DEPARTMENT OF ELECTRICAL ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE

I, SMRITI CHHABRA, Roll No. 2K12/C&I/019 student of M. Tech. (Control & Instrumentation), hereby declare that the dissertation titled "SORTING OF RETINAL BLOOD VESSELS USING NEURAL NETWORKS" under the supervision of Dr. BHARAT BHUSHAN, Associate Professor of Electrical Engineering Department Delhi Technological University in partial fulfilment of the requirement for the award of the degree of Master of Technology has not been submitted elsewhere for the award of any Degree.

Place: Delhi

Date: 31.07.2014

Smriti Chhabra

(2K12/C&I/19)

Dr. Bharat Bhushan

Associate Prof., EED, DTU

ABSTRACT

In recent years, computation techniques are emerging with a good speed in the medical field. A significant branch of biomedical field is said to be Ophthalmology; it needs automated techniques for identification of pathology.

In health care units, since new technologies are increasing, their goal is to reduce the visits to specialised doctors. They aim at efficiency of doctor which means how many cases a doctor can deal with proper diagnosis. Also, it might reduce the cost of overall procedure for both patients and clinic.

Diabetic retinopathy being one of the diseases which if encountered with at later stage might cause serious problems like blindness. So, we need to detect it as soon as we can, and this is done with the help of segmentation of retinal images. There are many retinal diseases which are characterized by modification in retinal vessels. Retinal blood vessels are of two types: arteries and veins. Irregular wide veins are the symptom of Diabetic Retinopathy which leads to low Artery to Vein ratio(AVR).

The technique used here is implementing a supervised classification for retinal vessel detection. These vessels go through few stages like pre-processing, feature extraction and hence classification is done.

The classifier used is Neural Network which is fast, has good convergence speed and properties like generalization, non-linearity and more are discussed. Two types of feed-forward Neural Network are used, Back propagation and Probabilistic Neural Network. It has to be found out which is better for this classification problem and why.

ACKNOWLEDGEMENT

I want to express my sincere thanks to the mentor of my project **Dr. Bharat Bhushan**, Associate Professor, Department of Electrical Engineering, Delhi Technological University. He has helped me from starting and has boosted me to work hard.

I sincerely acknowledge **Prof. Madhusudan Singh, HOD,department of electrical engineering**, Delhi Technological University, New Delhi and I wish to thank all faculty members who helped to finish my thesis in all aspects.

I would like to give special thanks to **Mrs. Devanjali Relan**, researcher in University of Edinburgh without whose support, it would have been difficult to do the work. She has provided me with essential MATLAB tool which was important for the completion of the project.

I want to thank **Mr. Rishi Relan**, researcher in Vrije Universiteit Brussel whose constant motivation kept me going. He helped me in understanding the basic concepts of Neural Network.

I am honour bound to my family for their continuous love and support.

Smriti Chhabra
(2K12/C&I/19)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS USED

AVR- Artery to vein ratio

NN- Neural Network

BP- Back Propagation

PNN- Probabilistic Neural Network

ANN- Artificial Neural Network

DR- Diabetic Retinopathy

PDR- Proliferative Diabetic Retinopathy

NPDR- Non-Proliferative Diabetic Retinopathy

DRIVE- Digital Retinal Images For Vessel Extraction

FOV- Field of view

KDD- Knowledge Discovery in database

SCG- Scaled Conjugate Gradient

MLP- Multi Layer Perceptron

MABP- Mean arterial blood pressure

Pdf- Probability density function

I/p- input

O/p- output

LSTM- Long short term memory

CGB- Conjugate Gradient Backpropagation

ROC- Receiver Operating Curve

CHAPTER 1

INTRODUCTION

Many systemic diseases and eye diseases show prominent changes in retina. Cardiovascular diseases like stroke and coronary heart disease are threatening the human life worldwide [50].New studies show that narrowing of retinal blood vessels are the early makers of cardiovascular risk [31]. Artery-vein classification is required to extract biomarkers separately from arteries and veins to detect different diseases. As a number of other anatomical structures refer to the process of vision, my work would focus on classification of arteries and veins in retinal images.

1.1 RETINOPATHY

Due to gradual or little damage caused to retina of an eye leads to Retinopathy. If the patient does not know about the disease, vascular occlusion can occur over short interval of time. Vascular occlusion is a blockage of blood supply running in the vessels; it might be a clot or in any other form [51].

Blockages are classified as retinal vein occlusion or retinal artery occlusion which can be found in the centre or branches of arteries and veins. Classification depends upon if the occlusion occurs in an artery that carries blood from the heart or vein that carries the blood which is to be re-oxygenated towards the eye [51]. This blockage leads to various vision disturbances like blurred images are formed on the retina (portion of eye where image is formed) or distortion in one eye. Blood vessels hence should not obscure the retina and disturb vision and allow proper amounts of oxygen and nutrients without affecting the transparency of the retina.

Retinopathy is of various types like:

Diabetic retinopathy - due to diabetes;

Hypertensive retinopathy – high blood pressure also called arterial hypertension;

Retinopathy of prematurity – due to prematurity of the newborn;

Radiation retinopathy – due to exposure to ionising radiation;

Solar retinopathy – due to direct sunlight exposure;

Sickle cell disease – hereditary blood disorder

and many more.

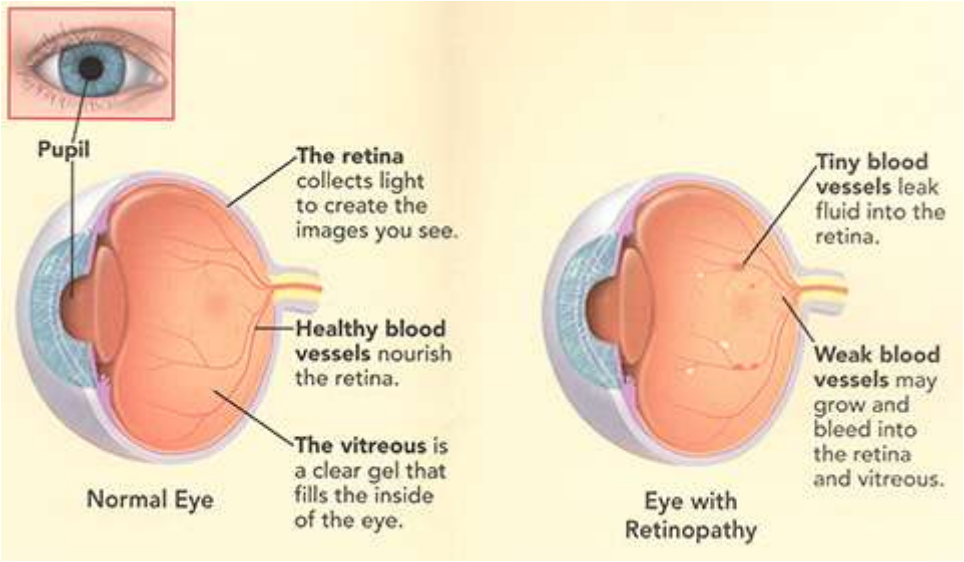


Figure 1.1: Difference between normal eye and eye with retinopathy

An ophthalmologist says malfunctioning of vasculature of eye can lead to serious diseases. A single mistake during classification of vessels propagates through and brings the whole vasculature in an inconsistent state. Hence, we should have some safe and non invasive methods to detect the disease in its earliest stage. Manually, it takes long to detect the onset of any systemic disease but we need an efficient and automatic system to analyse the retinal images so that diseases' cure can be done.

An important part in classifying retinal blood vessels is that automatic and easy computing methods should be used [44]. With the increasing availability of digital cameras, automatic analysis of such digital images may relieve little burden of retinopathy screening for ophthalmologists.

1.2 RETINAL IMAGES

The retinal microvasculature shares structure and physiological characteristics with the vessel anatomy in other parts of human body. Imaging technique as retinography,

provide non invasive view of the blood vessels present in retina. Thus, the retinal images have become a tool for the study and diagnosis of several diseases.

Since the first change in retina due to onset of retinopathy appears in vessels, changes in vessel structure can affect arteries and veins differently. As retinopathy progresses, shape of arteries and veins change strangely, example Focal narrowing for arteries and beading of veins. One of the early signs is called arteriolar narrowing, in that ratio between arteries and veins (AVR) diameters decreases [35].

An automatic method for classification of retinal vessels is based on feature extraction in major vessels. Features are extracted in regard to their colour and size since arteries are brighter and thinner than veins and show a light reflection in central part of them.

A lower AVR associates with high blood pressure, increased risk of stroke, diabetes and hypertension. The current process of calculating AVR is very tedious and highly operator dependent. Firstly the vessel type is determined, and then diameter of each vessel is done manually which is very time consuming. It also varies from one inspection to another even though same grade has done it; hence reproducibility is a major concern [31]. It normally takes around twenty minutes to analyse one image which is not feasible for large scale research studies and clinical utility. Therefore, there is a need of a more efficient system that is precise in classifying vessels automatically.

1.3 NEURAL NETWORK CLASSIFIER

In this dynamic environment, all organisms need to interact with each other in an intelligent manner and there helps our brain. This is a complex organ which can deal with a large amount of input information in an efficient manner.

To imitate such kind of machine, we have to create an artificial brain in the Artificial Intelligence field. Since the brain has large number of nerve- cells called neurons. These neurons together form a connection with each other to form a network, hence named as Neural Network.

Sub- fields of AI, study of ANN tries and models the biological neurons; after forming network with the artificial neurons.

In health care research, machine learning methods provide inexpensive diagnosis and classification. It is one of the important decision making task for real world problems.

Classification is used when we have to define for a predefined class or any other particular group attribute.

For the purpose to be fast, easy and reliable, it has to use the computer technology. In this field, one can store and process large amounts of data accurately and quickly. It can access raw data which is stored in the digital format. Many soft computing techniques are available for easy computation like genetic algorithm, neural network and fuzzy logic control.

As the brain process the information, same way an artificial neural network works. It is composed of highly interconnected processing units (neurons) together to solve a specific problem. Neural networks have been applied in many biomedical researches.

Neural networks have the ability to work on any imprecise data and to extract patterns efficiently. Its computational speed is very high. It has more characteristics like:

1. It has the ability to learn through given through given data for training (adaptive learning).

2. The network can be utilised for real time operations as computations are carried out in parallel.

3. It is self organised as it represents the state during learning.

4. If there is any fault or error in coding, it has the ability to retrain itself.

1.4 DATASET USED

The DRIVE data set is taken from Image Sciences Institute.

The drive database is established so as to depict about the classification of arteries and veins in retinal images.

1.5 LITERATURE REVIEW

1.5.1 Artificial Neural Network

The motive of any classifier is to categorize the sequence of data from pre processor into different classes. For any change that classifier does, it also includes software which is self adaptive and is trained for correct classification.

In AI (artificial intelligence), many criteria of adaptive software are developed. ANN is very successful classifier and its basis is dependent on the mimicry of the structure of the human brain. The following chapter has the detailed information of the history of evolution of ANNs.

Based on algorithms and maths, Warren Mc Culloch and Water Pitts in 1943 created a mathematical model for neural networks. This model was called as threshold logic. This model changed its way for the research of NN in two different manners. First approach was aimed at the processing of brain and its biological functions and second was related to NN in the field of AI. The first model created by them had one output and two inputs and it was found out that with one input active, a neuron cannot be active. A binary output is evaluated for input being of equal weight. The output was zero till inputs were summed to give a threshold value [1].

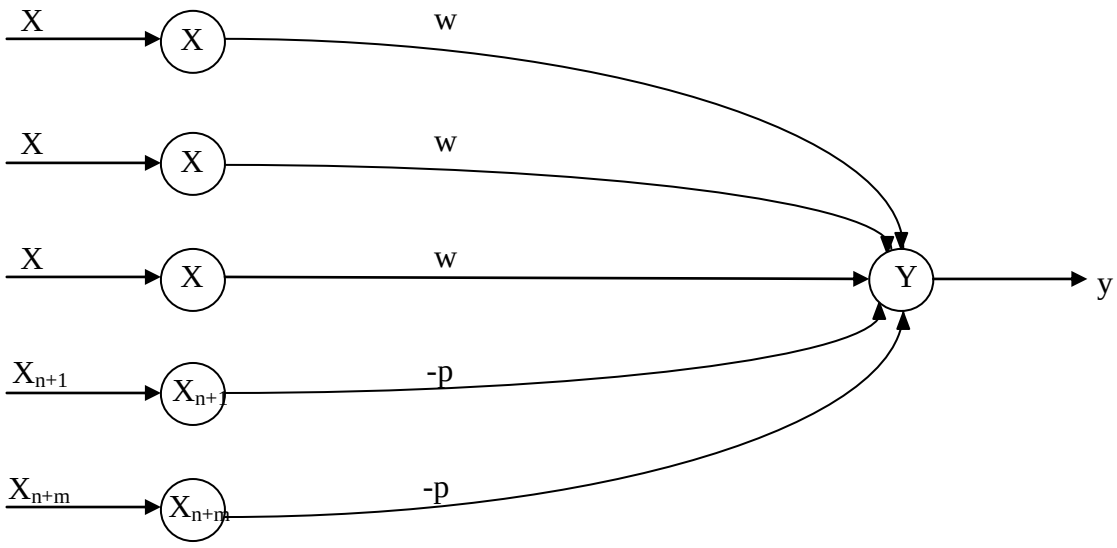


Figure 1.2: McCulloch- Pitt's neuron model

Donald Hebb was the psychologist who created a learning method based on assumption in late 1940s said to be as Hebbian Learning. Since the method is entirely grounded on suppositions hence it is considered under the category of unsupervised learning. In 1948, ideas were in use of practice.[3]

In 1954, Farley and Wesley A. Clark for the very first time were using methods to simulate the Hebbian network (at MIT). They started with computational machines followed by calculators [4].

In 1956, many other machines for computation were created by Duda, Habit, Holland et al.[5]. Perceptron was then created in 1958 by Frank Rosenblatt[6]. This algorithm was based on two layers and doing simply addition and subtraction was started being applied for Pattern Recognition. He applied the method of trial and error and thus interconnected perceptrons were changing weights randomly. But we use the model of Mc Culloch and Pitts' as it works like a voltage to frequency translator because of electrochemical process of neuron. It fires at higher frequency as get discharged by the chemical reaction and thus forming a threshold value..

In 1960, a mathematical method was developed for adopting weights by Widrow and Hoff. At that time gradient search method was used so that the squared error minimizes and later this algorithm was known as Least Mean Square.

In 1969, research on NN deteriorated when Marvin Minsky et al.[8] published on machine learning. Their important issues which lead to this downfall were (i) the ex-or circuit was not able to be processed with single layer NN. (ii) NN took time and computers were not efficient to run for that long time. This issue continued till we got fast processing computers. Research was put on hold as it was said that with usage of perceptrons only linearly separable problems could be solved. Using Perceptron' 2-separable problem and n-1 nodes were considered to solve the problem. Perceptron was published but research was not funded till the time a method for n- separable problem was discovered.

At the University of Toronto, Geoffry Hinton had the BP algo and unsupervised methods [36]. These were similar to Neocognitron by Kunihiko Fukushima and were used to train highly and deep non- linear architecture in 1980.

In the mid of 1980s, "connectionist learning" came into practice. In 1984, Duckworth describes the history of computer programmers till date. He has mentioned the changes in technology and development according to change in mindset of humans. Whatever are the limitations, descriptions and usage of AI are present has also been given [15].

NN is iterative process in which firstly data cases were presented and correspondingly weights were adjusted every time. This process repeats itself till the time all weights are fixed. Thus, this is also known as parallel distributed processing. James Mc Clelland

and et al. in 1986 give the full demonstration to simulate neural processes with the help of connectionism. [14].

NN are being used in AI. They are considered as simple models with neural processing in the brain. Its also true that to what degree NN matches the brain architecture, its function is debatable.

Nguyen and Hoff developed the mathematical model for adapting weights. It was based on minimising error of squared sums. On the assumption that we got a desired response, gradient method was thus applied. LMS is the name given to this algorithm later and is used in many formulations. To minimize the error, it was provided. Computational time was decreased with Selfridge's work. But by using LMS, time was further decreased with the help of feasible perceptrons.[16]

NN were overtaken by machine learning by SVM and more in 1990s. many simple methods were introduced like linear classifiers. Then in 2000s, deep learning was introduced which created a spark again.

Learning is considered one of the important issues in research area of AI. KoCabas[17] gives the definition for learning, intelligence and discussed about how machine leraning is important now, its history and future scope. Various methods of learning are described based on knowledge. It tells us about the theory of classifier like GA has symbol level learning and NN has device level system learning, the above mentioned ways are decided according to i/p, learning algorithms or devices and o/ps. Also, discussion about limits of learning and relationship of learning and relationship of learning and knowledge representation is done with the conclusion in the end.[18]

Introduction of SCG with super linear convergence rate is done. Algorithm has its base on the class of various optimization techniques which are known in the numerical analysis like CGM. This algorithm uses information of NN 2^{nd} order system with 0(n) memory usage (n- no. of weights in network). Moller presented the comparison between performance of SCG has more speed than BP and hence more convergence and less error. An optimum step- size is also important as it reduces overall complexity of NN architecture. So if lesser complex system, weight space is correctly defined and hence the conclusion was drawn that SCG is more efficient than BP[19].

CMOS is used to create these computational devices and in fact now recent efforts tell that for very large scale functions and applications, nano- devices will be created. If this gets successful; it would be a new era for neural computing because all would be dependent on learning rather than programming.

Yang J et al. gave an approach for pattern classification with near to minimal architecture of NN. Constructive learning algorithm were presented which helped in overcoming the need for ad hoc. It also included choices for network technology which can thus search for suitable weights in earlier fixed network architectures. Few algorithms with zero convergence (under certain assumptions) to zero classification involves binary to binary mapping.[22]

David H. Hubel and Torsten Wiesel identified "standard architecture of vision"[27]. ICJNN 2011 Traffic Sign Recognition competition was won with fast processing GPU implementation of many pattern recognition contests, cells being the inspiration [28]

Human brain can solve complex visual problems like image interpretation, object recognition etc. since last few years, research is being done on the understanding of human brain. Presently, human operator assigns actions for these kind of models on which high level human vision can operate. But still there is no model present which can directly account for low- level visual i/p (i.e. pixels). So after understanding this issue, Pulkit Agarwal and et al. explored two types of models computer vision and machine learning. It is found out that both these classes can correctly predicts the brain activity in high-level visual areas and that also without any need of hand annotation or tags for images. It can be directly done from pixels and first time the mapping is obtained. Thus its an area where more exploration can be done regarding brain function in both machine learning and computer vision.[29]

1.5.2 History of Retinal Images and their use

Eye has an optimal property which do not allow for the direct depiction of retina. When at an appropriate angle, light is shined so blurred reflection making pupil appears red known as red reflex and is known for centuries. So, a need of technique arose which could get a focus image on the retina.

French physician Jean Mary was the first who examined over a live cat in water; the vessels of retina were visible outside. In 1823, Czech scientist Jan Evanelista Purkinje

invented the ophthalmoscope as its need arose. Charles Babbage did its reinvention in 1845. He also found the link between programmable computer and retinal imaging. Again in 1851, Helmholtz reinvented the instrument[30]. Finally it became famous among ophthalmologists and first retinal images were published in 1853 by Trigt.

This was an attraction for all physicians knowing infectious diseases they wanted to examine through eyes. In 1891, Gerloff[33] showed images of retina depicting blood vessels. Gullstrand in 1910 developed fundus camera[34] which is used till date. The fundus imaging method was cost- effective, safe and primary method of retinal imaging.

Focussing on vessel segmentation for the first time, Matsui et al. were the first to give a method for retinal image analysis[35]. In 1984, first method to test segment's abnormal structure was published by Baudoin et al. It described about microaneurysms, cause of diabetic retinopathy [36].

In 1990s, the field was vastly changed with the development and expansion of this area.

Since its difficult to access microcirculation and direct visualisation; methods were developed in 2008 to measure and check abnormalities in the retinal vasculature with the help of computer- assisted image and retinal photographic imaging techniques. The final assessments from these methods always help in clinical, cardiovascular, metabolic outcomes etc. Hence it can be said that Retinal Vascular Imaging is a research tool which play a major role in cardiovascular risk. Liew G. and et al. reviewed various applications where research is going on.[37]

Susan Lewallen et al. have made the effort collectively to pen down the ocular features in many populations in 1999. Ocular fundus pathology is common and has importance in prediction of symptoms. 735 patients were examined in Kenya and gambia with dilated patients pupils by the method of direct and indirect opthalmoscopy. On this basis, 5 clinical features retinal whitening, unique vessel abnormalities, cotton wool spots, papilloedema and haemorrhages are found out. So, pictures of these descriptions are presented. [38]

In 2003, alterations in vessel network of retina were firstly observed. Changes in arteries and veins affected the retinopathy and showed whether the disease is mild or strong. So an automatic method for diagnosis of retinopathy was required by distinguishing arteries and veins. E. Grisan et al. developed a new algorithm to classify

vessels exploiting the peculiarities of retinal images. The method of a divide et imperia was used which performed robust classification analysis. The final results are based on the features of arteries and veins were compared with manual classification. For 443 vessels, error of 12% was found.[39]

D. Relan et al. in 2013 classified the retinal vessels into arteries and veins based on the coloured features. Gaussian Mixture Model, unsupervised classifier i.e. GMM-EM method was used. From 35 images, 406 vessels were processed and 92% were correctly classified. Result of this classification was compared with two trained human operators. It was done so as to validate the obtained result. Compared to the previous methods used for the classification, this yielded the best results[40].

Generalized and focal arteriolar narrowing, retinopathy and any cuts are the abnormalities which indicate vascular damage because of aging, hypertension etc. Studies say that these can be observed in 2-15% of non diabetic population and are actually associated with increase in blood pressure. T.Y Wong believes that arteriovenous nicking, narrowing seem to be markers of hypertension of both past and present blood pressure levels. There is no direct link of these strokes, cardiovascular molarity, atherosclerosis or ischemic heart disease but only with the help of data it can be said. In 2001, many new technologies and techniques are being worked upon which would help in future telling us about the risk indicators for cerebrovascular diseases.[41]

Retinal ganglion cells and their axons are served by retinal blood vessels for nutrition. To measure the diameter of healthy and glaucoma eyes, Nguyen et al. studied the calibres of temporal blood vessels in 1989. They evaluated at optic disc border; 2mm from optic disc centre. Out of 346 normal eyes, 173 and patients having primary glaucoma out of 562, 473 eyes were examined. 15º coloured stereo optic disc pictures were used. Normal eyes have larger inferior temporal vessels. The diameter of retinal vessel diameter is independent of patients' age and optic disc. Vessel caliber in glaucoma group is comparatively small as per normal eyes; diameters of vessels decrease with increasing stage of glaucoma. Vascular supply in the superficial retinal area is being reflected in parapapillary retinal vessel diameter. This is correlated with nerve fibre layer thickness and local gangolian cell density.[42]

At branch point of arteries, diameters are evaluated so as to optimize circulatory efficiency and design principles and maintaining shear stress in the network. N.

Chapman et al. had an aim to check if the optimality at these bifurcations get affected with atherosclerosis in individuals or not. Retinal images of men with normal blood pressure and abnormal ankle brachial index(n=3), healthy controls(n=8) were analyzed. These were matched and compared for age and clinical blood pressure with controls. Men having peripheral disease had bad metabolic reports. So it was found out that with comparison to healthy men, men with peripheral vascular diseases, at bifurcation the junction exponents were deviating from the optimum value. Though bifurcation angles of retinal arteriolar did not have any difference for both men. Abnormalities in the arteriolar network of retina is called atherosclerosis[43]

In 1982, Tachibana observed that change in prfusion pressure with response of change in retinal arterial diameter made him to measure it with the help of fundus camera in 65 normal subjects. Change from comfort posture to an erect position, there is reduction in MABP and very small but an important increase was found out in arterial diameter with reduction in the retinal pressure in all subjects. This reduction(% change in diameter/△effective MABP) was correlated with advancement of age($p<0.01$). no important correations were observed between magnitude of diameter(range-60-140μ), systemic blood pressure(75-110 mmHg) and retinal arterial reactivity. Hence with these results; it was proved that retinal artery has an auto regulatory mechanism which gets influenced by aging[44].

In 2003, Huiqi et al.[44] proposed a piecewise Gaussian model taking the central reflex feature and hence described the vessel profile. It was evaluated by curve fitting over 505 segments of vessels. Classification result for arteries was 82% and for veins was 89%.

Likewise in retinal images which were divided into four quadrants with optic disk being centred, GRisan et al.[45] extracted features. Arteries and veins were almost same in every quadrant. Features were mean of hue values and variance of red values; with classifier as fuzzy clustering, 443 vessels were classified with 87.6% result.

In 2005, H.F.Jelinek et al. proposed a method in which optic disk was eliminated and then features were extracted from vessels. With the help of decision table, the best result was reported 70%[46].

In 2004, DRIVE database was published. It has 20 images each in training and testing data set and for every image, vessel map is sorted manually[47]. Meindert Wiemiejer et

al.used information of intensity and derivatives to yield results in 2009; 12 features wre extracted in HSL colour space and green channel of RGB space. A KNN classifier was used for this method. Manually, for separation blue colour was assigned to veins and red for arteries for the comparison by ophthalmologist. The result was 88% under ROC curve by automatic method for centreline pixels.

Segmentation of Retinal Vessels is divided into two parts : pixel processing and tracking method. In two-stage region procedure, retinal vessels were segmented.2-D wavelet transform was used for fundus images in with KNN classifier for determining probability of every pixel in the vessel. Training based method for vessel profile model was used with NN for every pixel in[48].

Wide irregularities are the symptom for DR which results in low AVR. Hence the classification of arteries and veins in retinal blood vessels is necessary. Not only arteries and veins but also curvature, colour, reflectivity and diameter serve as the diagnostic indicators.

1.6 OBJECTIVE OF THE PROJECT

The aim of this project is based on the Pattern Recognition or Classification. Classification has to be done to detect whether the present blood vessel is an artery or vein. For, that the DRIVE dataset from the Image Science Institute is downloaded. Work has been done earlier on this dataset with various techniques and here, the Neural Network Classifier is used for supervised learning. The two types of Feed- forward network used are Back propagation and Probabilistic Neural Network. The classification is required for measuring AVR. It helps in detecting systemic disease at earlier stage so that its diagnosis can be done in time.

1.7 ORIENTATION OF THESIS

In chapter 2nd , biological background of retina and its functions are discussed.

In chapter 3rd, concepts of neural network and pattern recognition are provided.

Followed by 4th chapter, in which the methodology of the work done is proposed and,

In chapter 5th, results are being displayed and discussed .

Then, the conclusion and future scope of work is described in 6th chapter.

CHAPTER 2

BIOLOGICAL BACKGROUND OF RETINA

Eye is sensitive to a portion of electromagnetic energy travelling through space that is called light. The human eye is sensitive to only a small portion of this spectrum known as visible spectrum.

Within this spectrum, a long wavelength appears as red and as the wavelength decreases, the colour changes. The middle wavelength appears as green colour and shorter wavelength as blue. The colour created by given wavelength is called as the hue.

2.1 HUMAN EYE

The human eye is that sense of organ which reacts to light and has many purposes such as the following:

a) Helps us to see stimuli

b) Minute details can be recognized

c) Enable dark and light adaptation

d) Facilitate saccadic movements

e) Enable us to see colours

The eye is like a camera. It has a convex lens on which light is incident and hence passes through it. The light passes through a transparent protective structure, curved smaller frontal unit called cornea. It is approximately 8mm in radius.

Due to transparency of cornea, we can see the centre of the eye, black circular portion called pupil whose size varies with lighting condition. Amount of light passing through lens is regulated by iris; a structure whose shape adjusts to permit any object at varying distances. The projection of the image is finally formed at retina.

Eye consist of three coats; having three transparent structures.

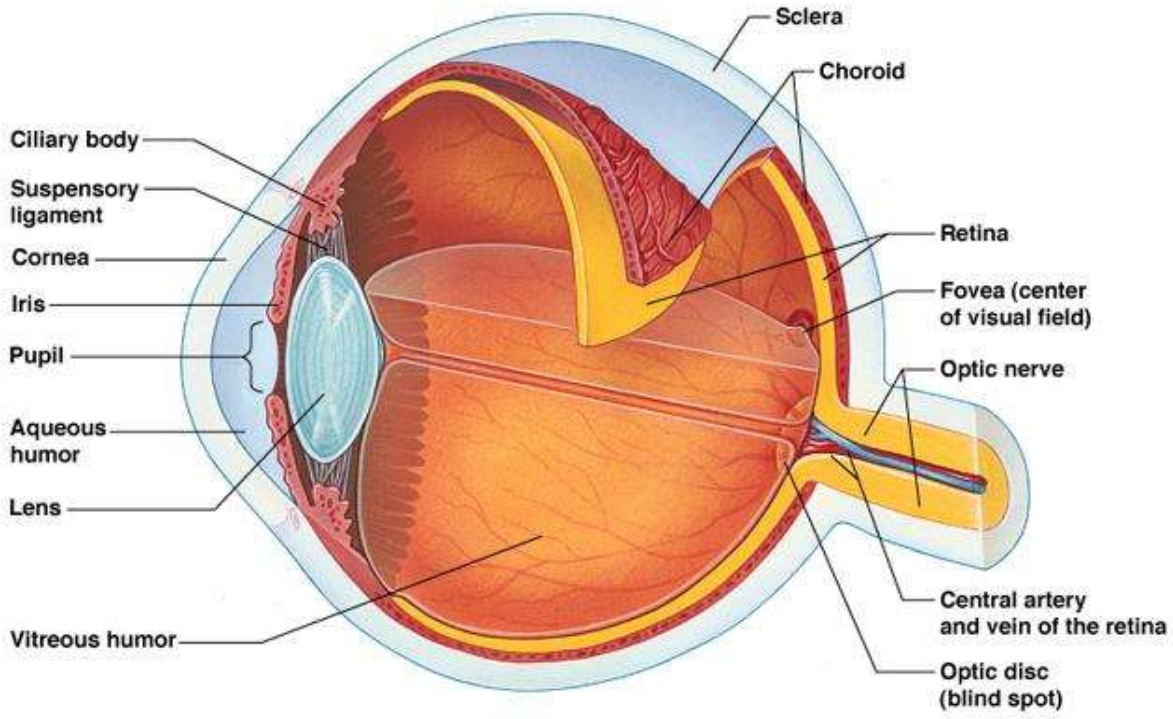


Figure2.1: Human eye

1. Fibrous tunic, outermost layer, composed of sclera and cornea. Cornea admits and refracts the light. Cornea has approximately 1.38 refractive index. Sclera protects the inner parts and provides shape to them.

2. Vascular tunic or, middle layer consists of iris, choroid and ciliary body. Choroid helps in blood supply and absorbs scattered light. Iris is the coloured portion and regulates the amount of light entering the eye light entering through pupil passes through crystalline lens. It has an approximately refractive index as 1.40. Ciliary body alters the shape of lens according to vision and secret aqueous humour.

3. Retina or nervous tunic is the innermost layer which gets its circulation from retinal vessels or choroid vessels and can be seen through an ophthalmoscope. It is the beginning of visual pathway.

Within these coats are vitreous body, aqueous humour and flexible lens.

Aqueous humour: - fluid that is contained between cornea and iris; and iris and lens.

The lens is made up of transparent fibres and suspended by suspensory ligament to ciliary body.

The vitreous body present behind the lens, is a clear jelly and connected via pupil.

To study the effect of systemic diseases and retinopathies on the retina and its vasculature, we need to understand the architecture of retina and its ocular structures.

At the back of an eye, there is a sensory layer called retina, it has three main layers :-

a) first layer has nerve cells that are sensitive to light and are further classified as rods and cons. These cells convert light energy into nerve impulses. Retina contains approx 100 million rods and 6.5 millions cons.

A1) cons- they lie in the centre area of retina known as fovea; work best in bright light, coloured, and black and white vision too. They pick up fine details and produce colour sensations.

A2) rods- they lie outside fovea and work best in lower levels of illumination. They help us in a better vision even in night or dim light and black and white light.

b) bipolar cells- they make synaptic connection between rods and cons.

c) ganglion cells- optic nerve is formed of these fibres. The nerve cells act as mediators between the central nervous system's part and optical signal received at retina and therefore dealing with visual sense.

2.1.1 Structure of Retina

Retina has three components:-

1) Optic disc (or optic nerve head)- it is near the centre, oval shaped and is main part of eye. It is also called blind spot and is the part where optic nerve exits the eyeball. Thousands of electrical signals which are interpreted are to be transferred from retina to brain through disc. Its abnormalities can be a mark for few diseases such as astigmatism and glaucoma.

A multilayered outgrowth of the brain is referred as neural portion of retina. All vessels appear to converge here. Photoreceptors are specialised to transduce light waves into

receptor potential. This visual data is processed before transmitting nerve impulses to the brain.

2) Macula- fovea located in the middle of retina is responsible for high resolution vision. Concentration of cone cells at fovea is responsible for sharp vision . Any abnormality is this region would lead to blindness.

3) Vasculature – supplies blood to retina; it also has same physiological and anatomical characteristics with coronary circulation and cerebral. Hence, changes in any part of the body will affect it. It can be viewed directly by ophthalmoscopy or digital retinal imaging.

The present scenario in research of automated retinal analysis targets in detecting above landmarks to aid diagnosis. The detection of these landmarks is important to establish a retinal reference coordinate system so that first system can analyze pathology entity and also enable mapping of lesion distribution.

The diseases can be of two types:-

I) Retinal abnormalities- ischemic diseases, retinal stresses, inflammation, radiation damages and hemodynamic diseases.

II) Abnormal blood circulation due to systemic dysfunctions- diabetes, cardiovascular diseases, hypertension, blood infection and cancers.

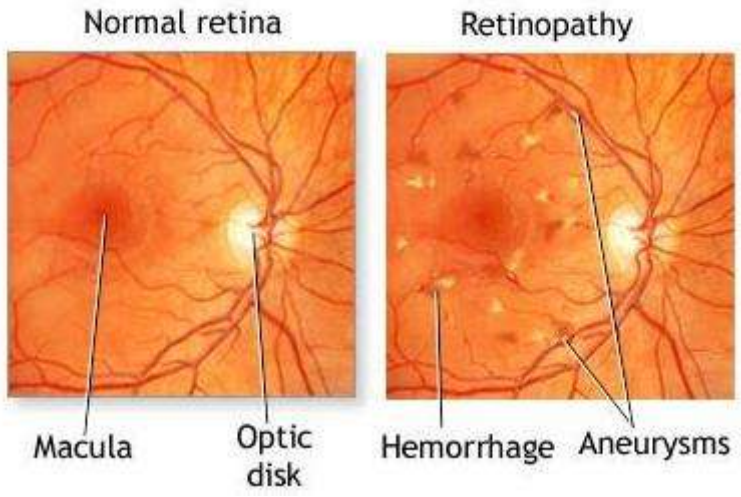


Fig 2.2: normal eye and eye with retinopathy

Retinal blood vessels are measured as entire network of blood vessels can be directly seen in the retina thus we examine it for pathological changes. If we use the ophthalmoscope through pupil, we can see a magnified image of retina and blood vessels. Changes in the structure of blood vessels indicate many types of diseases such as hypertension, diabetes and arteriosclerosis.

Blood vessels (arteries and veins) have many features on whose basis we can conclude like length, colour, diameter, tortuosity, and branching angle.

2.1.2 Ocular Circulation

For the nutrition of an eye, there are two separate vascular systems: retinal vessels and the uveal, or ciliary, blood vessels.

The uveal blood vessels have choroid, ciliary body and iris. Important function of choroid is to serve retina. Retina depend on retinal vessels and choroid.

Ophthalmic artery, branch of internal carotid, ocular vessels is derived from it. This artery branches into two or three posterior ciliary arteries, central retinal artery and few anterior ciliary arteries.

10 mm behind the eyeball is optic nerve where central retinal artery enters. It appears at about 1.5 mm in diameter at the optic disc. This branches into four major vessels; each supply into one quadrant of retina.

In the inner two- third of retina, retinal vessels are distributed. In outer layers, (including photoreceptors) are avascular which are nourished from the choroid.

In fovea, is avascular zone; it enables light to central photoreceptors and not even a single blood vessel is encountered. In nerve fibre layer are arteries and veins located.

In a large part of retina, there are two layers of flat capillary networks. These have a diameter of 5-6 µm.

Retinal arterial diameter- 40-160µm

Central artery- 160µm

Inferior temporal branches and Superior temporal - 120µm

2.2 PROCESSING OF VISUAL INFORMATION

In each sensory system, the receptor neuron deals with various kind of energy- mechanical, chemical or electromagnetic. These all exhibit different receptor proteins and look different from each other. They convert a stimulus into an electrochemical nerve impulse from environment so that brain can understand.

The first step begins with stimulation of photoreceptors (rods and cons). Then the incoming light is focused on the retina and the information passes to the next layer of retina. Finally it goes down to optic nerve and then deeper to the brain.

Most of the refraction occurs at cornea. This is because it has a shape of converging lens. Lens changes the shape as per the action of ciliary muscles.

Cornea has a bulging shape which refracts light as it does to a double convex lens. Real, small- sized and inverted image is formed and brain has the ability to interpret the signal easily.

The concentration of cone is greatest at a diameter of 0.25mm approximately; though nerve cells being all over through retina. The region mentioned here is known as fovea centralis which is considered as the optimal location for the image formation.

Since the distance between the detector (retina) and refractor (cornea) is fixed; eye must be able to change the focal length so as to focus both far away and nearby objects from the retinal surface.

Ciliary muscles squeeze and contract the lens into convex lens for nearby objects an d they relax to flatter shape for distant objects.

Features are detected by the neurons at various levels. Three types of feature detectors are:-

1. Simple cells- cells within visual system which respond to specific shapes (vertical, horizontal).

2. Complex cells- neurons in visual cortex which respond to stimuli having a particular direction and orientation.

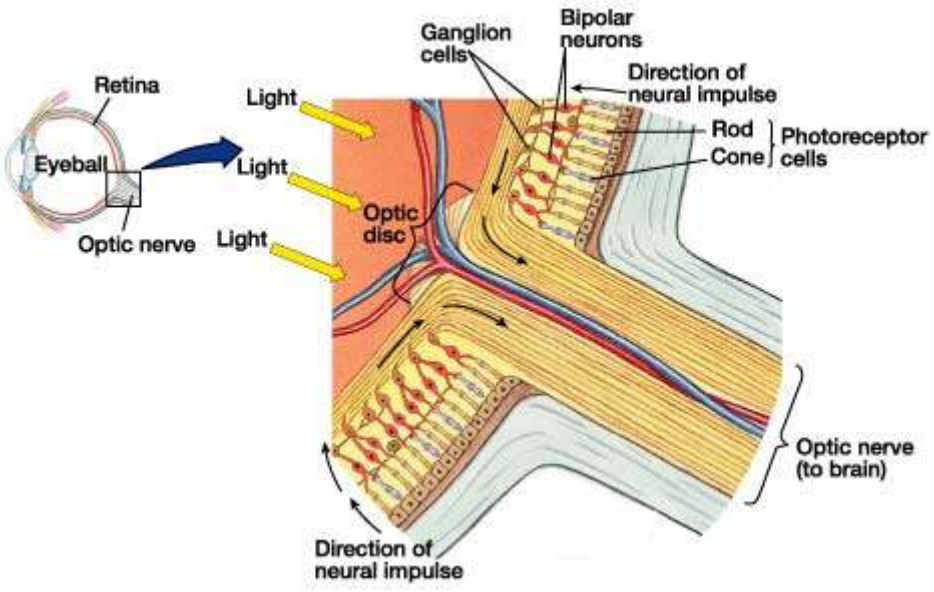


Figure2.3: Path of passing information from eye to brain

3. Hyper complex cells- neurons in visual cortex which respond to complex aspects like shape, width and length for visualising stimuli.

2.3 RETINAL PATHOLOGIES

Diabetic eye disease refers to group of problems in which people with diabetes face the complications of diabetes.

Diabetic eye disease may include:-

1. Diabetic retinopathy- retinopathy is referred as damage to small blood vessels in the retina.

2. Cataract- it develops at earlier age in people with diseases. It is clouding of eye's lens.

3. Glaucoma- a person with diabetes is likely to suffer through it. In eye, there is an increase in fluid pressure of the eye. The pressure builds up in anterior chamber and slows down the drainage of aqueous humour. This leads to loss of vision and optic nerve damage.

2.3.1 Diabetic Retinopathy

Eye is one of the first place where diabetic retinopathy becomes apparent. So, examination of various characteristics features can help us in diagnosis of various diseases.

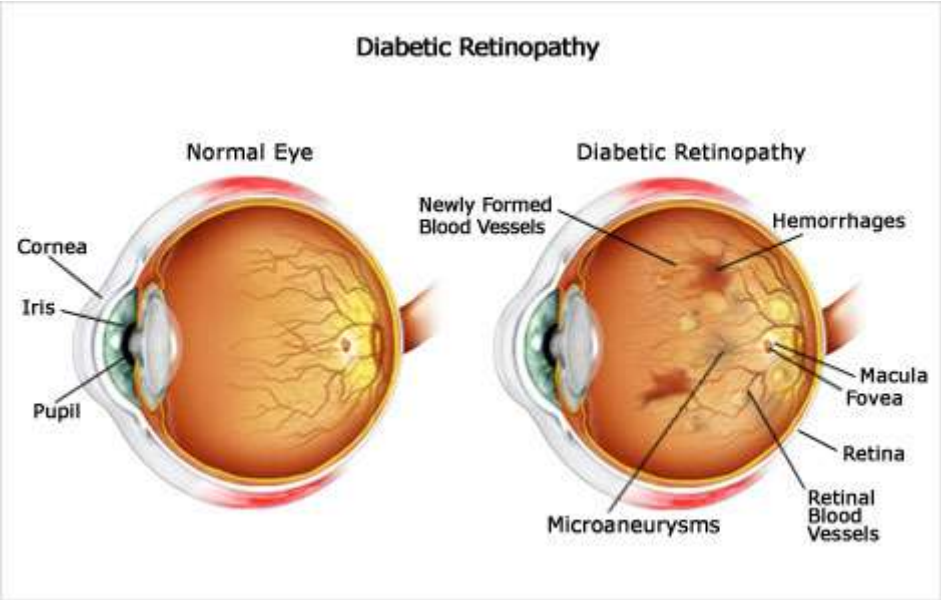


Figure 2.4: Difference between normal eye and eye with diabetic retinopathy

Diabetes is also known as diabetes mellitus. It is a group in which high blood sugar level in a group are prolonged over a long period. It is due to pancreas which can't produce enough insulin.

Diabetes is the inability of metabolic system so as to control the blood sugar levels which results into hyper glycaemia. As the micro vascular structures in the retina are easily attached by blood sugar level hence they are severely damaged. Pericyte cells supporting the vessel walls die and thus the wall becomes permeable and fragile.

It is damage to retina which happens due to occurrence of diabetes and can lead gradually to blindness or severe vision loss. If the patient has diabetes for ten years or more, 80% of them will be affected with this systemic disease. 90% of the above mentioned cases can be reduced with proper treatment and monitoring of the eyes.

Two stages of diabetic retinopathy:-

1. Primary stage- NPDR

2. Secondary stage- PDR

(NPDR) - Retinal blood vessels when not able to withstand high concentration of sugars like fructose and glucose in the blood, gets damaged.

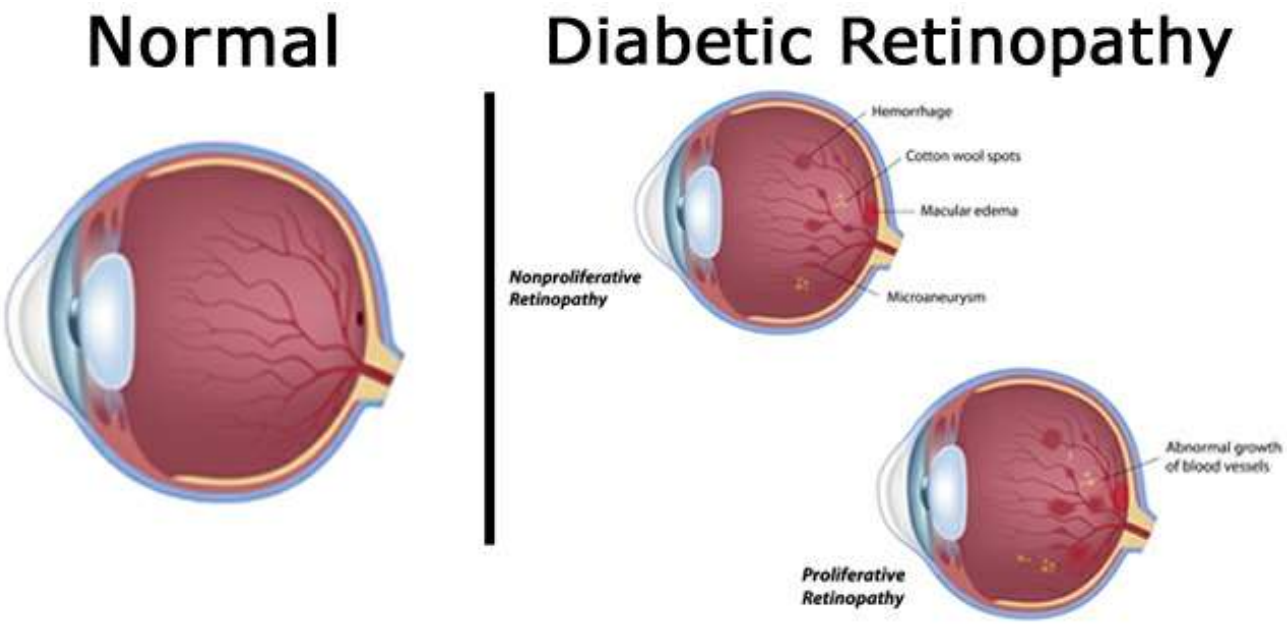


Fig 2.5: Types of diabetic retinopathy

Mural cells on the interior lining of vessels are damaged, these help in formation of blood retinal barrier; blood leakage is also prevented. Since deterioration of endothelial lining and mural cells occur, walls of blood vessels become weak and hence leak or break down. This leakage may lead to few pathological conditions like protein fluid(deposits known as hard exudates), edema(accumulation of fluid) and cotton wool spots(capillary blockage).

In few cases, it might burst leading to haemorrhages. As the disease and damage speed up large haemorrhages will appear. Not only leaking blood but vessels leak proteins, and lipids which cause small bright dots (exudates) to appear.

Small parts of retina have less supply of blood so these areas look like fluffy whitish blobs also called cotton wool spots on the retina.

One of the clear signs of presence of diabetic retinopathy is the micro aneurysms' (tiny capillary wall swellings) appearance as small red dots between large vessels of retina. This may result in scar tissues formation; causing retinal traction or detachment by contraction.

Advancement in (NPDR) conditions lead to a (PDR) stage. This results into retinal ischemia therefore poor nourishment of retina.

Due to responses exhibited by human body, nourishment signals are being sent to unnourished areas for oxygen supply. In the areas with less blood supply new vessels grow with more oxygen. Neovascularisations, these new vessels have a greater risk of causing large haemorrhages and rupturing than normal vessels. These blood vessels formed are weak and fragile; not normal. Its formation might be on ocular structure, iris in vitreous humour or on the retinal surface. These blood vessels might leak, bleed or breakdown and hence result into fibro vascular proliferation. Because of this scar tissues are formed leading to retinal detachment. If the vessels are formed on iris; filter responsible for fluid drainage from eye is blocked. This may cause more pressure in eyes which ultimately results to neovascular glaucoma and hence blindness.

CHAPTER 3

NEURAL NETWORK CLASSIFIER AND PATTERN RECOGNITION

The motive of any classifier is to categorize the sequence of data from pre-processor into different classes. For any changes the classifier does, it also includes software which is self adaptive and is trained for correct classification.

In Artificial Intelligence (AI), many criteria's of adaptive software are developed. ANN is very successful classifier and its basis is dependent on the mimicry of the structure of the human brain. The following chapter has a detailed description of the history behind evolution of ANN's.

3.1 CONCEPTS OF NEURAL NETWORK

3.1.1 Data Mining

It is defined as the step for analysis in KDD. It is the processing of computation to discover patterns in data sets. It is processed with the combination of AI, machine learning, and database systems. The aim is to transform the data se in an understandable form by extracting information from data set for further use. It involves various algorithms to analyze data. These algorithms try and fit the model to data. After the algorithm examines the data, it determines a model which is closest to the features of the data which is examined. The KDD process includes steps like transformation, pre-processing, selection, evaluation and data mining. And followed by data mining functions are regression, prediction, summarization, classification, time series analysis, association rules, clustering and sequence discovery. The problems of classification and prediction may be solved with the help of neural networks.

3.1.2 Classification

The problem of dividing number of patterns or data into different classes

C= {C1, C2,..,Cm}, where [m- no. of classes] is known as classification.

It is defined as the problem in which one can identify that to which new observations lie in which categories. It is done on the basis of :-

Training set- contains the data or instances for which we know the category. Set of quantifiable properties can be analyzed by individual observations. Those may be in terms of features, explanatory variables.

Testing set- it has the data to verify in the end after the classifier being trained.

An algorithm which can implement classification is said to b a 'classifier'. It is referred to a mathematical function which is implemented by classification algorithm and hence mapping the input data to a particular category.

In the thesis given, the problem is that how to build a network and make it learn to classify any input into m classes, given there are P patterns.

Classification is also dependent o the transfer function being used in hidden layer. If its sigmoid or threshold function, the network becomes a global classifier.

When we talk about machine learning, classification is taken as a part of supervised learning (where the training set of correct identified results is available).

When we work with areas or statistics where a huge amount of data is to be analyzed; it becomes important to categorize the data- points into different sub- groups. It seems to be a difficult task for human beings; who are not often able to identify that to which data point the class belongs to because the data contained is in large amount. Hence, we need a digital classifier.

There are various ways of designing a digital classifier. A common character for all classifiers is that their work follows supervised learning (classifier is trained on some data whose output is known and again same kind of data is used to test the knowledge of training data for classifying the new data).

It is necessary for any classifier that it can sort and generalize data which it might not have faced before based on the sub-group which is most alike.

3.2 BIOLOGICAL NEURONS

Biological neurons are the building blocks of NNs and are to be imitated; so its important to understand their functions. In ANN, maximum terminology is originated from these only. ANN creates machines which work in the same manner as human brain

components do. In fact, the operation of ANN is simplified comparative to the human brain.

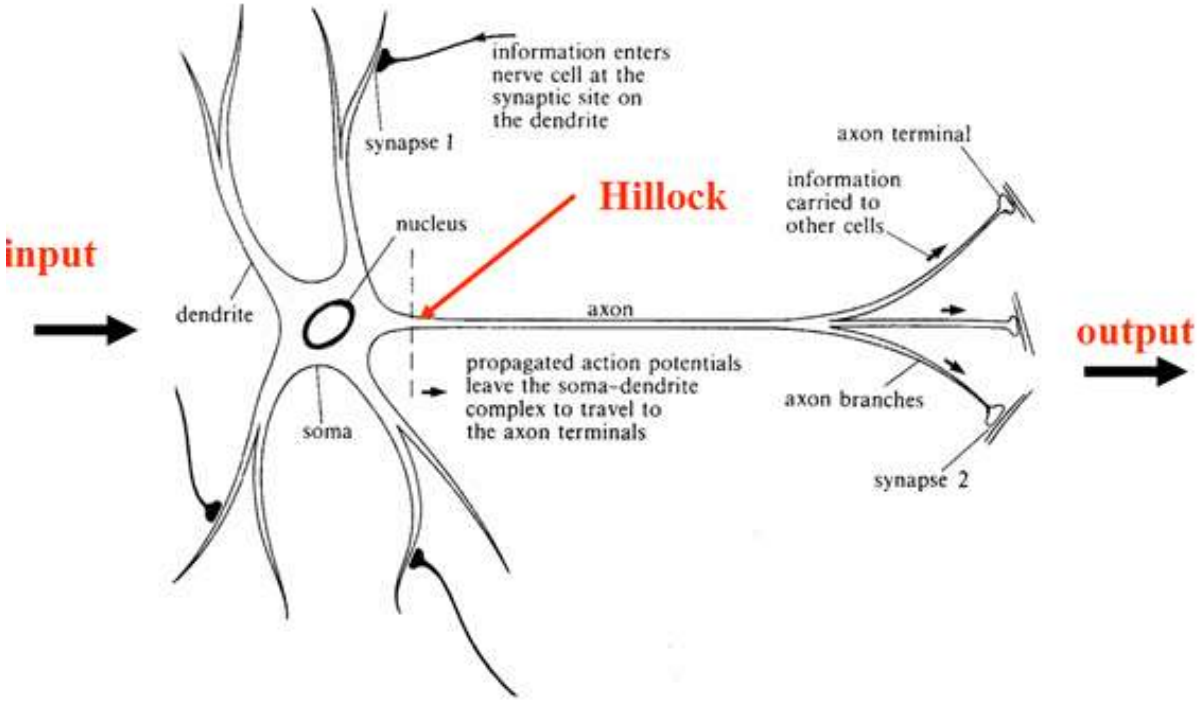


Figure 3.1: biological neuron

The actual processing units; neurons (of the brain) are compared to silicon- chips and we see they are slow. Their rate of computation is also simple. A powerful network is only formed if lot of these units are together. In any silicon-based computer, there is only one processing unit but in NN, these neurons work in parallel. Each neuron may be connected to over 1000s of other neurons. The structure of each of them is similar. Cell body (soma) of a neuron has branched fibres, called as dendrites. One and more axons may extend away from all to other neurons and hence branches in the end.

A negative electrical charge (-70mv) is the resting potential for neurons. The potential can be rose with the help of certain stimuli of other neurons. Threshold value (-55mv) is reached when neuron fires a spike (electric pulse along its axon). The axon branches at the end are connected to other neurons via dendrites; this is called synapse. A neuron sending the spike is called pre-synaptic neuron and one receiving is called post-synaptic neuron. Change can be on negative or positive depending on what type of synapse it is(positive- raising potential, negative- lowering potential).

When the potential is raised, neuron fires and hence synapse is called excitory. On lowering potential, its hard for neuron to fire and hence synapse is called inhibitory. Since neurons want to stay at the resting potential, this change is temporal and gradually fades away. Refractory period is referred as the time interval after neuron fires the spike and recovers back.

There is a variety of synapses and neurons. Few types of neurons have large axons which can influence other regions of brain. Few other neurons with short axon and dendrites can only compute locally. Not necessarily axons have to form synapse with dendrites. Sometimes synapses are formed with cell body of other neurons creating a bigger influence on it. When some of them form inhibitory synapses (with other axons) so as to present the spike being propagated from it.

Brain is heterogeneous mixture of neurons and similarly a biological NN is also built with a mixture of these neurons. But since we get lot of loops in the network; these are called recurrent networks with both positive and negative feedback.

3.3 NEURAL NETWORKS

For the simplification of classification or prediction problems, neural networks are being introduced. These are the models (simpler) of biological neuron system of human body. They are also called parallel distributed system. It is made up of interconnected elements of neural computation which have the capability of learning and hence acquiring knowledge and thus making its presence for use.

There are mechanisms which can enable NN to acquire knowledge. We have lots of NN architectures depending on the learning mechanism they use and many other features. The process of learning is said to be as training and to solve any problem with acquired knowledge is termed as inference.

NNs are the simplified versions of Central Nervous System and also have been inspired by computation of human brain. The constituents of human brain are entities called neurons; they tend to perform computations like logical inference, cognition and pattern recognition and so on. Brain is composed of millions of neurons which might or might not perform all the tasks.

ANNs (Artificial neural network) is the technology which imitates the simplified version of neurons of the human brain. The technology is also referred as Neuro-

computers, Connectionist networks, or Parallel Distributed Processors. Neurons are also called neurods, nodes and Processing elements (PEs). The work over ANN is being inspired. With the way human brain (part of biological nervous system) processes the information. So, the goal is to design a system which can handle problems as the biological brain with ease.

So we can say that NN models are relatively electronic modes as the structure of the brain. With the help of biological nervous system, we see that NN is a network of interconnected elements. There is always a level of abstraction. It gives different variety of neurons and thus we have different models. Neurons might or might not be abstracted accurately. We have to overall see the performance of neuron in the end.

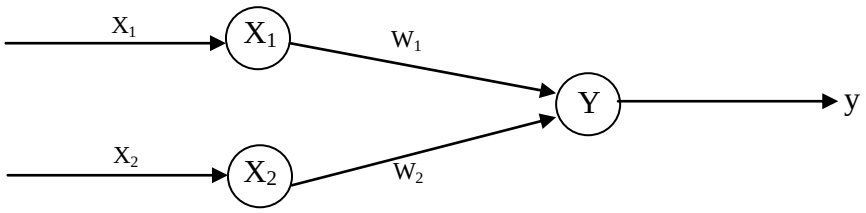


Figure 3.2: Architecture of a simple artificial neuron net.

NNs are used in the connection with various applications. There may be tasks like classification/ pattern recognition or function approximation. As in a network can be asked to classify an input pattern if it belongs to number of available classes, or maybe to produce an output value for an unknown function of (one or more) input values.

A difficult feature for our network is its ability of how it would help mapping the inputs to desired outputs without telling it the rules. So, they instead adjust and manipulate internally the connections depending on the desired mapping and hence give results previously not seen. So, we can say that they capture patterns in sets of any data and these patterns are required to perform computations.

There are different types of NN, each with advantages and disadvantages depending on the problem given. The network I chose for classification is a feed forward neural network. It is characterized as the information flows from input and forwarded to output and no feedback is there.

Table 3.1: Representation of parts in both biological and artificial neuron

Biological neuron	**Artificial neuron**
Cell	Neuron
Dendrites	Weights or interconnections
Soma	Net input
Axon	Output

All neural networks are made up of lots of highly interconnected nodes (units) or neurons. Each of these units receives input from the outside world or another network and hence calculates the output based on them. With each connection, a synapse is associated with weights. The manner in which these units are connected, they communicate , exert and have vital influence on its properties.

In the feed-forward network, units' arrangement is done in layers. It has different processing levels first being input layer and last being output layer . An intermediate level of hidden units is hidden layer which is far away from the outside world. First hidden layer is near the input layer and there can be as many layers(hidden)depending on the problem and last layer is nearest to output layer. If the network do not have any hidden layer, it is simply called a simple perceptron and if more than one hidden layer thus its called multi layer perceptron(or X-layer perceptron). In a feed-forward network, the layers are connected from low to high and thus no feedback is allowed. This gives the stability to those systems.

3.3.1 Properties of Neural Network

i) Learning

It is the ability of network that it can learn from the examples assigned inputs are thus learnt to give desired outputs. It involves the manipulation of parameters present inside; maybe weights so that overall response behaviour is given by set of training examples. Any training set includes an input and output pattern so a network is trained by giving input to network and then see what corresponding output we get. If the ultimate output is not what is desired, the internal weights are modified so that difference between desired and actual output decreases. This is continued till the point where value of

weight is such that giving an expected output. They have the property of displaying the mapped patterns of input to associated output patterns.[]

ii) NN learn with the help of examples. Hence the NN classifier is being trained. Training is done before analysing the problem for results. After training, the classifier has the ability of testing any new data or unknown data.

iii) NNs have the ability of processing information at high speed, in parallel and distributed manner.

iv) Non-linearity

As read earlier the output space is mapped from input space of a network, this mapping is associated as

Y= f(x) = vector valued function

It transforms an input vector to output vector y. These both can have any dimensions.

f- mapping itself is a combination of the mapping performed by simpler units in parallel. Resulting neuron is non-linear, as information processed in every neuron is n on-linear. Hence, for physical system it's a very good property.

v) Fault tolerance

Like the human brain is a wonderful example of fault tolerant system. This system means that it will keep on working even if other or any part stops working. The brain works all the time though neurons, die off as the natural process of human body. In case there is any serious damage, only then there is change in its performance. Though it does come to a lower level bu does not drop to zero hence it is said to be graceful degradation. Also brain is a distributed and parallel structure; so ANN tries to copy our brain structure and thus incur lot of properties of brain. They have the ability to recall complete patterns from partial, noisy or incomplete patterns. Hence, we can say it is a fault tolerant and robust system.

vi)Generalization

NNs also have the ability to predict new outputs from the past methods; so they possess the property of being generalised. The good part is a network can learn without prior

knowledge from various examples. Hence we say it produces good results for inputs which are not encountered during training.

3.3.2 Neural Network Characteristics

In NN, network is the interconnection of neurons in different layers of the system. Each system has following layers:

I/p layer has the i/p neurons which transfers data through synapses to next i.e. hidden layer. Then via hidden layer the data is forwarded to output layer with the help of more synapses. Stored synapses value called weights ; help to manipulate both input and output to various layers. Three characteristics of ANN are:-

1. the architecture:- no. of layers and nodes are present in each of these layers.

2. a learning mechanism is applied so that weights of connections can be updated.

3. an activation function is used in different layers.

3.4 ARCHITECTURE OF FEED FORWARD NEURAL NETWORK

When for the graph is assigned an orientation we call it a directed graph or a diagraph. A feed-forward network has the directed but acyclic graph. Since NN graphs are restricted to flow in a particular direction; hence diagraphs are important neurons (i/p-o/p) are represented on vertices and synaptic links represented on edges in any graph. Weights are attached to links labelled on edges.

a) Single Layer Feed-forward Network

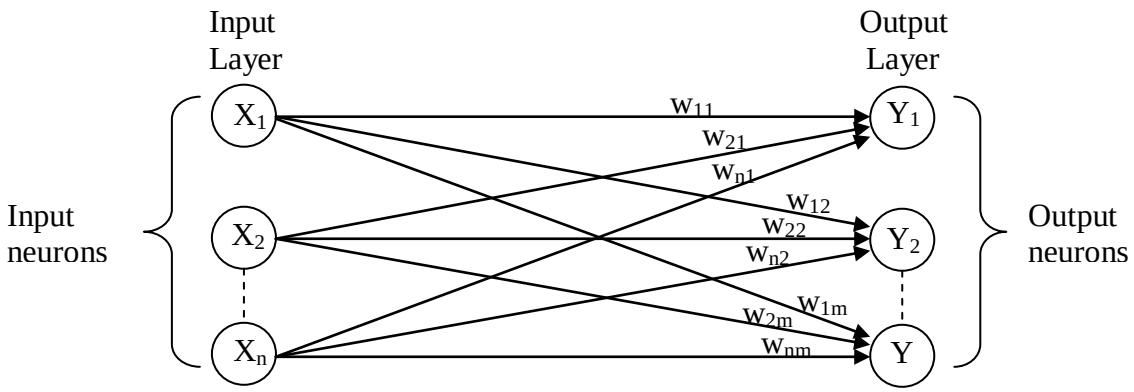


Figure 3.3: Single Layer Feed-forward Network

It consists of only two layers: i/p(receiving i/p signals) and o/p(neurons receiving o/p signals)

Weights are connected from i/p to o/p neurons via synaptic links but not vice versa. This kind of network is referred to feed-forward network. Output layer only performs computation hence it is called single layer despite having two layers.

b) Multi layer feed-forward network

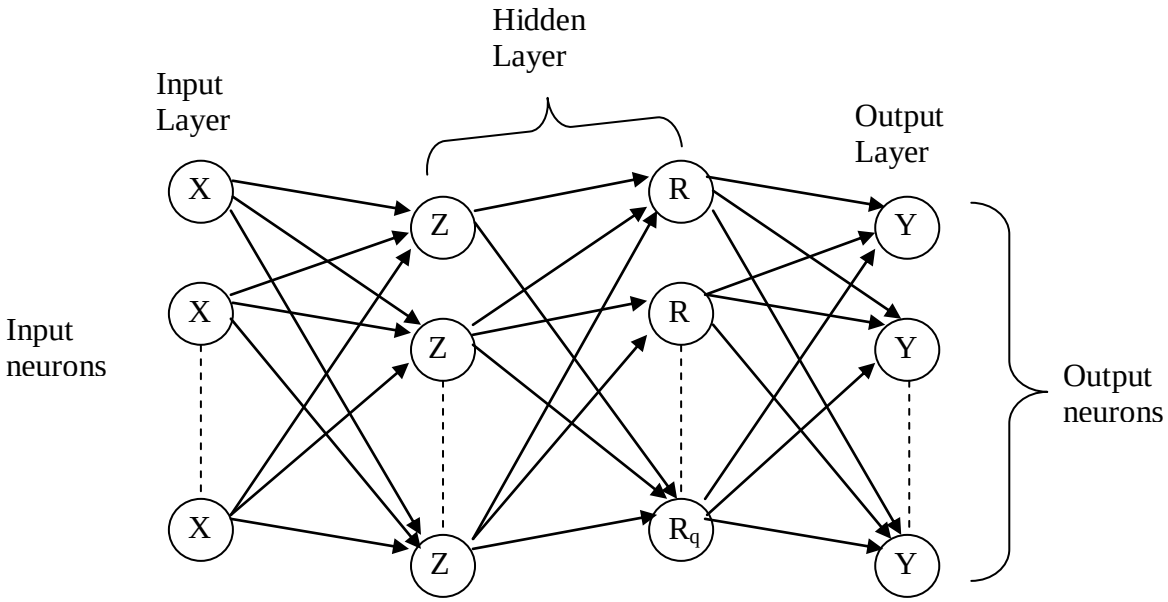


Figure 3.4: Architecture of Multi Layer Feed- forward network

It has three layers: i/p, o/p and hidden. Hidden units are present in hidden layer which conducts intermediate calculations or manipulations which are required before giving the input information for output layer.

Weights on link of input to hidden layer are termed as i/p- hidden layer weights. In the same way, hidden to output layer has hidden-o/p layer weights.

3.5 CLASSIFICATION OF FEED-FORWARD NETWORKS

3.5.1 Backpropagation network

It is the method of training a multi-layer NN. Supervised algorithms are based on error learning algorithms which are compared with reference and thus output is obtained. On the basis of error signal, the synaptic connection weights are modified so as to improve

the performance of the system. In this method, an assumption is made that a desired answer is known as “a priori”.

The concept of this algorithm is gradient descent method i.e. weight up gradation method. The error is sent back to the hidden unit.

In these three layers, input has ‘x’ nodes, hidden has ‘z’ nodes and output layer has ‘y’ nodes. Sigmoidal functions are used for activation functions for both hidden and output layers but for input layer, linear activation function is used.

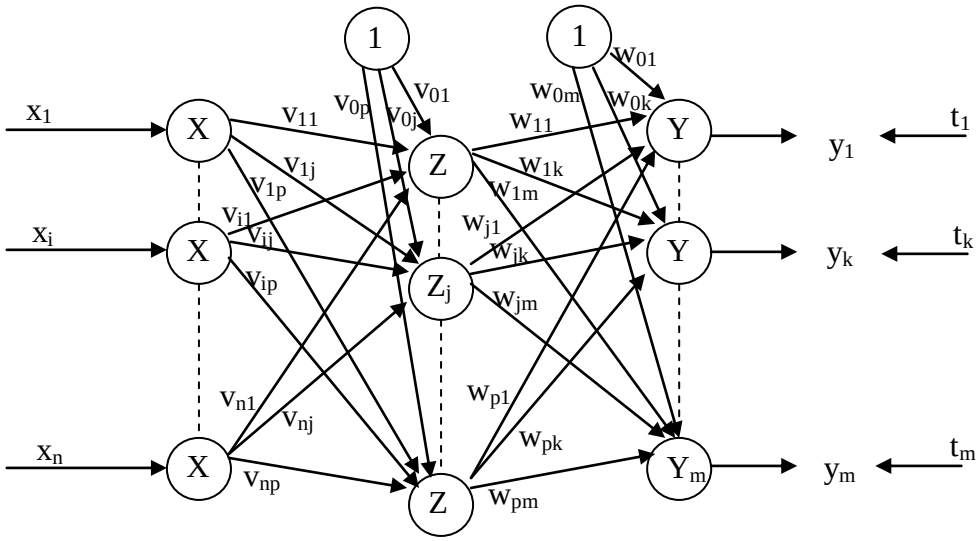


Figure 3.5: The architecture of BP

The training is done in three steps:

- i/p training pattern is feed forwarded
- calculation and sending back the error
- and weights are updated.

Training Algorithm

Step 0: Initialize weights and learning rate

Step 1: Perform Steps 2-9 when stopping condition is false.

Step 2: Perform Steps 3-8 for each training pair.

Feed Forward phase (Phase I):

Step 3: Each i/p unit receives input signal x_i and sends it to the hidden unit (i= 1 to n).

Step 4: Each hidden unit z_j(j= 1 to p) sums its weighted input signals to calculate net input :

$$z_{inj} = v_{0j} + \sum_{i=1}^{n} x_i v_{ij} \quad (3.1)$$

Calculate o/p of the hidden unit by applying its activation functions over z_{inj} (binary or bipolar sigmoidal activation function) :

$$z_j = f(z_{inj}) \quad (3.2)$$

and send the output signal from the hidden unit to the input of output layer units.

Step 5: For each output unit y_k (k = 1 to m), calculate the net input:

$$y_{ink} = w_{0k} + \sum_{j=1}^{p} z_j w_{jk} \quad (3.3)$$

and apply the activation function to compute output signal

$$y_k = f(y_{ink}) \quad (3.4)$$

Back Propagation of error (Phase II):

Step 6: Each output unit y_k(k= 1 to m) receives a target pattern corresponding to the input training pattern and computes the error correction term:

$$\delta_k = (t_k - y_k) f'(y_{ink}) \quad (3.5)$$

On the basis of the calculated error correction term, update the change in weights and bias:

$$\Delta w_{jk} = \alpha \delta_k z_j; \quad (3.6)$$

$$\Delta w_{0k} = \alpha \delta_k \quad (3.7)$$

Also, send δ_k to the hidden layer backwards.

Step 7: Each hidden unit (z_j, j= 1 to p) sums its delta inputs from the output units:

$$\delta_{inj} = \sum_{k=1}^{m} \delta_k w_{jk} \quad (3.8)$$

The term δ_{inj} gets multiplied with the derivative of $f(z_{inj})$ to calculate the error term:

$$\delta_j = \delta_{inj} f'(z_{inj}) \quad (3.9)$$

On the basis of calculated δ_j, update the change in weights and bias:

$$\Delta v_{ij} = \alpha \delta_j x_i; \quad (3.10)$$

$$\Delta v_{0j} = \alpha \delta_j \quad (3.11)$$

Weight and bias up gradation (Phase III):

Step 8: Each output unit(y_k, k=1 to m) updates the bias and weights:

$$w_{jk}(new) = w_{jk}(old) + \Delta w_{jk} \quad (3.12)$$

$$w_{0k}(new) = w_{0k}(old) + \Delta w_{0k} \quad (3.13)$$

Each hidden unit(z_j, j= 1 to p) updates its bias and weights:

$$v_{ij}(new) = v_{ij}(old) + \Delta v_{ij} \quad (3.14)$$

$$v_{0j}(new) = v_{0j}(old) + \Delta v_{0j} \quad (3.15)$$

Step 9: Check for the stopping condition.

.

3.5.2 Probabilistic neural network

It is a feed forward NN, derived from Bayesian network and a statistical algorithm called Kernel Fisher discriminant analysis. Gaussian activation function is generally used because of its easy computing with efficient results characteristics. Its operations are organized into a multi layered feed forward network with following four layers as shown in the figure

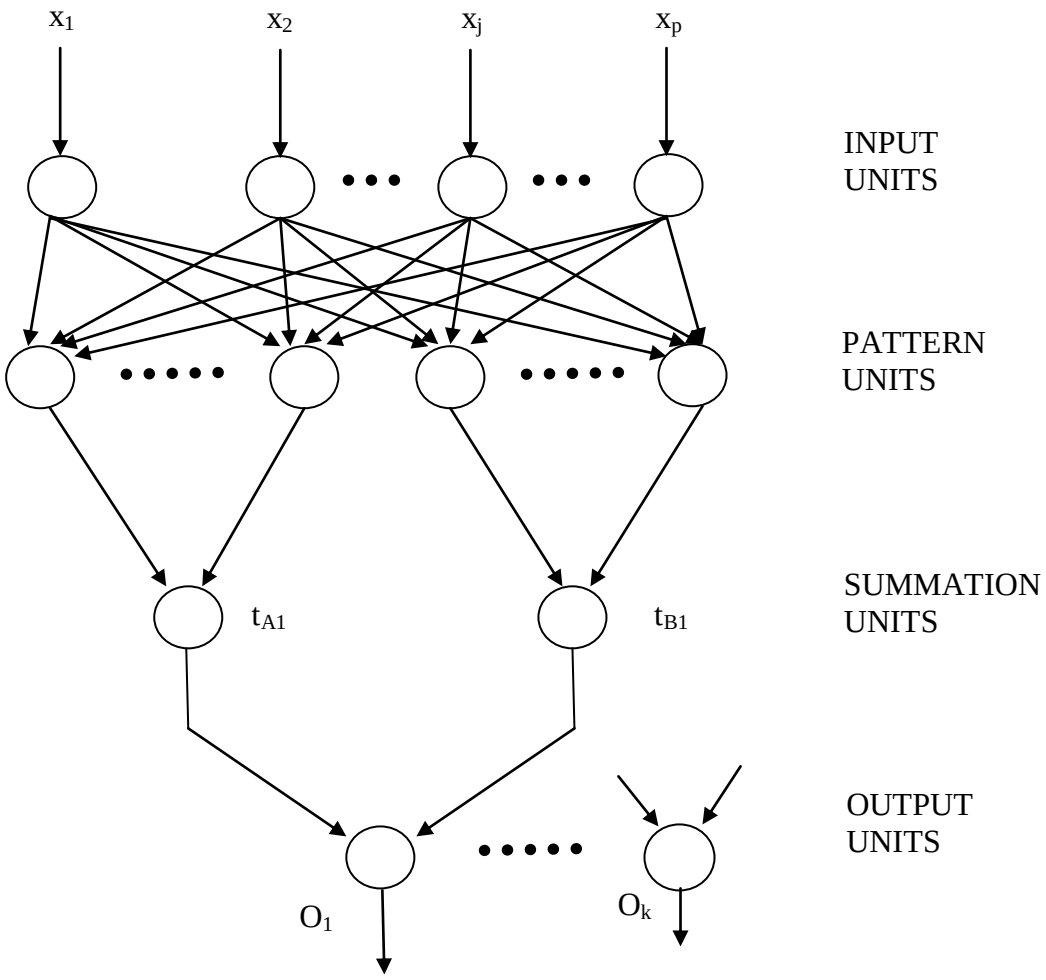


Figure 3.6: Architecture of PNN

1. Input layer- I/p nodes are present which are used for further measurements as sending the data to the next layer.

2. Pattern layer- It has one neuron associated with each of the training data set point. Euclidean distance is calculated from centre point of neuron to data set. This gives a vector output which tells us about how near or far is the training i/p from i/p. It saves the value of predictor variables of the data in the pattern neuron.

3. Summation layer- Now the above stored neuron values with respective weights are added up so that we can get the results corresponding to each neuron .

4. Output layer- It compares the evaluated results to weighted vectors of the target. One with the maximum probability of a class is assigned one class (+) and the other class is assigned as(-). Hence the data set is evaluated into two different classes.

a) THE PNN MODEL

Parzen's estimate and the Bayesian decision criterion –

This method is non- parametric identification procedure. By super positioning of number of windows, estimation of pdf is synthesized.

g( ) → copies of function, called as kernel.

$$f(x) \cong f_n(x) = \frac{1}{n\lambda} \sum_{i=1}^{n} g\left(\frac{x-x(l)}{\lambda}\right) \quad (3.16)$$

x → dummy arguments for points.

x(l)→ patterns from training set

$\lambda \rightarrow$ function of n such that

$$\lim_{n\to\infty} \lambda = 0 \text{ and } \lim_{n\to\infty} n\lambda = \infty \quad (3.17)$$

The estimator is same in quadratic mean:

$$\lim_{n\to\infty} E\,|f(x) - f_n(x)|^2 = 0 \quad (3.18)$$

density f(x)→ continuous

In the above equations whenever there is an increase in training set, estimates are more nearly smoothened to correct density.[] These results are got when we are using Kernel functions.

If the pdf of classes to be distinguished are known, Bayesian decision criteria is supposed to be optimal classifier. So, any of the class can be depicted as :

$$c = \text{argmax}_k\{\gamma^{(k)} f^{(k)}(x)\} \quad (3.19)$$

C→ class label o/p

x→ vector of input space

$f^{(k)}(x)$→ density of kth class

$\gamma^{(k)}$ →have prior probability and risk coefficients.

B) Neural Implementation

Structure of PNN is reflected by Baye's criterion. First layer as in figure is of input units , then second is of pattern units. With each pattern unit is 1 window whose size does not have any parameter. It might be called smoothening pattern.

$$Z_t = x(l).w \tag{3.20}$$

$$y_t = g(Z_t) \tag{3.21}$$

For l- pattern units;

zl = net i/p

yl = output

x(l) = l- th input vector

w- weight of input weights

g ( )- activation Kernel function

This all defines the pattern layer with single window. Next is summation layer which sums the windows of every class and hence required density is estimated. Then a comparison is done between the evaluated probabilities in the i/p pattern at the o/p layer and therefore the class with maximum probability assigns to one class.

C) The Training Algorithm

PNN is a fast technique and need only localized parameter adaptation whereas other training algorithms like BP need to change at every learning step. For every new

training pattern, i/p weights are updated so they match with the patterns . Hence the activation function is placed in the centre so as to use the desired Kernel function.

No adaptive parameters are present in summation layer. Relative frequencies of patterns in training set give weights in the o/p layer. These values are not calculated from the patterns. Hence no iterations are produced and thus this is said to be one- shot training.

3.5.3 Generalization performance

1. Overfitting

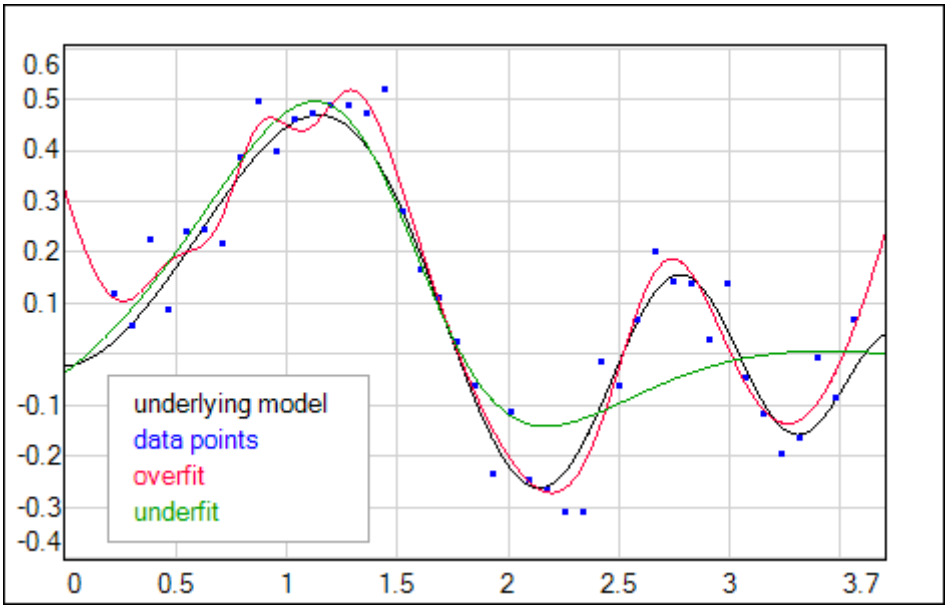


Figure 3.7: graph representing over fitting and under fitting

As described in figure , training a NN in many ways is equal to fit the curve for a set of data points. This curve is representing i/p-o/p mapping which is the result of network learning the training data points. The training error corresponds to total distance between curve and the points. Comparatively, a test error E_{tes} can be defined as distance between curve and set of the test data points not previously seen but from same population as taken for training data.

$$E_{tes} = \sum \| y_p - d_p \| \tag{3.22}$$

Test error is seen as an estimation of total generalization error in network as P is the total summation in test set. If generalization capability is good for NN, it will give good i/p- o/p mappings also for data points which might differ from training set. Therefore, a test error is found out.

‘black curve’- o/p of properly trained network.

‘black curve’ is not passing through all data points in training set but as a whole it gives good approximation for the data.

‘red curve’- fits the training points exactly

‘red curve’ gives the idea of poor location of data points not present in training set. The particular behaviour is said to be ‘over fitting’. Over fitting does pose a problem practically for all types of NN. The basic aim for any BP algorithm is that it minimizes training error at any cost. At some point of time, network might reach a state where the generalization is not able to improve anymore. From this point, if training moves on and past the point, network will begin to over fit and gradually will loose the ability to generalize.

2. Early Stopping

Whenever over fitting occurs, we have to stop training with that generalization error; this method is said to be ‘early stopping’. Though it sounds simple but its not the reality. This is because we do not have any idea if we have reached that point or not and only test error can tell us about that optimal point. As in following figure, a smooth and continuous functions are presented. If that would be the case, we can try and stop the training process before the test set begins to increase.

Practically, lot of local minima are present in the error curves and thus its more difficult to tell if we have reached the best point or not. Choosing E_{tes} has its main idea of approximately the generalization error for the unseen data in test set. Thus for this issue present is the validation set. This test provides the help to us by interrupting training whenever the error reaches its minimum and hence using classification error of test data so as to measure performance of network.

3. Dimensionality Control

In a MLP, each weight is represented by a degree of freedom thus improving generalization through it is a problem of choosing an optimal number of mode required in hidden layer of network.

So another way of reducing over fitting is to reduce degree of freedom for network.

For the above figure

$$y = P(x) + \varepsilon \qquad (3.23)$$

$$\downarrow \qquad \downarrow$$

polynomial some random noise

Curve ‘black’ represents the best approximation of data set. It is LMS fit of a 3rd order polynomial.

Curve ‘red’ shows symptoms for over fitting.

Not only fitting P(x), but using extra degree of freedom to fit noise in training data gives poor realisation as a result.

Curve ‘green’ do not have enough amount of free parameters so as to fit the data and hence suffers from the problem of under fitting.

A small network might not be complex as to partition the complicate pattern space and similarly large network may over fit gradually leading to poor generalization. These are all consequences of curse of dimensionality.

4. AMOUNT OF TRAINING DATA

In figure , we see curve ‘red’ shows some oscillating behaviour. This is due to data set; it has lot of degrees of freedom to fit the set. It does not compromise as other two graphs does.

This is because we can add data points for training set and eventually we would reach when no free parameters are present to fit the curve. After this point the curve stops oscillating and improves generalizations.

Hence it can be said that for MLP, large network require more data so as to maintain good generalization capability.

$$N = O\left(\frac{W}{\varepsilon}\right) \qquad (3.24)$$

$N \rightarrow$ No. of training points

$W \rightarrow$ No. of weights in n/w

$\varepsilon \rightarrow$ Fraction of permitted errors on test data set

Overall, it can be said that generalization is depending on both size of network and training data set.

3.6 PATTERN RECOGNITION

Pattern recognition is the act in which we can recognize the raw data or the category of any pattern. It include examples like identifying a face, to read handwritten characters, to understand the spoken words, to decide if apple is ripe or not by its smell. So, to recognize all this one need to have knowledge for it.

For all the tasks, there was the need for designing and building machines which would recognize patterns. For fingerprint identification, accurate pattern recognition, DNA sequence identification and much more so we need automatic system for accurate classification.

Set of i/p values is termed as i/p pattern and of o/p is o/p pattern. NN learns the relationship of input and output as:

$x=(x1,\ldots,xn) \rightarrow y=(y1,\ldots,ym)$ (3.25)

1) Model

It may be mathematical; shows and explains the features and characteristics it has. All models are unique and hence have different characteristics too.

2) Pre- processing

Without loosing any important information, the operations are simplified for further use of the model. A model may be in the form of images, data set or speech.

3) Segmentation

It is the process in which models are separated from their original position for further use.

4) Feature Extraction

There is a feature extractor which aims at reducing data by evaluating properties or features. These resulting features and their values are finally passed into a classifier. It computes the data present and thus gives a final decision.

These algos aim at giving an answer to the i/ps which match with the training data. Even if complete data is not provided, it checks for the previously known patterns.

Hence we can say it categorizes the pre- existing and new data into sub- classes by assigning them labels for whichever class they belong to.

3.7 LEARNING

For any type of network, learning is very important as the classifier learns from examples or the data and thus gives the output. Learning is divided into parts depending on the kind of information given to network.

3.7.1 Supervised learning

In this both input and output patterns are presented. Each input pattern is taken to train the network which is linked with desired output pattern or target. A teacher has to be present during learning session as comparison is done between the corrupted output network and correct expected output so as to determine the error. Regression and Pattern Recognition are tasks performed under this category.

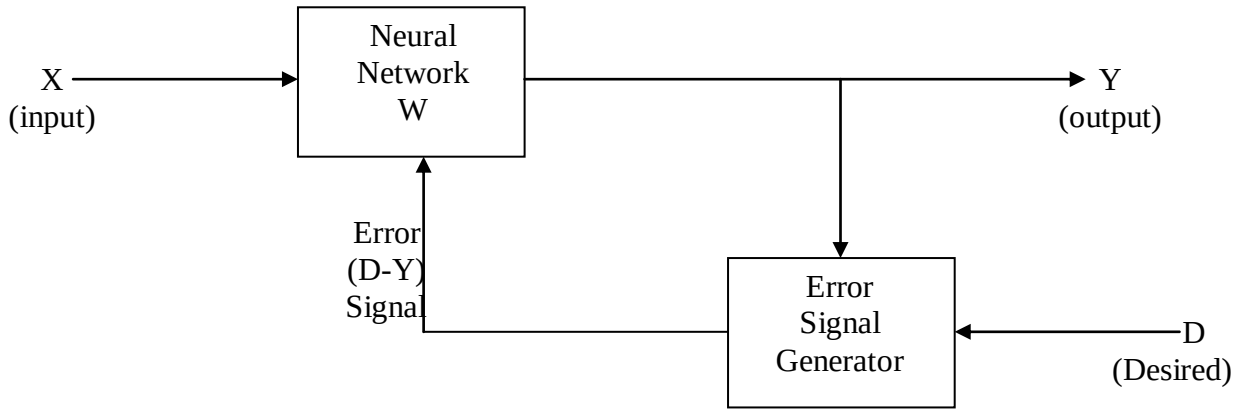


Figure 3.8: Supervised learning

The desired pattern(o/p) is known as target.

$$T= \{tl1,tl2,\ldots tlnl\} \quad (3.26)$$

This learning has two components :-

During learning, network's performance is measured and for the improvement in performance, an algorithm is used. Error function is used hence giving qualitative error.

$$E = \frac{1}{2P} \sum_{P=1}^{P} \sum_{i=1}^{M} (T_i^P - V_{LA}^P)^2 \quad (3.27)$$

M- no. of o/p units, P- no. of patterns

It is very popular as its easy to manage mathematically by squaring the distance.

3.7.2 Unsupervised learning

In this learning, target p/p is not presented to network. This system learns by adapting to the structural features in the i/p pattern as if no teacher can present desired patterns. Clustering, filtering and compression fall under this category.

3.7.3 Reinforced learning

In this technique, no expected answer is given; it is assumed that a teacher is present only for giving indication if the output which is calculated is correct or not. This information is used and thus help network in learning process.

3.8 ACTIVATION FUNCTIONS

- These functions limit output of all neurons in NN to certain output value in range of [-1,1] or [0,1].
- Any activation function for the back- propagation net has to be differentiable, continuous and monotonically non- decreasing.[]
- Various activation functions are:-

- Linear :

  $f_i(s) = cs$ (3.28)

- Threshold or Step :

  $f_i(s) = \begin{cases} 1 & \text{if } s > T \\ 0 & \text{otherwise} \end{cases}$ (3.29)

- Hyperbolic Tangent :

  $f_i(s) = \left(\frac{1-e^{-cs}}{1+e^{-cs}}\right)$ (3.30)

- Gaussian :

  $f_i(s) = e^{\frac{-s^2}{v}}$ (3.31)

3.9 GRADIENT DESCENT

Training is the process in which network weights are adjusted so as to give a desired function. It is one of the examples of 'machine learning'; or specifically said it's a survey of various machine learning techniques[]. A short review of gradient descent algorithm is given with its background. Also, Scaled Conjugate Gradient (2nd order method) is described.

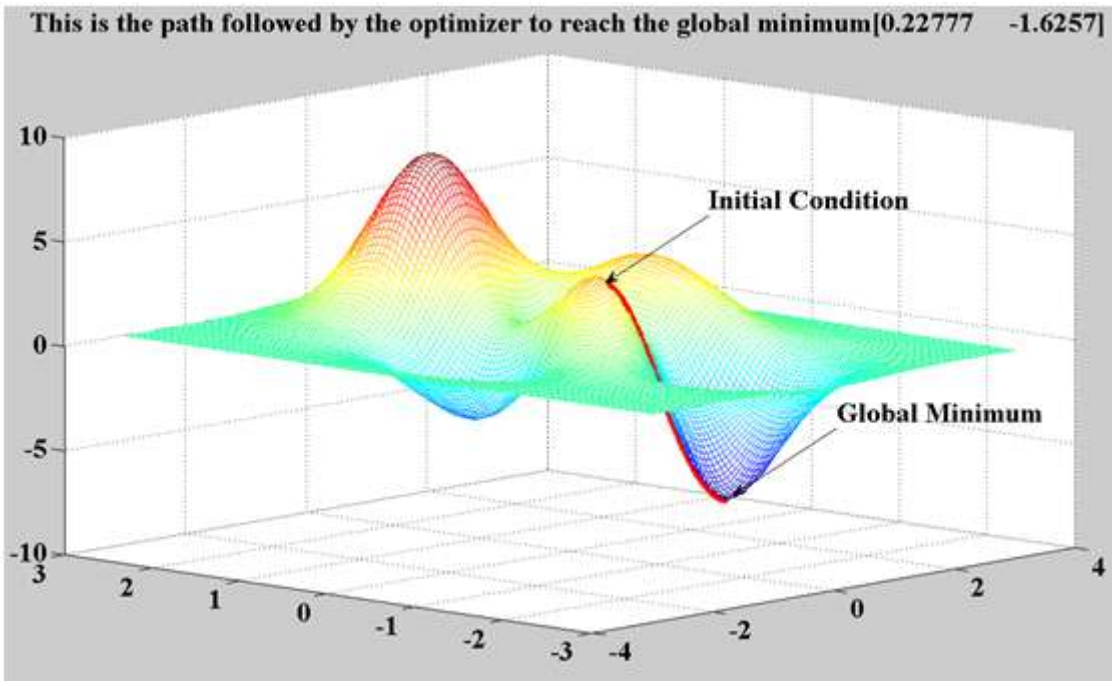


Figure 3.9: Graph for gradient descent method

Using Taylor approximation (1st order), w. r. t 'w', change in error with o/p as:

$$E(\mathrm{w}+\Delta\mathrm{w}) \approx \mathrm{E}(\mathrm{w}) + \Delta w^{T} E'(w) \tag{3.32}$$

$$\eta \rightarrow \text{step size /learning rate}$$

For minimising an error function is the method of Gradient Descent. The method says if error decreases

$$\Delta\mathrm{w} = -\eta \frac{\delta \mathrm{E}(\mathrm{w})}{\delta \mathrm{w}},\ \eta > 0 \tag{3.33}$$

η can be found by line search method. If the eta is chosen properly at every step; this method is referred as steepest descent method. Its of following two types:

1. off- line mode: accumulation of partial gradient vectors, the gradient vector is formed for every pattern present in training set.

2. on-line mode: the method is performed on every partial error function successively and hence linked with each pattern of training set.

$$\Delta w = -\eta \frac{\delta E^P(w)}{\delta w} = -\eta E'^P(w),\ \eta > 0 \tag{3.34}$$

where $\frac{\delta E^P(w)}{\delta w} \rightarrow$ error gradient (with pattern P)

When η tends to zero with time, during 1 epoch, movement in weight space is obtained with an off- line update. The network adjust weights after every presentation of any imput pattern so if for same pattern an improved result would be got. This is called an epoch i. e. full presentation for all patterns.

Generally, learning rate accelerates convergence, as the paths in these weight spaces differ.

Hence, on- line mode is more preferable over off- line mode; training set gets large and has redundant information. It gets a clear path specially for classification problems where targets can just be approximated.

Theoretical discussions regarding issues containing on-line and off-line techniques is found []. In real, problems are confronted with on-line methods if patterns are differently distributed from different classes.

Gradient descent method usually finds weight position which is minimum for the class, if lot of patterns form a large class. This weight position not necessarily the best position for other classes so search for new weights is carried on till the error becomes smaller and hence it gives the position of local minima.

3.9.1 Scaled- Conjugate Gradient

The disadvantages of Back Propagation are its slow convergence and a need to set few learning parameters. Few methods like 2^{nd} order Taylor approximation had been adapted[]. Method like Quick- Prop2 by[] and Conjugate Gradient have become popular. The main idea of this algorithm is to produce for 2^{nd} order and with non-interference of direction of search. This method is a variation of the standard conjugate algorithm [].

The whole space is spanned from d_t to d_t+1 and error is minimised starting from dt to dt+1

So, weight update is given by:

$$\Delta w_t = \epsilon_t d_t \tag{3.35}$$

$$\epsilon_t \rightarrow \text{step size}, d_t \rightarrow \text{search direction}$$

$\epsilon_t$is found by line search and is time consuming because it involves calculation of first derivative or the error. But in scaled c g, step size can be found by scaling mechanism and hence consumption of the time is less.

$$d_{t+1} = -E'(w_{t+1}) + \beta_t d_t \tag{3.36}$$

$w_t \rightarrow$ vector containing all weight values
$t \rightarrow$time step

$$\beta_t = \frac{\left|E'(w_{t+1})\right|^2 - E'(w_{t+1})^T E'(w_t)}{\left|E'(w_t)\right|^2} \tag{3.37}$$

Step size

$$\epsilon_t = \frac{-d_t^T E'(w_t)}{d_t^T s_t + \lambda_t |d_t|^2} \tag{3.38}$$

where s_t is

$$s_t = \frac{E'(w_t + \sigma_t d_t) - E'(w_t)}{\sigma_t} \tag{3.39}$$

ϵ_t minimises 2^{nd}order approximation to error function.

λ_t is the scaling parameter.

$$\Delta_t = \frac{E(w_t) - E(w_t + \epsilon_t d_t)}{E(w_t) - E_q(w_t + \epsilon_t d_t)} \tag{3.40}$$

$E(w_t)$ is real error,
E_q is quadratic approximation of error.

CHAPTER 4

METHODOLOGY

This chapter describes the procedure followed for the extraction of arteries and veins in the retinal blood vessels. A flowchart is presented in the end for easy understanding of the methods used.

4.1 MATERIAL

The digital images are being taken from Image Sciences Institute.

A DR programme was held at Netherlands. Total four hundred(400) Diabetic subjects' screening population consisted between age 25-90 years. Randomly forty(40) photographs were selected, out of which only seven(7) show signs of early mild Diabetic Retinopathy and thirty three showed no sign for it. Images have been JPEG compressed.

With a 45 degrees FOV(field of view), images were captured using a Canon CR5 non-hydria tic 3 CCD camera. Each image has been acquired using eight(8) bits per colour plane at seven hundred and eighty six by five hundred and eighty four (786*584) pixels.

With a diameter of approximately five hundred and forty pixels(540), FOV of each image is circular. The images are cropped around it. The total forty(40) images are divided into two sets: (20 each).

Training images- only single manual segmentation of vasculature is present.

Testing images- double manual segmentation is present. One is used for the gold standard; other can be used for the comparison for computer generated segmentations done by independent human observers. These all human observers had been trained and instructed by Ophthalmologists. They were asked to mask seventy percent (70%) certainty if they were vessels.

These images are actually kept for making clinical diagnosis. In terms of the privacy of all the patients, there names and details are not enclosed rather no information is being provided to anybody.

4.2 IMAGE PRE- PROCESSING

If at the time of result, there are any contrasts or colour variability it would effect the classification of results in the end. Thus image pre processing is required. Firstly, background illumination is compensated as in[42]; it is based on median filtering. This is done for illumination correction. Now this channel after being processed to extract centreline pixels. This is done in[41] and thus every vessel is represented by a set of centreline pixels.

4.3 FEATURE EXTRACTION

It is quite impractical to deal with raw data if there are large sets. It is just not because of the volume of data set but information in a single data point can be hidden if it's a sequence. Hence, feature extraction should be used.

In this, large data is being described in terms of few features. The good part is we can talk about an object in terms of its height, colour, weights etc. This might not give us 100% information but gives enough information to recognize it.

In the statistical analysis, it can be done with the help of algorithms which run on the input data and thus returning the features described by it. There is a lot of variety of features existing but are used depending on the problem. In case of digital features, features are calculated in frequency and time domain both.

After the centreline pixels were found, four coloured features were extracted MR, MH, MG and VR. These are extracted around the centre of centreline pixels. So, NN classifier uses these features for final classification into two classes: arteries and veins.

The final pixel classification is done based on following logic:

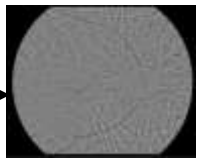
1. if label= 1; assign it 'a' (artery)

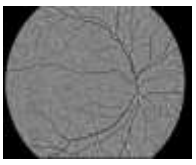
2. if label= 0; assign it 'v' (vein)

Hence, each pixel is thus classified whether artery or vein.

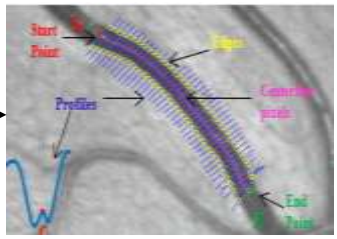
4.4 FLOWCHART OF THE TECHNIQUE

Illumination correction





Extraction of centreline pixels



Training Neural Network Classifier

Feature extraction

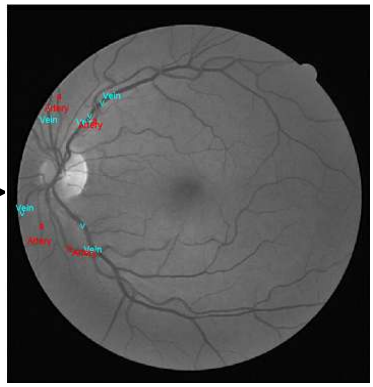
Mean of red (MR), Mean of green (MG), Mean of hue (MH), Variance of red (VR))

Testing the Data

Pixel Classification

Conversion of Pixel to Vessel

Final labelling & classification



CHAPTER 5

RESULTS AND DISCUSSION

To classify arteries and veins, Neural Network classifier is used. Feed forward network is the method used with Back Propagation and Probabilistic Neural Network and their results are being displayed.

NN toolbox is used with the command line functions for training the classifier for both Back Propagation and Probabilistic Neural Network.

Total number of samples are divided for:-

Training- network adjusts according to error

Validation- it measures generalization and may pause training if generalization after a point does not improve.

Testing- it provides with the evaluation of network performance during the training and after it also.

Few terms should be known for considering and understanding the results:

- Confusion plot- Elements on the diagonal row represents the no. of samples giving correct results and shown by green colour while red colour represents not correctly classified percentage of samples. Blue colour gives us the overall percentage of correct classification.
- ROC curve- Blue line in the graph shows the ROC curve. As the value of threshold changes, it tells us about sensitivity and 1- specificity where sensitivity is true positive rate and 1- specificity is false positive rate.Point being nearer to the left upper side show how correct is the response.
- Cross Entropy- it should be minimized for good results of classification. If the value is near zero, its best
- % error- tells us about misclassified samples,
  0 giving no misclassification.
- Performance plot tells us about at which iteration we are getting the best validation value.

- Regression plots the linear value w.r.t o/p. If its nearer to 1, the samples are giving that much correct response.
- Error histogram represents the error calculated. The large peak indicates the minimum error as it is nearest to the targeted value and the small peaks represent the incorrect responses.

1. BACK PROPAGATION NETWORK

Firstly the network architecture is created as shown below:

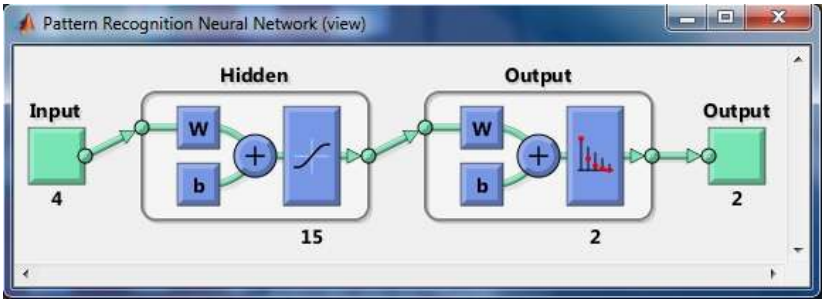


Figure 5.1: Network architecture of feed forward network

As it is seen, that the network has one hidden layer and number of neurons chosen are 15 for optimum results. Sigmoid function is the activation function used in the hidden layer as per requirement. Total no. of input units in input layer are 4 and corresponding outputs are 2 i.e. the problem would be classified for two classes- 'artery' and 'vein'.

Neural network Pattern Recognition tool (nprtool) is used for this classification. It helps us in selecting data, creating and thus training a network. After this the performance is evaluated using confusion matrices and regression value.

The network is trained with trainscg (scaled conjugate gradient backpropagation.).

The scaled conjugate gradient method has been explained earlier in chapter 5. Neural network training tool has also been shown and then given is the best validation performance for the number of iterations taken is 1000 by default.

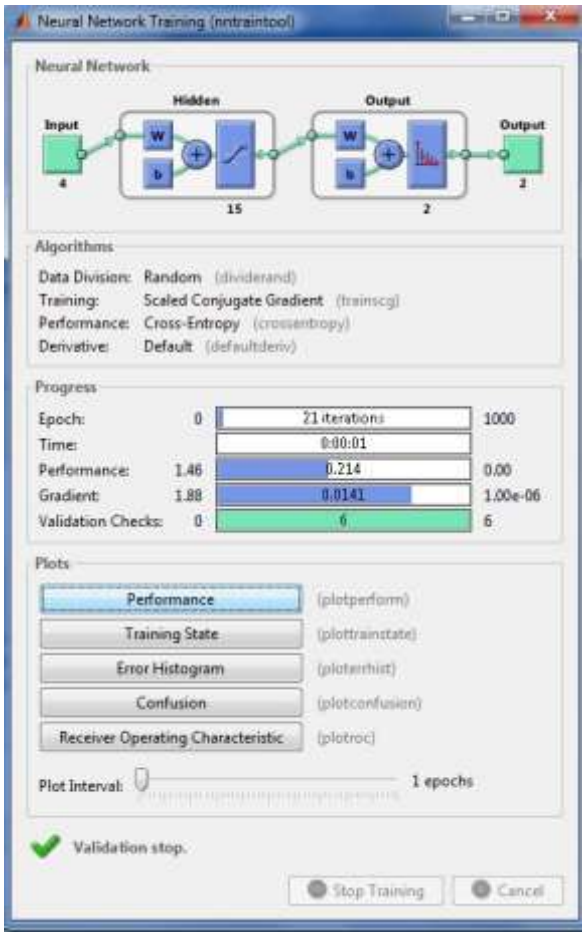


Figure 5.2: nntraintool

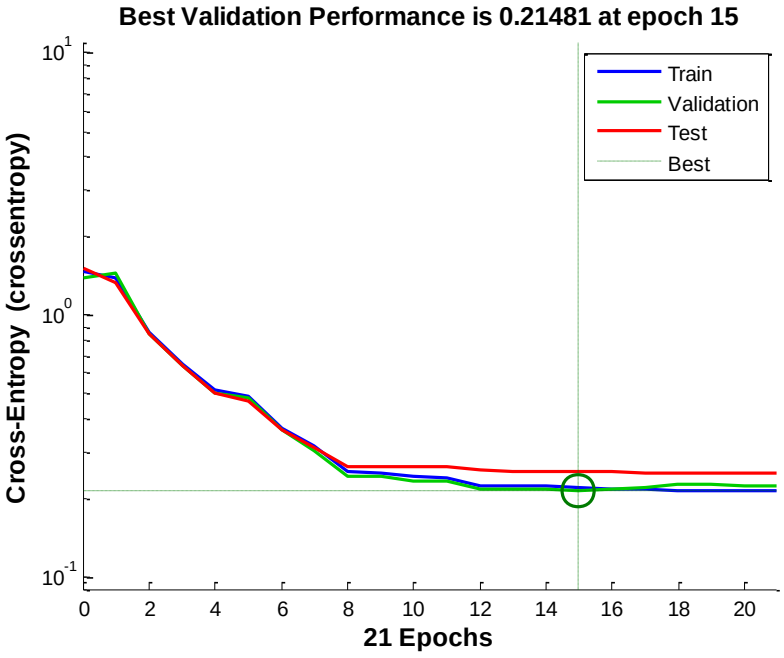


Figure 5.5: Cross- entropy error graph

8.2 PROBABILISTIC NEURAL NETWORK

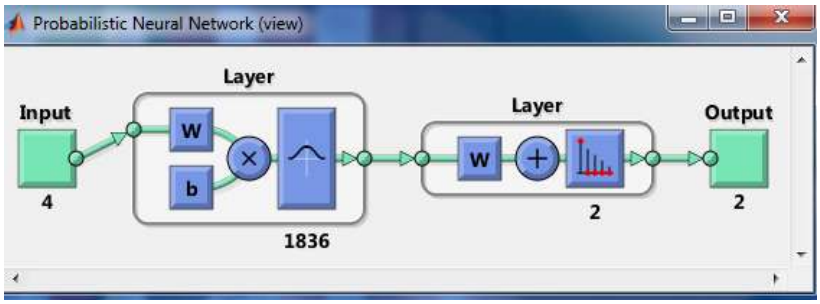


Figure 5.6: Network architecture of PN network

The architecture of PNN is shown which has taken Gaussian function as the activation function for pattern layer where 1836 are the no. of input samples and next layer is summation layer showing that two summation units are formed out of all classifying 'artery' and 'vein' so two outputs are shown.

This method is faster than BP but again only if the network is once trained earlier as for new samples which need to be trained, the technique takes time to learn. In this algorithm, supervised learning is being used.

5.1 COMPARISON OF RESULTS

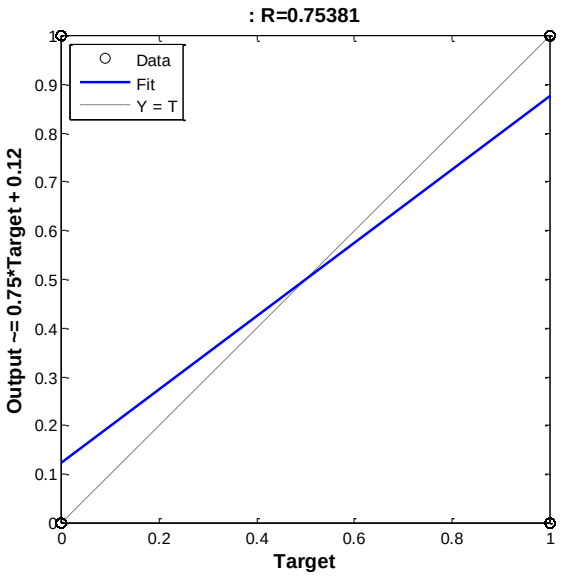


Figure 5.7(a): Regression (BP)

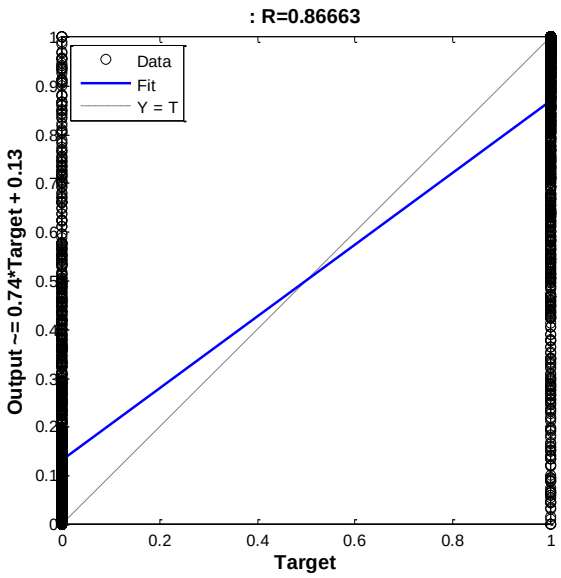


Figure 5.7(b): Regression (PNN)

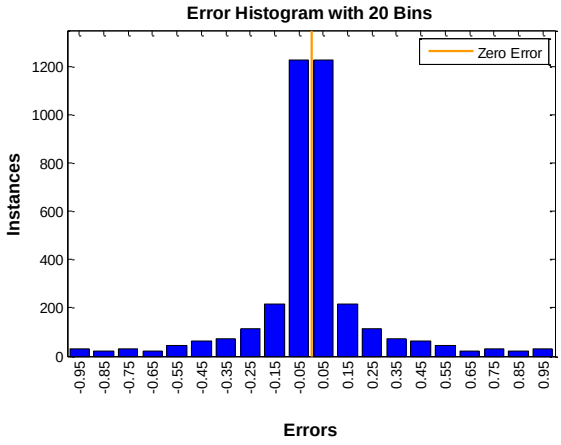


Figure 5.8(a): Error histogram (BP)

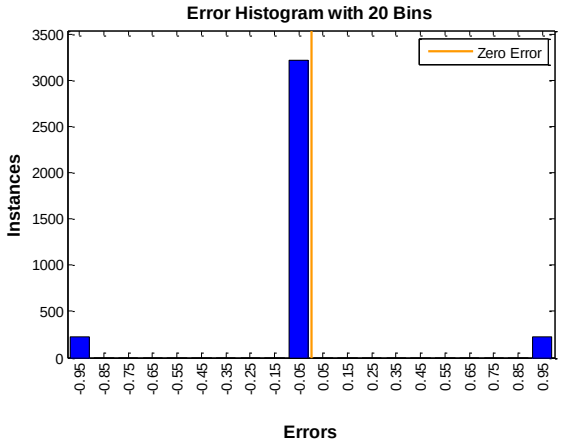


Figure 5.8: Error histogram(PNN)

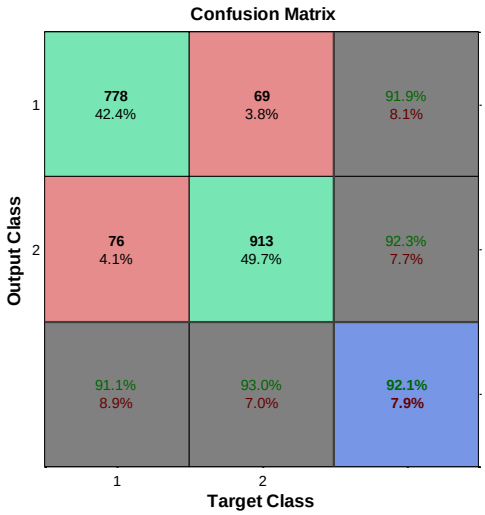


Figure 5.9(a): Confusion plot (BP)

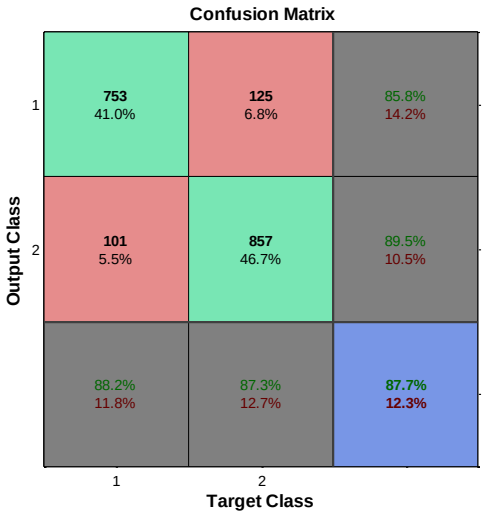
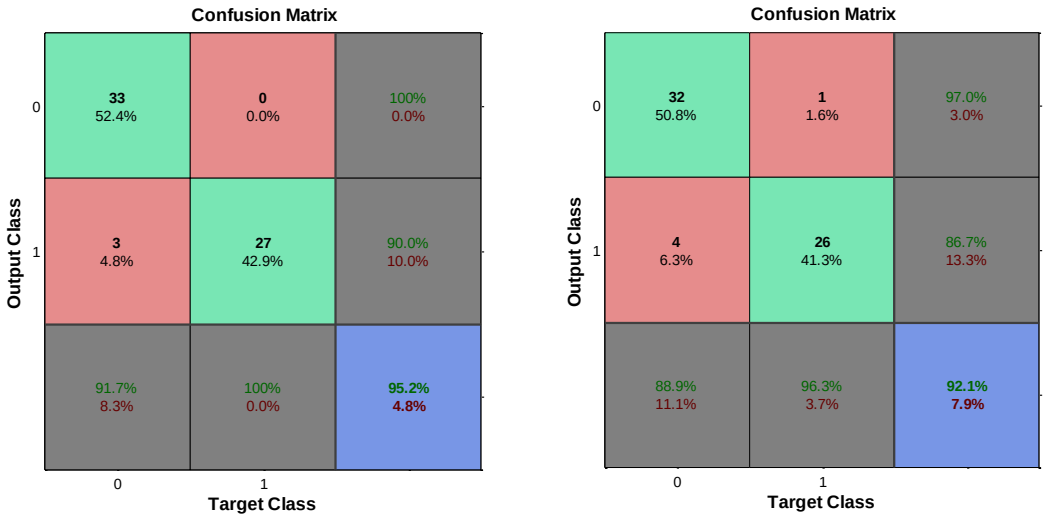


Fig 5.9(b): Confusion plot (PNN)

As it seen from the figures, that regression value of BP is more nearer to 1 than PNN. Also, total truly classified percentage for PNN is 87.7% and BP is 92.1%.

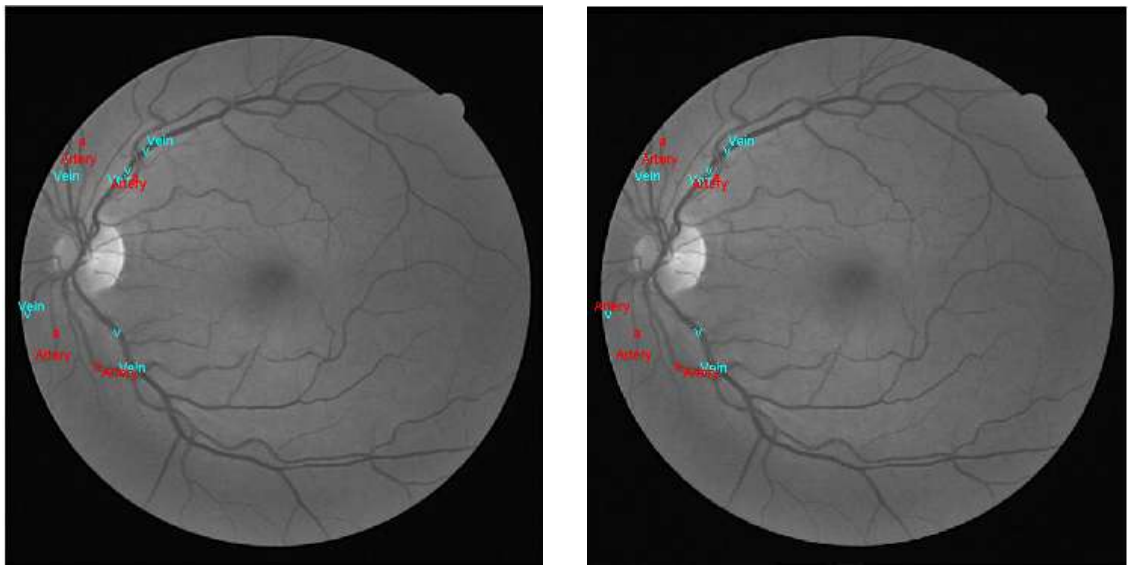
Corresponding outputs after pixel classification for randomly 4 images are shown below with their confusion plots, and corresponding outputs for vessel classification are given in the image of eye classifying arteries and veins. As it can be seen arteries are represented by red colour and veins by fluorescent green colour. So, order wise from first to fourth image for both the networks, the results are being displayed below:-

1. Image 1



(a) (b)

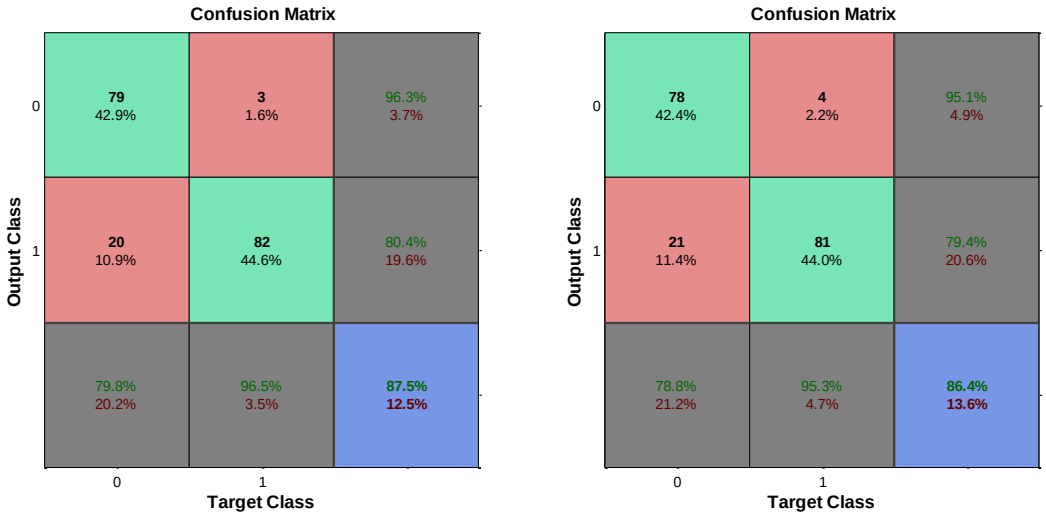
Figure 5.10: Pixel classification Result (a) BP (b) PNN



(a) (b)

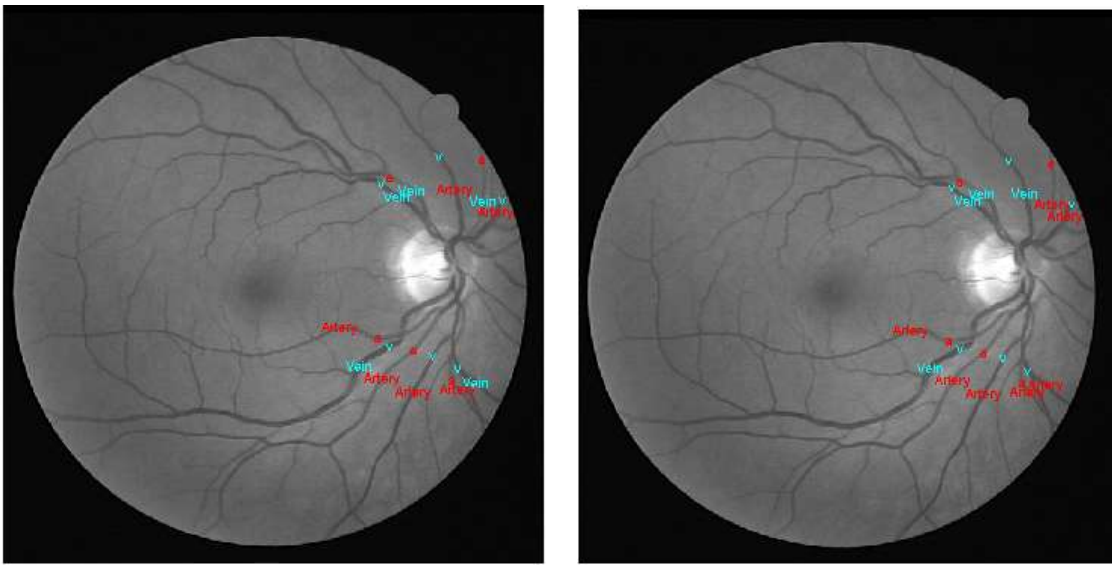
Figure 5.11: Vessel classification Result (a) BP (b) PNN

2. Image 2



(a) (b)

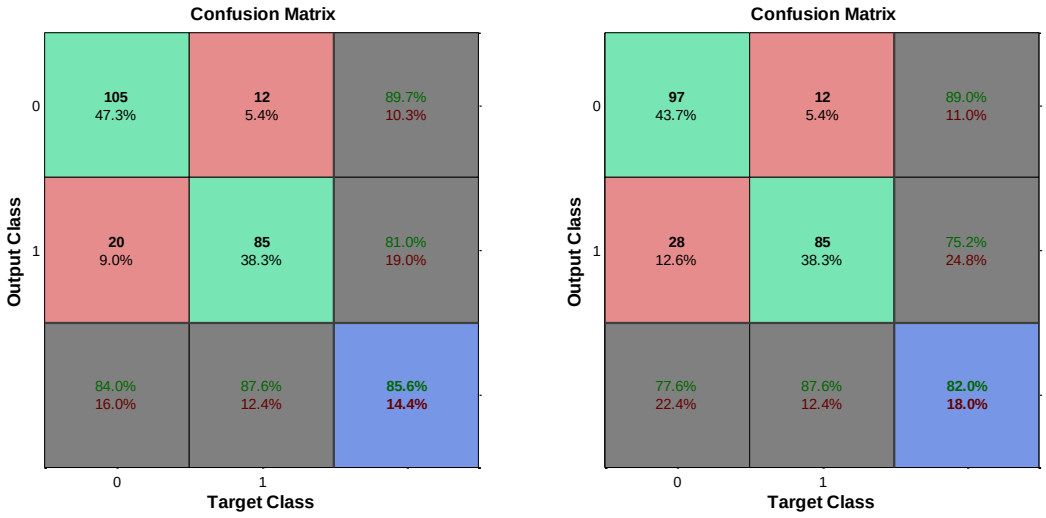
Figure 5.12: Pixel classification Result (a) BP (b) PNN



(a) (b)

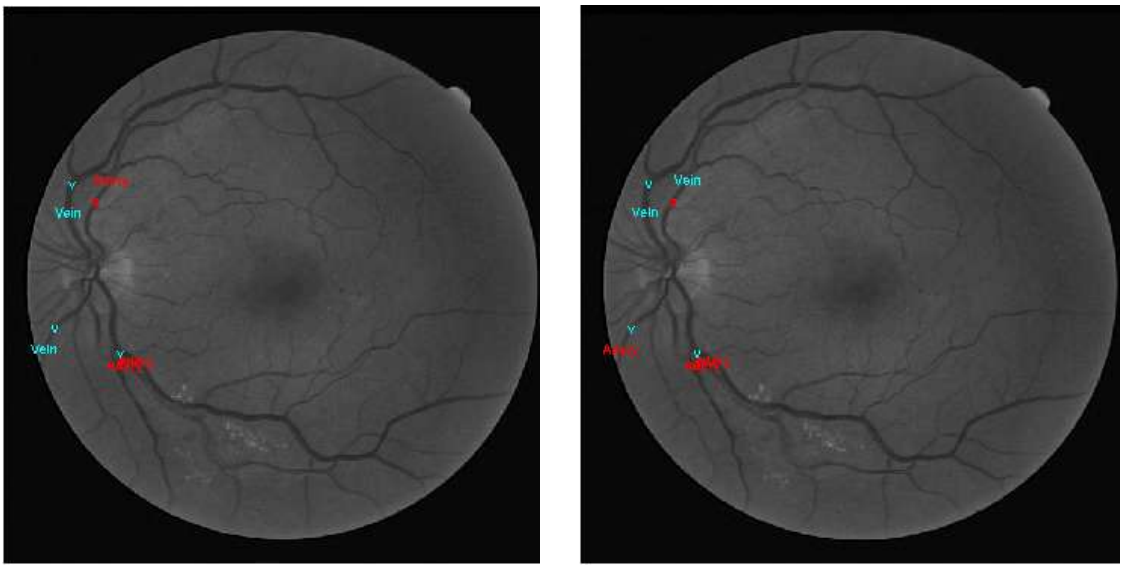
Figure 5.13: Vessel classification Result (a) BP (b) PNN

3. Image 3



(a) (b)

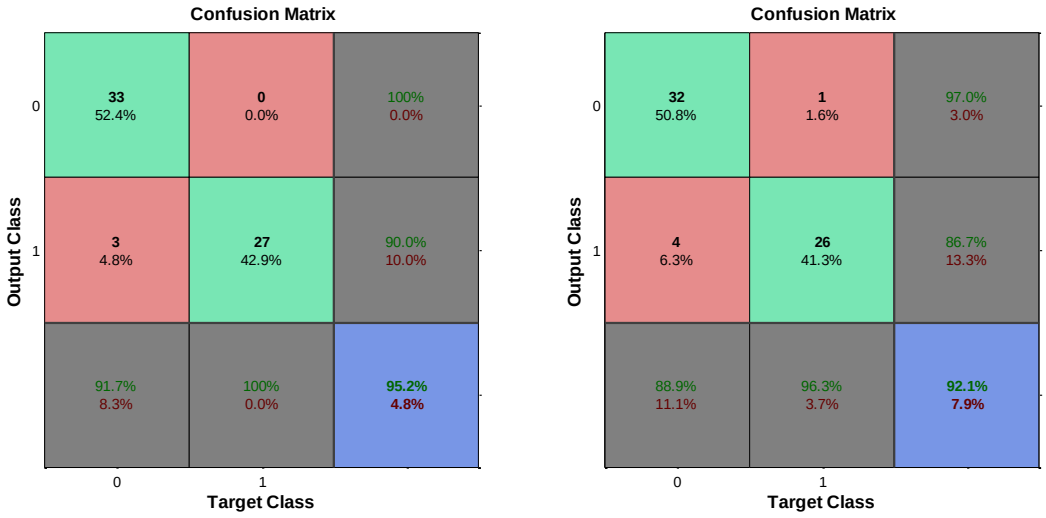
Figure 5.14: Pixel classification Result (a) BP (b) PNN



(a) (b)

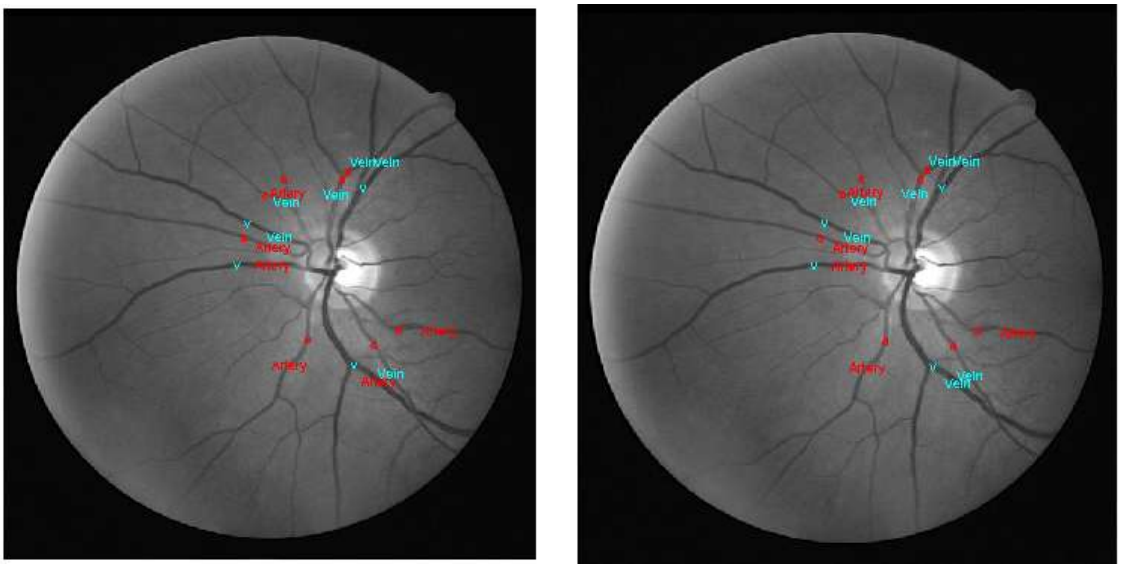
Figure 5.15: Vessel classification Result (a) BP (b) PNN

4. Image 4



(a) (b)

Figure 5.15: Pixel classification Result (a) BP (b) PNN



(a) (b)

Figure 5.16: Vessel classification Result (a) BP (b) PNN

Table 5.1: Comparison of results

	BP	PNN
Regression	0.86663	0.75381
Performance value	0.2241	0.1231
Pixel classification (%)	87	84
Vessel classification (%)	64	62.7
Correctly classified samples in training state (%)	92.1	87.7

From above results in the table and figures displayed, it can be seen that Back Propagation network is more efficient than Probabilistic Neural Network. It is a generalised network which can be used for various kinds of problems and thus this is more efficient in the project undertaken.

CHAPTER 6

CONCLUSION AND FUTURE SCOPE OF WORK

6.1 CONCLUSION

Many eye diseases usually lead to blindness which cannot be cured on time as the diagnosis is done very late on the patients. In my work, NN classifiers are used after being trained so as to develop an automatic diagnostic tool. Physicians should detect the eye abnormalities with this tool. Results in the end show that the NN classifier is itself better than statistical classifiers.

Both BPN and PNN yield good results 87% and 84% respectively. The classifiers work on the DRIVE dataset. Firstly the classifiers are trained, and then features are extracted from the pre processed retinal images. Testing is done after the extraction and hence this gives us pixel classification for both the classifiers. Then these pixels are converted to vessels and hence vessels are being assigned with the result of either ‘a’ (artery) or ‘v’ (vein).

But PNN classifier takes more time than BPN as it takes long in training the classifier for the first time. PNN also uses more space to save the results comparative to another. BPN is a generalized network though PNN is not. Both are used for feed forward multi layer network. PNN uses Gaussian function as the activation function for simple computation. So, overall while comparing the results of both classifier I can say that BPN is more efficient in my problem. The classification result of my work is finally showed through images and blood vessels are labelled as arteries and veins.

Accuracy of the result is depending on factors like feature extracted or the parameter used. If few changes are made in network architecture or the training algorithm, speed of execution of network increases. Also if some parameters are changed or added, it would lead to better performance and results in the end.

6.2 FUTURE SCOPE OF WORK

Retinal blood vessels are indicators for treatment of eye diseases, clinical diagnosis and also for systemic diseases such as diabetes, hypertension etc. These diseases manifest themselves in retina and are observed in fundus imaging. It is essential to classify

vessels into arteries and veins for the discovery of biomarkers in retinal vasculature. Changes in vasculature appear with the onset of a systemic disease and thus it affects both arteries and veins differently. One of the examples is early sign of retinopathy leads to lower AVR. This leads to narrowing of retinal arterioles which are linked with hypertension. Since, AVR is one of the best predictor of stroke and other cardiovascular diseases, so there is a need of automatic systems for vascular characterization which is a computation method for classifying arteries and veins.

NN classifiers are suitable for pattern recognition problems with supervised learning. But sometimes unsupervised learning might also be required. So, NN is one of the optimum classifier which can thus be used. The classification of arteries and veins in blood vessels is used as biomarkers for the detection of many systemic diseases. A lot of research is going on in the medical area since long and will continue for long. Not just the sorting of vessels but also AVR ratio is calculated for detecting retinal diseases. As the features are extracted from three colours, there may be many more features not just mean, standard deviation and variance. Features based on length, diameter, shape, luminosity can also be extracted and used for better performance and thus classification. Hence, for detection of biomarkers, many new techniques with different methodology can be done.

REFERENCES

[1] M. W. and W. Pitts, "A Logical Calculus of Ideas Immanent in Nervous Activity," *Bulletin of Mathematical Biophysics,* vol. 5, no. 4, p. 115–133, 1943.

[2] C. S. Flick, "Centenary of Babbage's ophthalmoscope," *Optician,* pp. 113,246, 1947.

[3] D. Hebb, "The Organization of Behavior", New York: Wiley and sons, 1949

[4] .B. Farley and W. Clark, ""Simulation of Self-Organizing Systems by Digital Computer"," *IRE Transactions on Information Theory,* vol. 4, pp. 76-84, 1954.

[5] N. Rochester, J. H. Holland, H. L. Habit and W. L. Duda, ""Tests on a cell assembly theory of the action of the brain, using a large digital computer"," *IRE Transactions on Information Theory,* vol. 2, pp. 80-93, 1956.

[6] F. Rosenblatt, ""The Perceptron: A Probalistic Model For Information Storage And Organization In The Brain"," *Psychological Review,* vol. 6, no. 65, pp. 386-408, 1958.

[7] S. D. G., Pandemonium: a paradigm for learning,"Mechanisation of Thought Process," in *Synopsium held at National Physics Labortary,* London, 1958.

[8] M. Minsky and S. Papert, *"An Introduction to Computational Geometry"*, MIT Press, 1969.

[9] D. S. Halacy and C. Babbage, London: Crowell- Collier Press, 1970.

[10] M. Matsui, T. Tashiro, K. Matsumoto and S. Yamamoto, " A study on automatic and quantitative diagnosis of fundus photographs.," *Nippon Gakka Zasshi,* vol. 77, no. 8, pp. 907,918, 1973.

[11] B. Webros and P. J., *"Beyond Regression: New Tools for Prediction and Analysis*

in the Behavioral Sciences", 1975.

[12] H. Tachibana, F. Gotoh and Y. Ishikawa, "Retinal vascular autoregulation in normal subjects," *STROKE,* vol. 13, pp. 149-155, 1982.

[13] C. E. Baudoin, B. J. Lay and J. C. Klein, " Automatic detection of microaneurysms in diabetic Fluorescein angiography," *Epidemiol Sante,* vol. 32, no. 3-4, pp. 254-261, 1984.

[14] J. D. Bolter, Western Culture in the Computer Age, 1984.

[15] D. Rumelhart and J. McClelland, *" Parallel Distributed Processing: Explorations in the Microstructure of Cognition",* Cambridge: MIT Press, 1986.

[16] B A. and E. Rosenfeld, *Neurocomputing: Foundations of research,* Cambridge: MIT Press, 1987.

[17] S. Fahlman, D. Touretzky, G. Hinton and T. Sejnowski, "Fast Learning Variations on Back-Propagation: An Empirical Study," *Connectionist Models* , pp. 38-51, 1988.

[18] J. Jonas, G. Gusek and G. Naumann, Parapapillary diameter of retinal vessels;estimating the size of the optic papilla, Klin Monatsble Augenh, 1988.

[19] N. D. and B. Widrow, "Improving the learning speed of -layer neural networks by choosing initial values of adaptive weights," in *IEEE*, 1990.

[20] S. Kocabas, "A review of learning," *The Knowledge Engineering Review,* vol. 6, no. 3, pp. 195-222, 1991.

[21] R. Battiti, "First and Second-Order Methods For Learning; between Steepest Descent and Newton's method," *Neural Computing,* vol. 4, no. 2, pp. 141-167, 1992.

[22] A. Alm and W. Hart, Physiology of the eye, Clinical Application, U.S.A: Mosby Year book, 1992.

[23] M. M., *A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning*, 1993.

[24] M. Moller, "Efficient Training of Feed Forward Network," Computer Science department, Aarhus, 1993

[25] H. Kolb, R. Nelson, E. Fernandez and B. Jones, The Organization of the retina and visual system, 1995.

[26] Keeler.C.R, "150 years since Babbage's ophthalmoscope," *Ophthalmol,* vol. 115, pp. 1456-1457, 1997.

[27] Athseda, "Cardiovascular,lung and blood diseases", US, 1998.

[28] Lewallen, S. Harding and S. Ajewole, "A Review of the Spectrum of clinical ocular fundus findings in P.Falciparum malaria in african children with a proposed classification and grading system," *Science Direct,* vol. 6, no. 93, pp. 619-622, 1999.

[29] M. Perz, M. Hughes, A. Stanton and S. Thom,"Retinal blood vessel segmentation by means of scale-space analysis and region growing," *MICCAI,* vol. 1679, pp. 90-97, 1999.

[30] A.Colchester, C.Taylor, "State scale analysis for charcterization of retinal blood vessels,in medical image computing and Computer Assisted Intervention" Springer, 1999.

[31] P. R., "Neural network"," *IEEE Transactions,* vol. 2, no. 2000.

[32] T. Y. Wong, R. Klein, B. E. Klein and J. M. Tielsen, ""Retinal microvascular abnormalities and their relation to hypertension,cardiovascular diseases and morality"," in *IEEE*, 2001.

[33] Dr.M.H.Dunhan, "Traffic engineering with AIMD in MPLS networks", Data Mining , Introductory and Advanced topics, Prentice Hall, 2002.

[34] C. N. Deliomo, M. Sartini, M. Witt, A. Hughes, S. Thom and R. Pedrinelli,

"Peripheral Vascular Disease is associated with abnormal arteriolar diameter relationships at bifurcations in the human retina," London, 2002. [1]

[35] E. G,. A. Ruggeri, "A divide et imperia strategy for automatic classification of retinal vessels into arteries and veins" in *IEEE*, University of Padova, Italy, 2003.

[36] Staal, J, Abram,M.,M.Niemeijer,M.A.Viergever, " Ridge based vessel segmentation in colour images of the retina," in *IEEE* Transactions on Medical Imaging,2004.

[37] Hinton, G. E Osindero, S.; Teh, Y. W. (2006). "A Fast Learning Algorithm for Deep Belief Nets". *Neural Computation* **18** (7): 1527–1554.

[38] S. Kumar, *Neural Network A classroom approach*, Tata Mc Graw Hill. S. L. Young, M. Hagenbucnner, A. C. Tsoi, F. Scarselli and M. Gori, "Document mining using Graph NN" in *IEEE*, 2006.

[39] J. A. Gardener, T. Kirchhof and B. Ryan, Retinal Vascular Diseases, Springer, 2007

[40] .G. Liew, J. Wang, P. Mitchell and T. Wong, Retinal Vascular imaging, a new tool in microvascular diseases., Cardiovascular imaging, 2008.

[41] I. Russell, *"Neural Networks Module"*, 2012.

[42] D. Ciresan., A. Giusti., L. Gambardella and J. Schmidhuber, " Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images", in *NIPS*, Lake Tahoe, 2012.

[43] C. Ciresan, U. Meier and Schmidhuber, "Multi-column Deep Neural Networks for Image Classification", in *IEEE Conf. on CVPR*, 2012.

[44] S. Rajasekaram and V. Pai, Neural network, Fuzzy logic and GA, Prentice hall of India, 2012.

[45] D. Relan, T. Macgillivray, L. Ballerini and E. Trucco, "Retinal vessel classification:sorting arteries and veins", in *IEEE*, Edinburgh, U.K, 2013.

[46] P. Agrawal, D. Stansbury, J. Malik and L. J. Gallant, "Pixels to Voxels:Modelling Visual Representation in the Human brain," 2014.

[47] "Vision RX," [Online]. Available: http://www.visionrx.com. [Accessed June 2014].

[48] "The my eye world," [Online]. Available: http://www.myeyeworld.com/files,2014.

[49] ."EMedicine Health," [Online]. Available: http://www.EMedicinehealth.com. [Accessed May 2014]

[50] "The Bathseda Retina," [Online]. Available: http://www.bethesdaretina.com/library.htm. [Accessed June 2014].