specificity and Accuracy

There are several terms that are commonly used along with the description of sensitivity, specificity and accuracy. They are true positive (TP), true negative (TN), false negative (FN), and false positive (FP). If a pixel is proven to be a vessel, and the given diagnostic test also indicates the vessel, then the result of the diagnostic test is considered as true positive(TP). Similarly, if a pixel is proven to be not a vessel, and the diagnostic test suggests not a vessel also, then the test result is true negative (TN). If the diagnostic test indicates the presence of a vessel that actually has no vessel, the test result is false positive (FP). Similarly, if the result of the diagnosis test suggests that the pixel is background for a pixel which is surely a vessel, the test result is false negative (FN) [43].

Table 3. 1: Vessel classification confusion matrix

	Vessel present	**Vessel absent**
Vessel detected	True Positive (TP)	False Positive (FP)
Vessel not detected	False Negative (FN)	True Negative (TN)

Sensitivity, specificity and accuracy are described in terms of TP, TN, FN and FP as follows :

$$Sensitivity = \frac{Number\ of\ true\ positive\ assessment}{Number\ of\ all\ positive\ assessment}.$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{Number\ of\ true\ negative\ assessment}{Number\ of\ all\ negative\ assessment}.$$

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

$$Accuracy = \frac{Number\ of\ correct\ assessment}{Number\ of\ all\ assessment}$$

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP}$$

As suggested by the above equations, sensitivity is the proportion of true positives that are correctly identified by a diagnostic test, it shows how good the test is detecting an object. Specificity is the proportion of the true negatives correctly identified by a diagnostic test, it suggests how good the test is at identifying normal (negative) condition. While, accuracy is the proportion of true results, either true positive or true negative, in a population, it measures the degree of veracity of a diagnostic test on a condition.

The numerical values of sensitivity represent the probability of a diagnostic test to identify the case, in fact have the condition. A test with high sensitivity tends to capture all possible positive conditions without missing anyone. Thus a test with high sensitivity is often used to screen for disease.

The numerical value of specificity represents the probability to diagnose a particular condition without giving false-positive results. For example, if the specificity of a test is 99%, it means: when we conduct a

diagnostic test on a patient without certain disease, there is 99% chance; this patient will be identified as negative.

3-5 : Convolutional neural networks (CNNs)

Advances in AI and deep learning have enabled the speedy evolution within the fields of computer vision and image analysis; this can be all created attainable by the emergence and progress of Convolutional Neural Networks (CNNs). A CNN could be a deep learning formula that may acknowledge and classify options in pictures for machine vision. It's a multi-layer neural network designed to investigate visual inputs and perform tasks akin to image classification, segmentation and object detection, which might be helpful for autonomous vehicles. CNNs may be used for deep learning applications in healthcare, such as medical imaging [46].

There are two main parts in a CNN:

- A convolution tool that splits the varied options of the image for analysis.

- A fully connected layer that uses the output of the convolution layer to predict the simplest description for the image.

The neurons among a CNN are split into a three-dimensional structure, with every set of neurons analyzing a tiny low region or feature of the image. In different words, each cluster of neurons focuses on distinguishing one part of the image. CNNs use the predictions from the

layers to supply a final output that presents a vector of likelihood scores to represent the chance that a particular feature belongs to a definite class [46].

There are three types of layers in a convolutional neural network: **convolutional layer**, **pooling layer**, and **fully connected layer**. Each of these layers has different parameters that can be optimized and performs a different task on the input data. Features of a convolutional layer.

- Convolutional layers: The convolutional layers function feature extractors, and so they learn the feature representations of their input images. The nerve cells within the convolutional layers are organized into feature maps.

Every neuron during a feature map contains a receptive field that is connected to a section of neurons in the previous layer via a group of trainable weights, typically named as a filter bank.

Inputs are convolved with the learned weights so as to work out a brand new feature map, and also the convolved results are sent through a nonlinear activation function. All neurons inside a feature map have weights that are affected to be equal; however, totally different feature maps within constant convolutional layer have different weights in order that several options may be extracted at every location . More formally, the kth output feature map Y_k can be computed as

$$Y_k = f(W_k * x)$$

wherever the input image is denoted by x; the convolutional filter concerning the kth feature map is denoted by W_k ; the multiplication sign on this context refers to the second convolutional operator, that is employed to calculate the inner product of the filter model at every location of the input image; and $f(\cdot)$ represents the nonlinear activation operator .

- Pooling Layers: scales down the amount of data the convolutional layer generated for every feature and maintains the foremost essential information.

It absolutely was common apply to use average pooling aggregation layers to propagate the common of all the input worth's, of a tiny low neighborhood of a picture to successive layers [46].

Formally, max pooling selects the biggest part within each receptive field specified

$$Y_{kij} = max\ (p,q) \in R_{ij}\ x_{kpq}\ ,$$

wherever the output of the pooling operation, related to the kth feature map, is denoted by Y_{kij}, x_{kpq} denotes the part at location (p,q) contained by the pooling region R_{ij}, that embodies a receptive field round the position (i,j).

- Fully Connected Layers: many convolutional and pooling layers are typically stacked on top of every different to extract additional abstract

feature representations in moving through the network. The totally connected layers that follow these layers interpret these feature representations and perform of high-level reasoning [46].

Chapter Four

Analysis and Results

4-1: Introduction

4-2: Segmenting DRIVE dataset

4-3: Conclusion, difficulty and obstecals

Chapter Four

Analysis and Results

4-1 Introduction

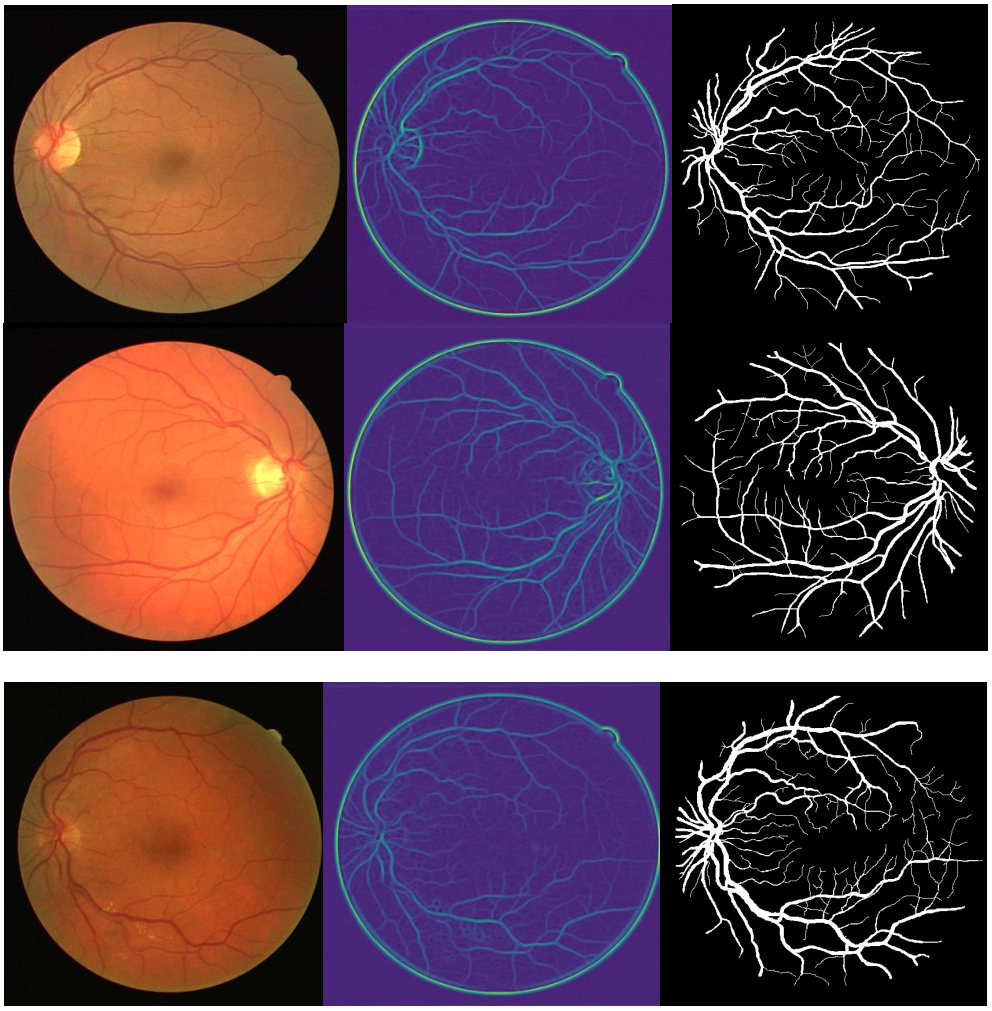
The thesis tried to review and understand the retinal blood vessels segmentation; methods, advantages, comparison between manual segmentation done by ophthalmologists with automatic methods, in addition to obstacles due to fundus images properties. As segmentation represents the first step in diagnosing eye disease, hence, its correct detection gives better chances to identify the case correctly. Huge number of algorithms are implemented to segment retina, some of them are supervised while others are unsupervised, but unfortunately, still no perfect segmentation exists. This motivated us to contribute to make a step in improving this type of segmentation.

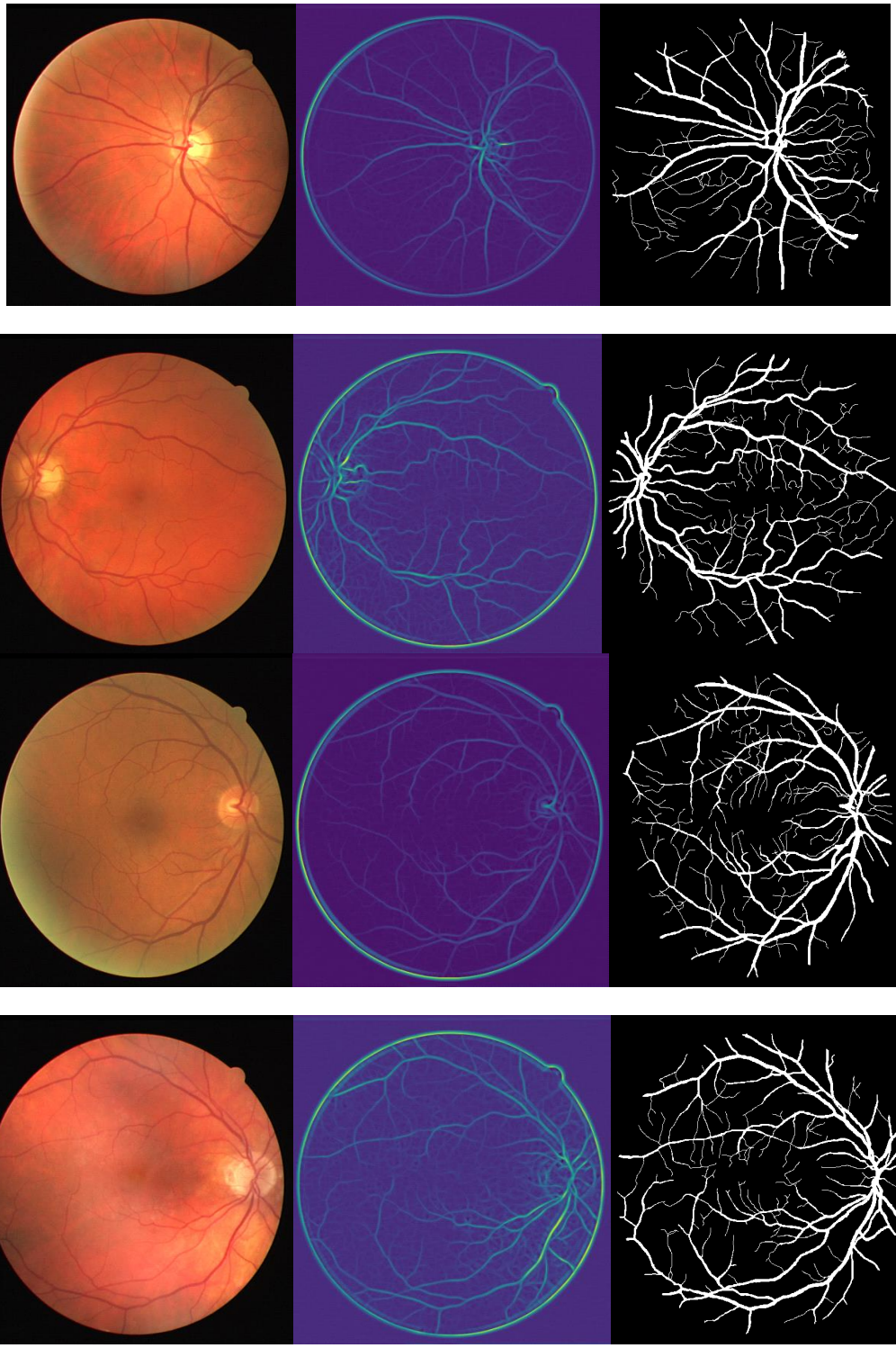
The implemented methodology was tested using retinal images from publicly available DRIVE dataset; and the results were compared against the corresponding ground truth images to calculate the performance measures of the algorithm. The supervised SVM method is evaluated in terms of sensitivity, specificity, and accuracy [44]. The results obtained for vessels segmentation method used gray-level and moment invariants-based features (see chapter 3).

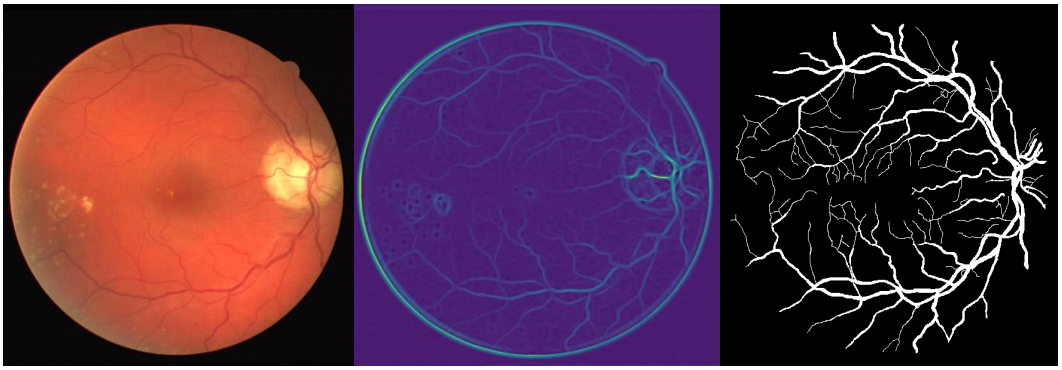
4-2 Segmentation DRIVE Dataset

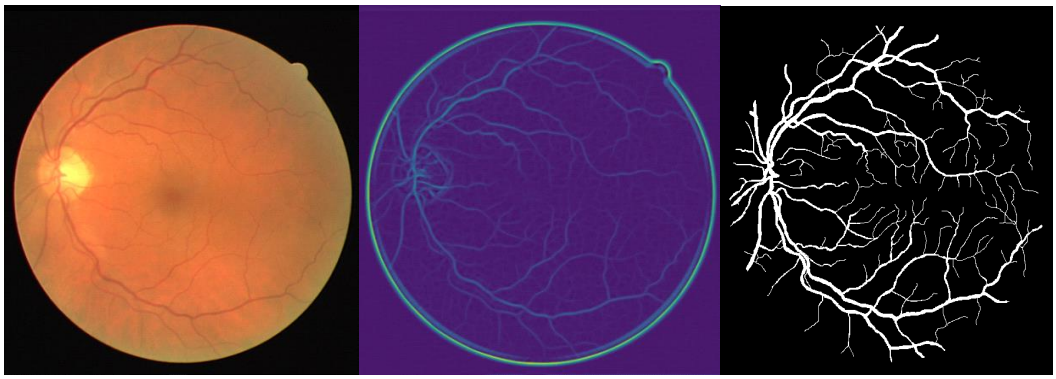
DRIVE dataset was chosen for evaluating the algorithm, this publicly available dataset has 40 fundus images divided into training and testing sets with experts segmented ground truth and being used for many state-of-the-art methods. For these reasons and more, we were motivated to select it.

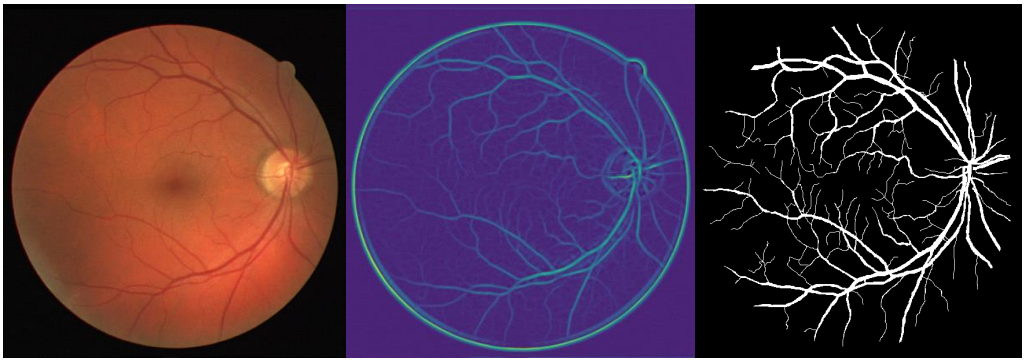
The segmented output of all DRIVE images are depicted by butting them together with their original RGB image, in addition to the neural network intermediate output image, this gives a visual comparison for the abilities of the suggested method, this can be seen in Fig. 4-1.

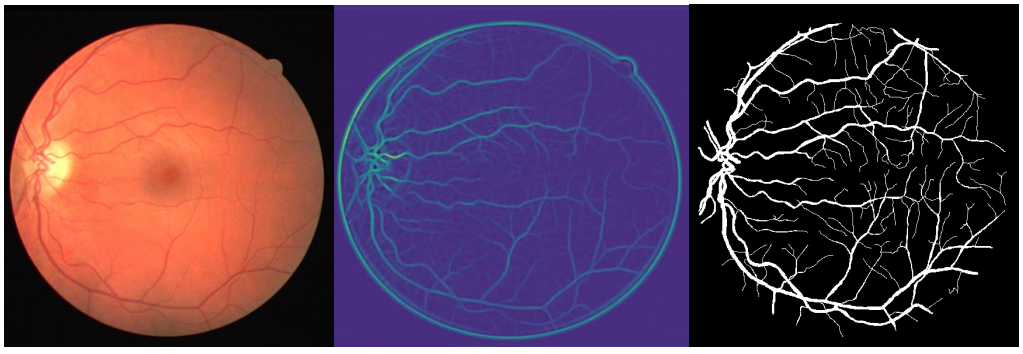


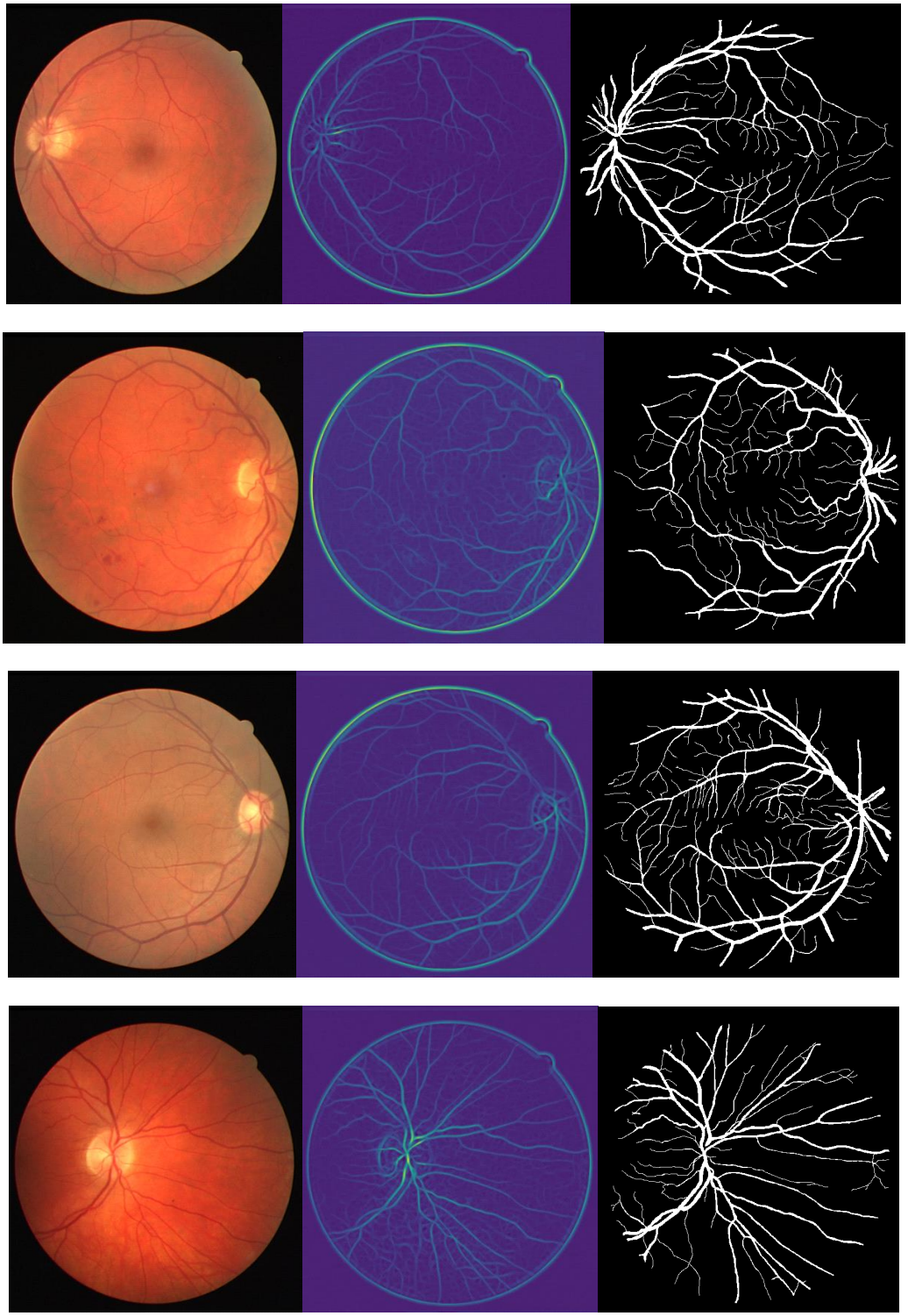


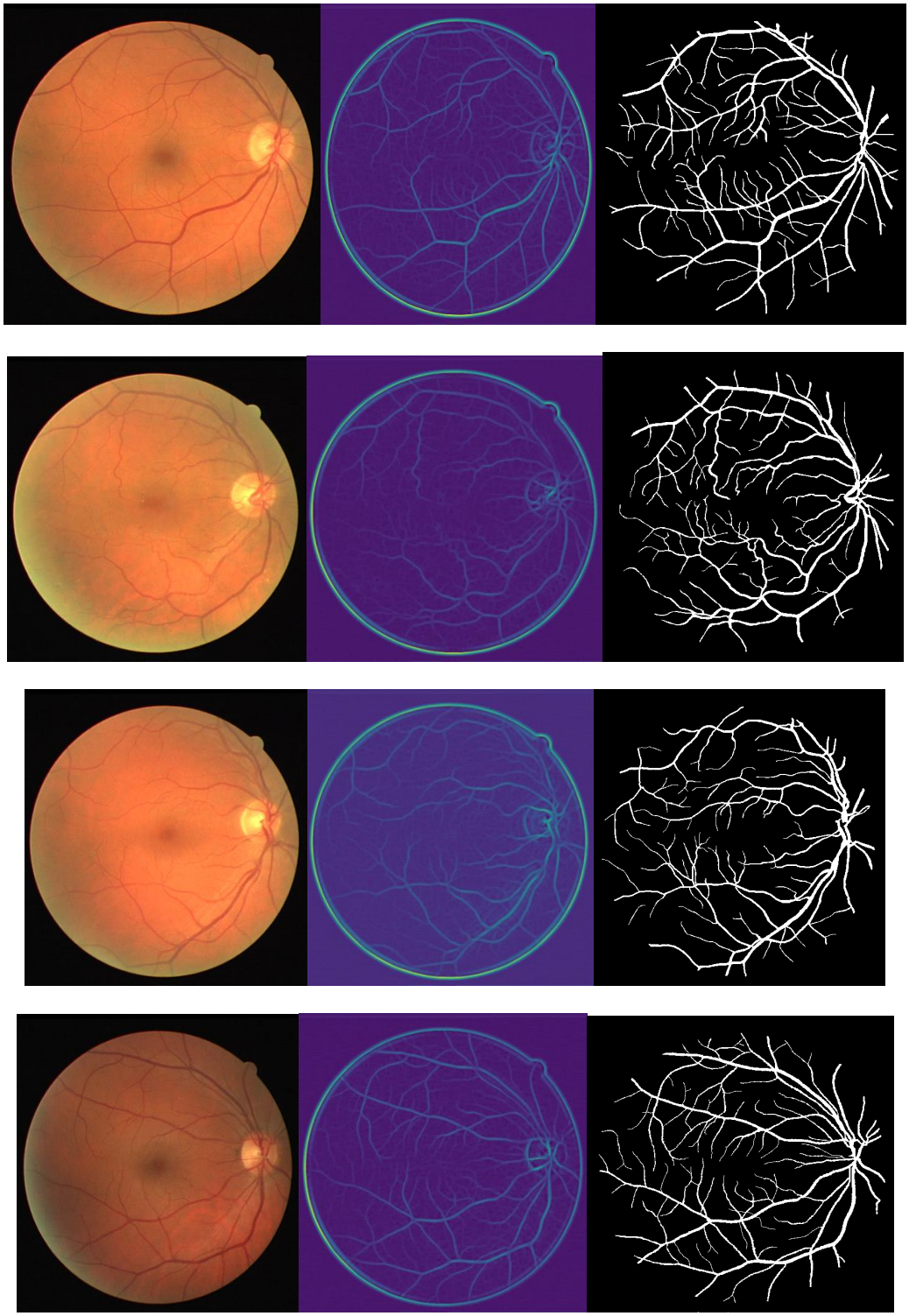


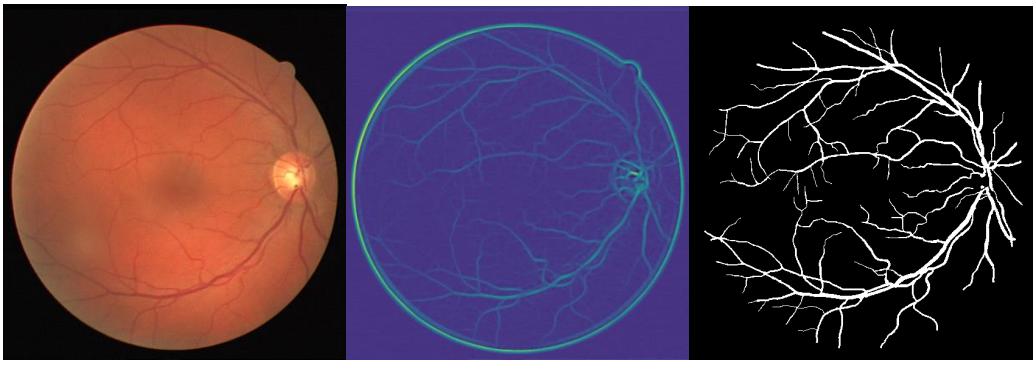












a **b** **c**

Fig. 4: 1 Segmentation of the DRIVE dataset by the proposed method:a) RGB, b) Neural network, c) segmented images

From numerical perspective, a list of values representing: true positive (TP), false positive (FP), false negative (FN), true negative (TN), and then calculating the sensitivity, specificity, and accuracy for each image of the Drive dataset, where the highest values: 96.5, 99.5, 94.8 respectively, this information is shown in Table: 4-1.

Table 4.1: Performance results for DRIVE images through the proposed method

Image #	TP	FP	FN	TN	Sensitivity	Specificity	Accuracy
1	77	25	5	290	93.90	92.06	92.44
2	12	4	9	20	57.10	83.30	71.11
3	39	14	8	363	82.98	96.29	94.81
4	37	13	22	25	59.67	62.85	60.82
5	9	16	5	89	36.00	74.70	82.35
6	141	6	46	75	75.40	98.68	92.02
7	71	21	204	935	25.80	97.80	81.72
8	8	4	25	49	24.20	93.70	66.28
9	8	6	2	10	80.00	60.00	69.23
10	3	1	4	28	42.86	96.55	86.11
11	107	51	47	641	96.48	92.71	88.42
12	16	5	5	118	76.20	95.90	93.06
13	49	7	40	60	66.10	85.70	69.87
14	8	6	2	10	80.00	60.00	69.23
15	20	13	3	15	82.60	53.50	68.63
16	10	14	7	41	58.80	90.90	70.83
17	12	6	3	40	86.60	86.90	85.25
18	26	10	6	75	81.20	88.20	86.32
19	15	1	10	184	60.00	99.46	94.76
20	21	18	12	221	63.63	92.46	88.97

Charts are added here to make the comparison easier, as you can see in Fig. 4-2, image #11 has the best segmentation from sensitivity point of view, while image #19 represents the best segmentation regarding specificity, and image #3 has the highest accuracy.

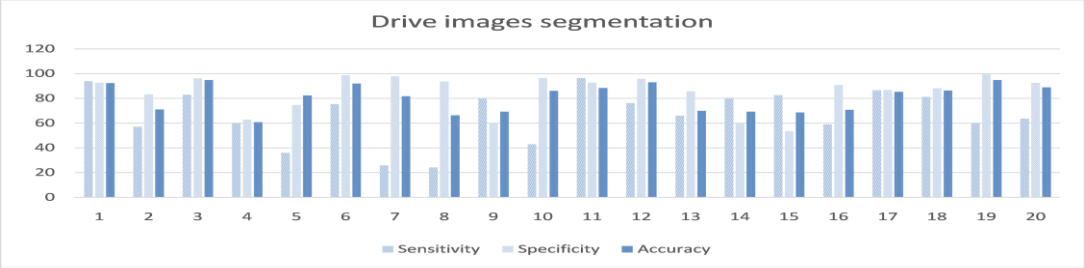


Fig. 4: 2 Representing the segmentation performance of the proposed method on Drive dataset:

4-3 Conclusion

There are available different types of blood vessels segmentation algorithms which some of them are mentioned in chapter 3, the issue addressed in this thesis is to introduce a method to segment blood vessels as accurate as possible from fundus images. To test and evaluate the proposed algorithm, fundus images are needed to train and test the method. Hence, the publicly available DRIVE dataset is selected to play an important role for evaluating and measuring the performance of the method. This dataset contains fundus images including the corresponding ground truth images. Due to its availability for public, the study and performance estimation of vessel segmentation algorithms has become easy.

When improving segmentation, it opens the horizon to develop the process of automating the diagnosis of eye diseases, which provides accuracy, quality and time, thus protecting patients from eye diseases before reaching the stage of blindness.

The proposed algorithm can be used for early detection of ophthalmic diseases as a result to its segmentation capabilities that can be done by this method. This method is in the category of supervised approach as it uses neural network for training purposes. The performance parameters given by Sensitivity, Specificity and Accuracy shows remarkable performance as compared by human ophthalmologist. The results clearly show that the proposed algorithm is effective for the segmentation of retinal blood vessels.

The algorithm has capability of detecting vessels from fundus images. As the manual segmentation of blood vessels is hard and time consuming, it would be better to use a fast, automated system which could detect efficiently various amounts of blood vessels. Besides saving time, it could decrease the number of experts required and increase the ability to segment large numbers of fundus images in a short period of time. Although the accuracy rate of algorithms is reasonable, improvements to detect the true vessel and background should be done. Hence, the automated vessel detection system is a step towards a fully trustworthy system to be achieved.

There are few areas for improving the algorithms' performance. The algorithm was evaluated using only DRIVE image dataset, which might be not enough for analyzing the algorithm to full extent. So, for algorithm evaluation, other available image databases could be included. In addition, the noise in fundus images usually lowers the overall results, if it is not handled properly. Hence, other preprocessing and post-processing strategies could be included along with the existing ones in order to improve the overall performance results. The study of these subjects is considered to be important for the future research.

4-4 Obstacles

Retinal blood vessels are divided into arteries and veins; this makes their properties different, consequently, additional variance in feature. While preparing and implementing this method some difficulties encountered the work, here are some of them:

- The available dataset consists of a limited number of images where a larger dataset is needed.

- Images in general are not of similar color properties, which makes more difficult to prepare a method to segment the vessels very correctly.

- The wide range of vessels width created another problem to the SVM to detect them.

- Light reflex in the large vessels confused the method.

References

1. A. Fathi, and A.R. Naghsh-Nilchi, '**Automatic wavelet-based retinal blood vessels segmentation and vessel diameter estimation'**. Biomedical Signal Processing and Control, vol. 8(1), pp.71-80, 2012.

2. A. M. Mendonça, and A. Campilho, '**Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction'**, IEEE Transaction on Medical Imaging, vol. 25(9), pp. 1200-1213, 2006.

3. A. Maharjan, '**Blood Vessel Segmentation from Retinal Images'**, School of Computing, Computer Science, June, 2016.

4. A. Osareh, and B. Shadgar, '***Automatic blood vessel segmentation in color images of retina'*. Iranian Journal of Science and Technology**, vol. 33, pp. 191–206, 2009.

5. B. Fang, and W. Hsu Lee, '**On the Detection of Retinal Vessels in Fundus Images**', 2003.

6. B. Tyson Smith, '**Lagrange Multipliers Tutorial in the Context Support Vector Machines'**, Memorial University of Newfoundland, Canada. June 3, 2004.

7. B. Tyson Smith, Candidate, '**Lagrange Multipliers Tutorial in the Context Support Vector Machines',** Faculty of Engineering and Applied Science, Memorial University of Newfoundland, Canada, June 3, 2004.

8. B. Wasan Cerutti, S. Ford, and R. Marsh, '**Vascular network changes in the retina with age and hypertension'**, vol. 13(12), pp. 1724-1728, 1995.

9. C. Burges, '**A tutorial on support vector machines for pattern recognition in Data Mining and Knowledge Discovery'**, Kluwer Academic Publishers, Boston, vol. 2, 1998.

10. D. Hubel, '**Eye, Brain, and Vision',** W. H. Freeman & Co., New York, vol. 22, p.240, 1988.

11. D. Mukherjee, Y. Deng, and S. K. Mitra, '**Region-based video coder using edge flow segmentation and hierarchical affine region matching'**, Proc. SPIE 3309, Visual Communications and Image Processing '98, 338 ,January 9, 1998

12. D. Phillips, **'Image Processing in C' .CMP**; ebook (2nd addition), pp. 125-132, 26 April 2000.

13. E.J Sussman, W.G Tsiaras, and K.A Soper, '***Diagnosis of diabetic eye disease'***. **The Journal of the American Medical Association**, vol. 247(23), pp. 3231-3234, 1982.

14. H.Gomez-Moreno, S.Maldonado-Bascon, and F.Lopez-Ferreras, '**Edge detection in noisy images using the support vector machines'**, pp. 685–692, 2011.

15. H. D. Cheng, X. H. Jiang, Y. Sun, and J. L. Wang, '**Color image segmentation: advances and prospects'**, Pattern Recognition, pp. 2259-2281, 2001.

16. H.R Taylor, and J.E Keeffe, '***World blindness: A 21st century perspective'*, British Journal of Ophthalmology**, vol. 85, pp. 261-266, 2001.

17.J. Allen, Y. Aloimonos, and TH. Dean, **'Artificial Intelligence'**, Addison-Wesley Publishing Company, Menlo Park, California, 1995.

18. J. Gao, J. Zhang, and M. G. Fleming, '**A novel multiresolution color image segmentation technique and its application to dermatoscopic image segmentation'**, Image Processing, vol.3, pp.408-411, 2000.

19. J. Soares, J. Leandro, R. Cesar, and M. Cree, '**Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification',** IEEE Transaction of Medical Imaging, vol. 25, pp.1214–1222, 2006.

20. J.F.Canny. '**A computation approach to edge detectors'**. IEEE Transactions on Pattern Analysis and Machine Intelligence, pp.34–43, 1986.

21. J.P.lewis, '**A Short SVM (Support Vector Machine) Tutorial'**, CGIT Lab / IMSC , U. Southern California , 2004.

22. K. A. Thomas, '**Image processing as applied to medical diagnostics'**, MSc dissertation, University of Oregon, 2010.

23. M. Al-Rawi, M. Qutaishat, and M. Arrar, '**An improved matched filter for blood vessel detection of digital retinal images',** pp. 262–267, 2007.

24. M. Batterbury, and B. Bowling, '**Opthalmology',** British Library Cataloguing in Publication Data , London , 2005.

25. M. Kadhm, Asst. Prof. Dr. A. Abdul Hassan '**Handwriting Word Recognition Based on SVM Classifier'**, Computer Science Department, University of Technology/ Baghdad, vol. 6, No. 11, 2015.

26. M. M. Mushrif, and A.K. Ray, '**Color image segmentation: Rough set theoretic approach',** Pattern Recognition. Lett, pp. 483–493, 2008

27. N. Cristianini and J.Shawe-Taylor, '**An Introduction to Support Vector Machines and Other Kernel-based Learning Methods'**, Cambridge University Press, 2000.

28. N. P. Emptage, D.M. S. Kealey, and F. C. Lum, **'Idiopathic Macular Hole'.** American Academy of Ophthalmology. September 20, 2014

29. N.Ahuja, and M .Yang. '**A Geometric Approach to Train Support Vector Machines'**. Proceedings of the 2000 IEEE Conference on Computer Vision and Pattern Recognition , Hilton Head Island vol. 1 , June, 2000.

30. O. Bousquet, S. Boucheron, and Gabor Lugosi, '**Introduction to Statistical Learning Theory'**, Max-Planck Institute for Biological Cybernetics Spemannstr. 38, D-72076 Tubingen, ¨ German, pp.175 – 178, 2007.

31. O. Rotem, H. Greenspan, and J. Goldberger, '**Combining region and edge cues for image segmentation in a probabilistic gaussian mixture framework'.** IEEE Conf. on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, pp. 17–22, June 2007 .

32. P. Ganesan, V. Rajini, '**Segmentation and edge detection of color images using CIELAB color space and edge detectors'**, Emerging Trends in Robotics and Communication Technologies (INTERACT),pp. 393-397, 2010.

33. R. O. Duda, P. E. Hart, and D. G. Stork, '**Pattern Classification'**, John Wiley, New York, 2001.

34. R.C. Gonzalez, R.E. Woods, '**Digital Image Processing'**, Addison Wesley, New York, 3rd ed., 1993.

35. S. Ertekin, **'Learning in Extreme Conditions: Online And Active Learning With Massive, Mbalanced And Noisy Data'**, A Dissertation in Computer Science and Engineering, 2009.

36. S. Marsland, '**Machine Learning: An algorithmic Perspective'**, A chapman & hall /CRC Press. New York, 2009.

37. T. Q. Chen, and Y. Lu, '**Color image segmentation: an innovative approach'**, Pattern Recognition, pp. 395–405, 2002

38. Th. Evgenuiu and M. Pontil, **'Statistical Learning Theory: a Primer'**, vision 38, New York. 1998.

39. U.T Nguyen, A. Bhuiyan, L.A Park, and K. Ramamohanarao, '**An effective retinal blood vessel segmentation method using multi-scale line detection'**, Pattern Recognition, vol. 46, pp. 703-715, 2013.

40. V. Jakkula, 'Tutorial on Support Vector Machine (SVM)', **School of EECS,** Washington State University, Pullman 99164, 1999.

41. V. Vapnik, S. Golowich, and A. Smola. '**Support vector method for function approximation, regression estimation, and signal processing'**, In M. Mozer, M. Jordan, and T. Petsche, editors, Advances in Neural Information Processing Systems 9, pp. 281– 287, Cambridge, MA, 1997.

42. W. Zhu, N. Zeng, Ning Wang, '**Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical'**, Health Care and Life Sciences Health Care, 2010.

43. X. You, Q. Peng Yuan ,Y. Cheung, and Y. Lei, '**Segmentation of retinal blood vessels using the radial projection and semi-supervised approach'**, Pattern Recognition, vol. 44, pp. 2314-2324, 2011

44. Y. Xu, H. Zhang, and H. Li, **'An improced algorithm for vessel centerline tracking in coronary angiograms',** pp. 131–143, 2007.

45. Y.W. Lim and S.U. Lee, '**On the color image segmentation algorithm based on the thresholding and the fuzzy C-means techniques'**, Pattern Recognition, vol. 23, no.9,pp. 935-952, 1990 .

46. Z.Wang and W. Rawat, **Department of Electrical and Mining Engineering, University of South Africa**, Florida 1710, South Africa, pp. 2355 – 2357 , 2017 .

جامعة النجاح الوطنية

كلية الدراسات العليا

النهج الرياضي الأمثل في تجزئة شبكية العين من خلال تصنيف آلة الدعم الموجه

إعداد

إسلام إبراهيم أحمد صويلح

إشراف

د. هادي حمد

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

2021

النهج الرياضي الأمثل في تجزئة شبكية العين من خلال تصنيف آلة الدعم الموجه

إعداد

إسلام إبراهيم أحمد صويلح

إشراف

د.هادي حمد

الملخص

ان عين الإنسان تعتبر جزءا حساسا وعرضة للإصابة للكثير من الأمراض .ونظراً لكونها جزءا مهما من الجسم وتحتاج إلى عناية مركزة، نجد ان هناك العديد من الخوارزميات يتم اقتراحها لتشخيص أمراض العيون والكشف المبكر عنها، وتعد معالجة الصور احدى تلك الأدوات الأساسية لمساعدة الأطباء في تحقيق هذا الهدف.

في هذه الاطروحة نركز على شبكية العين نظراً لان لها دورًا مهمًا في آلية رؤية الأشياء، فهي تتكون من العديد من الأجزاء مثل القرص البصري والشرايين والأوردة والبقعة .وتتأثر الأوعية (الشرايين / الأوردة (بعوامل مختلفة من شأنها ان تعرض العين للإصابة بأمراض بصرية.

إن هذا العمل يحتوي على طريقة اتوماتيكية لتحديد تجزئة شبكية العين اعتمادًا على" آلة الدعم الموجه (Support Vector Machine) "التي من خلالها تساهم في تحديد الاوعية الدموية في شبكية العين والتي تساعد في تشخيص الاعتلال الحاصل فيها بناء على تغيير شكلها مقارنة مع صور حقيقية سليمة، الميزات التي تساعدنا في هذا التصنيف هي ميزة تستند إلى مستوى الرمادي (Gray-level based feature) وميزات تستند إلى لحظات المتغيرات (Moment invariants-based feature). تم تنفيذ الخوارزمية على قاعدة بيانات DRIVE وهي قاعدة بيانات عامة على الإنترنت من خلال كتابة البرنامج على Python.

تم تقييم ومقارنة أداء الخوارزمية بنتائج حقيقية من قاعدة البيانات "DRIVE" باستخدام الحساسية والنوعية والدقة مع إعطاء النتائج 66.5 % ، 85.1 %، 80.6 % .