PREDICTING BREAST CANCER BY USING ARTIFICIAL NEURAL NETWORK

A MASTER'S THESIS

In

Computer Engineering

Atılım University

By

AYMEN FATHALLA H. ALHASADI

JANUARY 2016

PREDICTING BREAST CANCER BY USING ARTIFICIAL NEURAL NETWORK

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

OF

ATILIM UNIVERSITY

BY

AYMEN FATHALLA H. ALHASADI

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

THE DEPARTMENT OF COMPUTER ENGINEERING

JANUARY 2016

Approval of the Graduate School of Natural and Applied Sciences, Atılım University.

_____

(Prof. Dr. İbrahim Akman)

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

_____

(Prof. Dr. İbrahim Akman)

Head of Department

This is to certify that we have read the thesis "Predicting Breast Cancer by Using Artificial Neural Network" submitted by "Aymen FATHALLA H. ALHASADI" and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

_____

Asst. Prof. Dr. Erol Özçelik

Supervisor

Examining Committee Members

Asst. Prof. Dr. Erol Özçelik _____

Asst. Prof. Dr. Gökhan Şengül _____

Asst. Prof. Dr. Murat Saran _____

Date: 29.01.2016

I declare and guarantee that all data, knowledge and information in this document have been obtained, processed and presented in accordance with academic rules and ethical conduct. Based on these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Aymen FATHALLA H. ALHASADI

Signature:

ABSTRACT

Predicting Breast Cancer by Using Artificial Neural Network

FATALLAH, Aymen

M.S., Computer Engineering Department

Supervisor: Asst. Prof. Dr. Erol Özçelik

January 2016, 85 pages

Breast cancer is ranked the primary cause of death among women in the world. The goal of this study is to find a robust method for predicting recurrence and non-recurrence of breast cancer after surgery. The Wisconsin Prognostic Breast Cancer (WPBC) database which includes 194 samples is used in this study. The reason for choosing this database is due to the fact that it contains real data regarding breast cancer. In this thesis, breast cancer prediction is implemented by using Multi-Layer Perceptron (MLP) and Generalized Regression Neural Network (GRNN). The results of the artificial neural networks are also compared with the ones obtained by Support Vector Machine (SVM). The best performance is from obtained when GRNN method is used. The results and future work are discussed.

Keywords: Artificial neural network, Generalized Regression Neural Network, Support Vector Machine, Breast cancer prediction.

ÖZ

Yapay Sinir Ağı Kullanarak Meme Kanseri Tahmini

Fatallah, Aymen

Yüksek Lisans, Bilgisayar Mühendisliği Bölümü

Tez Yöneticisi: Yard. Doç. Dr. Erol Özçelik

Ocak 2016, 85 sayfa

Meme kanseri dünyada kadınlar arasında başlıca ölüm nedeni olarak yer almaktadır. Bu çalışmanın amacı, ameliyat sonrası meme kanseri tekrarını tahmin eden gürbüz bir yöntem bulmaktır. Bu çalışmada 194 örnek içeren Wisconsin Prognostik Meme Kanseri veritabanı kullanılmıştır. Meme kanseri ile ilgili gerçek veriler içerdiği için bu veritabanı seçilmiştir. Bu tezde, meme kanseri tahminini için çok katmanlı perceptron ve genelleştirilmiş regresyon sinir ağı işe koşulmuştur. Yapay sinir ağları ile ulaşılan sonuçlar ayrıca destek vektör makinesi ile elde edilenlerle karşılaştırılmıştır. En iyi sonuç, genelleştirilmiş regresyon sinir ağı yöntemi kullanıldığında bulunmuştur. Sonuçlar ve gelecek çalışmalar tartışılmıştır.

Anahtar Kelimeler: Yapay sinir ağları, Genelleştirilmiş regresyon sinir ağı, Destek vektör makinesi, Meme kanseri tahmini

DEDICATION

To my parents,
Fathalla, Karima,
and all of my family,
without whom none of my success would be possible

ACKNOWLEDGMENTS

Special appreciation goes to my supervisor Asst. Prof. Dr. Erol Özçelik for his supervision and constant support. His invaluable help of constructive comments and suggestions throughout the experiment and thesis works have contributed to the success of this research. I would also like to thank my examining committee members, Asst. Prof. Dr. Gökhan Şengül and Asst. Prof. Dr. Murat Saran for their constructive comments and feedbacks.

Sincere thanks to all my friends, especially Ashref BenzeigTariq Khalef, Abdulbasit Sultan, Salem Alessawy and others for their kindness and moral support during my study. Thanks for the friendship and memories.

I would also like to thank my parents, and elder brothers, specifically Mr. Farg, and Mr.Ossama. They were always supporting me and encouraging me with their best wishes.

Furthermore, I would like to express my gratitude to my brother-in-law Mr. Ramadan Bensalem and sister in-law Mrs. Ghada Bensalem for their unfailing support and help during my stay in Turkey.

Last, but not least, without the support of my wife Eman Bensalem. She always there cheered me up and stood by me through the good times and bad.

And also my children, Huor, Hood, Yusuf, and Bana, I would never have succeeded. Although they often had to endure my absence they seldom complained. I hope they will enjoy the effort being completed. Although, hardly aware of their stimulating role a word of gratitude is also spoken to the other members of my family and my friends. Unfortunately, my father is no longer with us to celebrate the completion of my thesis.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1
INTRODUCTION

1.1.Background of the Study

Cancer is a genetic disease that is characterised by cell growth. The tumour has the ability to destroy nearby tissues by invasion and spread to other parts of the body by metastasis. Despite significant improvements in early detection of cancer and new treatment strategies, more than sixty percent of the people diagnosed with cancer worldwide still died[1]. Breast cancer is one of the most commonly seen cancer types, which is unique to women. Figure 1.1 shows how breast cancer starts.

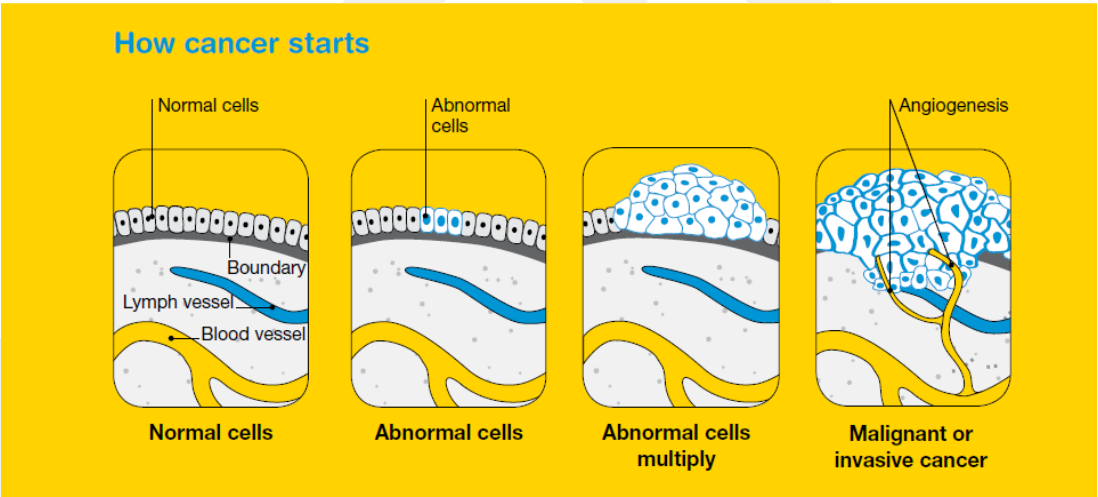


Figure 1. 1 Breast cancer state[2]

In this figure, the first image on the left are the normal cells while the second demonstrates the abnormality of the breast cells. After some period, the abnormal cells appear and the abnormal section will increase as this cancer gets multiple cells, which can be seen in the right two images. The final image illustrates how much the cancer cell took over. This is the most dangerous area of cancer.

The breast cancer starts once the cells in the breast start to grow and then can spread to nearby tissues by invading the body. Large groups of this out of control tissue are named tumours. They cannot spread or threaten someone's life, as some tumours are not actually cancer. These are called benign tumours. What can invade the body and spread to close tissues is considered as cancer and is named malignant tumours. In theory, malignant tumours in the breast are detectable by touching the breast. An unusual feeling in the breast can be a sign of breast cancer and should be treated cautiously.

Among the various kinds of cancer diseases, breast cancer is one of the most dangerous cancers; besides, it is considered the second biggest cause of death among women aged between 45-55 years. According to the American College of Radiology statistics, 1 out of 9 women might develop breast cancer during their lifetime [1]. Therefore early diagnosis of cancer plays a very important role. Diagnosing the breast cancer type and propagation velocity are the most important factors for successful treatment. Thus, early detection tests for breast cancer can save thousands of lives every year.

Approximately 269,000 breast cancer deaths occur in countries with low and middle incomes due to the fact that the disease is diagnosed at late stages which decrease the survival chance (lack of awareness of early detection and limitations in health services); however between 1989-1995 the breast cancer mortality rates declined by 1.4% per year and by 3.2% afterwards [1]. These declines have been attributed, in large part, to early detection [3]. Early tumour detection is the main cause of reducing mortality and the most effective treatment method to redress the disease without resorting to massive surgeries.

For early detection, mammography imaging is one of the most effective tools because of its ability to detect abnormalities more than 2 years before they occur as tumours. In fact, a tumour at the early stage appears as tiny shiny spots in the mammography images and these spots are calcium sediments named microcalcifications (MCs) and these are regarded as one of the most important signs of breast cancer. MCs have a subtle nature, in most cases unclear in the images and difficult to identify so this

constitutes a big challenge for radiologists due to the nature of the human vision system.

Conversely, the second type of breast cancer called masses may be easier to detect due to the nature of the size, shape and colour contrast. On the other hand, some types of masses can be difficult to detect, as they may look similar to normal breast tissue called parenchyma.

In order to ameliorate the radiologists' performance which has led to increased treatment efficiency and survival chance at the same time, reducing misdiagnosis, a lot of specialists in other technologies to help the radiologists are needed in reading mammography images and identifying suspicious regions in the breast and provide a second opinion for them to detect microcalcifications and masses earlier and to diagnose if those MCs or mass are benign or malignant. These technologies are Computer Aided Systems, which supply better detection and diagnosis techniques such as Computer Aided Detection (CADe) and Computer Aided Diagnosis (CADx).

Currently, most studies related to cancer susceptibility are either using the existing traditional probability testing or measurement techniques like statistical regression analysis. In this thesis, artificial neural techniques such as GRNN and multilayer perceptron will be examined for prediction of breast cancer. SVM is also used for the comparing its performance with the artificial neural network techniques.

1.2.Research Problem Statement

One of most frequent causes of dread among women in the world is that related to recurrence of malignant cases after a certain period; fear or worry of returning cancer. Recurrences of cancer usually develop within five years of treatment. However, 25 % of recurrences and half of new cancers in the opposite breast occur after 5 years [4].

In order to aid the physicians and medical scientists in making accurate decisions in the prognosis of cancer recurrence, data mining and statistical machine learning

techniques were used to analyze breast cancer dataset to predict cancer survivability and the results of each model's prediction and classification performances were also compared.

1.3.Goals of the Study

The goal of this study is to find a good and robust method for the prediction of recurrence and non-recurrence after surgery. The ultimate goal of this work is to develop predictive computer models for clinical decision support on several medical prognosis problems to decrease risks. It also aims to test the recurrence prediction model on a cross-validation data sample to counter artificial inflation in accuracy rates due to potential overtraining of the network, as well as to compare the performance of different classifiers(SVM, GRNN, and MLP) using clinical data.

1.4.Research Significance and Contributions

The main significance of this research is to determine if an accurate and transparent predictive model could be developed to serve as the basis for a future computerised screening tool. This research is using the GRNN, SVM and ANN with multilayer perceptron and comparing all these methods together. The core contributions of this work are both conceptual and practical. The conceptual contribution refers to the scientific references of learning and architecture for the models of intelligent response based on training technique. The practical contribution of this research work refers to the possibility of using the proposed computer prototype as a prediction tool that can be further developed into an automated technique for cancer detection.

1.5.Limitations of the Study

In this thesis, the breast cancer prediction is done and implemented on Wisconsin database. The limitation of this study is that this work is applied to three classifiers as MLP, GRNN, and SVM. The remaining part of this thesis is organized as follows:

Basic information about breast cancer and literature review is described in chapter 2. The classification and prediction methods are presented in chapter 3. The proposed algorithm and algorithm steps are described in chapter 4. The experimental result and quantitative comparison are presented in chapter 4, also the results are presented in chapter 5.

CHAPTER 2
LITERATURE REVIEW

2.1.Breast Cancer

The incidence of breast cancer varies from country to country around the world; in Hawaii, California, Canada with an annual incidence of 80-90 per hundred thousand, while, in the first place, the same value is only between 12-15 per hundred thousand in Japan. In European countries, the incidence is decreasing steadily southward and eastward from the western countries of the Nordic countries. This difference in disease incidence between countries of the world are seen in post-menopausal women, in particular, cross-country differences in the pre-menopausal period are very small. In the years after menopause mortality is becoming increasingly hard in Yugoslavia, which shows a rise while in the US, and Japan is steadily decreasing. The reason for this is still unknown[5].

However, the reported cases cannot reflect the full facts. The Ministry of Health's Cancer Registry of breast cancer, according to data from 1984-1990, are consistently ranked first in overall cancer incidence and cancer constitutes 24.3 - 35.5% of cases [5]. While breast cancer ranks first in Central Anatolia and the Black Sea regions, it ranks second in the regions of Marmara and Eastern Anatolia.Breast cancer incidence displays a rapid increase in Turkey. Breast cancer incidence had been previously calculated as 24.1/100,000 in 1993, and it is estimated that by 2010 the same rate raised to 50/100,000. These results show a two-fold increase in breast cancer incidence in Turkey over the last 20 years (9-12)[6].

- **Mortality**

Since the annual mortality rate is seen an increase, this increase was not as dramatic as in the incidence. Tracking Epidemiology and Results Program (US) states an increase in the frequency of 1980 breast cancer until 1989, the increase was detected in overall survival value, mortality rate shows that there are no changes. Besides the

increase in incidence, mortality has increased or decreased (despite the increase in effectiveness of treatment) emerged as the absence of conflict, in this case, is not yet disclosed [5].

Deaths are reported only from the provinces and districts. Therefore, death notifications are lacking. In 1996, the Ministry of Health has reported 3755 cases. Accordingly, face incidence is estimated to be 21.07 per thousand. However, according to the population of this value, it is considered to be below the expected numbers.

- **Risk Factors**

As in many types of cancers, breast cancer's underlying cause is still unknown but purportedly risk factors that increase the development of breast cancer genetics, reproductive, endocrine, way of life and can be grouped into environmental factors. Many risk factors for breast cancer clay of the relationship, according to the increase or decrease of the risk factors, the incidence also varies.

3% to 10% of breast cancers may be due to the BRCA1 or BRCA2, meaning these are related to genetic factors. For example, some patients who are under the age of 50 might likely to get the disease, especially if they have relatives who were diagnosed with breast cancer. If breast cancer runs in woman's family line, then regular checkup may be necessary in order to detect potential cancer at an early stage. In case a woman does have the mutation, then she needs to get a more rigorous screening in order to increase the chance of contracting cancer preventive (prophylactic) mastectomy. The decision to take the test should be discussed with a physician trained in counseling patients because genetic testing is a highly personal one.

By developing of breast cancer, it cannot be controlled by individual risk factors. Some of the risk factors are associated with external factors, however, there is no clear cause and effect relationship that exists. A woman who develops a breast cancer, there will be additional risk factors to consider. A woman treated with long-term hormone replacement therapy may avoid milk to prevent weight gain through exercise and

proper nutrition and one drink a day or less alcohol limit. Recent studies show that women who are at high risk of breast cancer have taken a drug called Tamoxifen for five years. The outcome demonstrated that the occurrence of cancer was reduced by about 50%. Tamoxifen is not a heavy treatment, yet some uncommon side effects such as hot flashes and vaginal discharge occur. Besides common side effects like blood clots, pulmonary embolism, stroke and cancer of the uterus are known to be experienced which can be life threatening. Tamoxifen is widely used to prevent, but may only be useful in some cases. Before it can be recommended for the prevention of breast cancer, some patients are also recommended to take vitamin D, which shows to be a way of treatment for a limited number of patients as the data suggests. Other ways of treatment of the disease are examining phytoestrogens (so large number of naturally occurring estrogens), vitamin E, vitamin C, and other drugs. Further testing of this material is also required before they can be recommended for the prevention of breast cancer. One of the best options to reduce the risk of death from breast cancer, specialists suggest regular screening mammography, breast self-exam and the need to learn to do a physical exam on a regular basis with your doctor.

Unfortunately, in the early stages, breast cancer may not have any symptoms. That's why screening is important to seek advice. The tumor grows in a variety of symptoms, including the production of

- Swelling or thickening of the breast or armpit,
- Changes in breast size and shape,
- Discharge from the nipple or nipple turning inward,
- Redness or scaling of the nipple or skin,
- Ridges or pitting of the breast skin.

We are using breast cancer mammography, clinical breast examinations, and tests to the screen of breasts. Screening mammograms are simply x-rays of each breast. X-rays are taken and a nozzle disposed between the two plates for a few seconds. If anything looks abnormal vision or better, it needs a large angle of view during the mammogram film. Mammograms often detect tumors before they felt they could be early signs of cancer and can detect small spots of calcium. Regular screening

mammography can reduce mortality in breast cancer up to 30 percent. The majority of breast cancer mammography is associated with abnormal findings. In mammography for a woman start from the age of 40 to take risks and have a strong family history or a genetic mutation that women who may want to have an early start to increase.

A woman aged between 20 and 39 must have a physical exam every three years. Furthermore, after age 40 every woman should perform a physical exam every year. Clinical breast exam medical examination for lumps and breast size and shape professionally made sense to look at change. About 15% of tumors can be felt but not seen with mammography screening on a regular basis. One can learn to do a breast self-exam during a physical exam. About a week after the end of each woman should perform breast self-examination once a month. If anything is found or any changes occur in the breast, one should contact their doctor.

Currently, some experimental screening methods are studied. These are magnetic resonance imaging (MRI), ductal lavage, ultrasound, optical tomography, a PET scan and digital mammography. Magnetic resonance imaging is an important matter in the clinical treatment of many medical patients. The medical information detailed magnetic resonance, computed tomography (X-ray CT), other imaging modalities such as mammography since EITC is available more commonly. There are two types of nuclear relaxation time (T1 and T2), and the intensity of the movement. Figure 1.2 shows a sample of MRI imaging system for breast cancer.

Figure 1. 2 MRI imaging system for breast cancer[7]

Breast cancer based on the currently available information on development ductal epithelium, atypical ductal hyperplasia does turn to pass through stages of breast cancer such as ductal carcinoma in situ end. This transformation lasts for decades. Bass milk duct system transmits the baseline (duct) passes in limited its basal membranes of cancer cells into connective tissue scroll later. At this stage, the tumor cells are capable of damaging what they encounter with blood vessels and the lymphatic metastasis. It is estimated that it takes about 8 years for a gram of cancer to occur in the breast.

2.2.Type of Breast Cancers

There are three types of breast cancer:

A. Benign: if the increase in the signal intensity is monotonic over the period after the injection.
B. Suspicious: if the peak signal intensity is observed before three minutes and preserved for the rest of test.
C. Malignant: if there's an immediate decrease in the signal intensity right after the peak is reached [8].

In figure 2.1 these three types are shown.

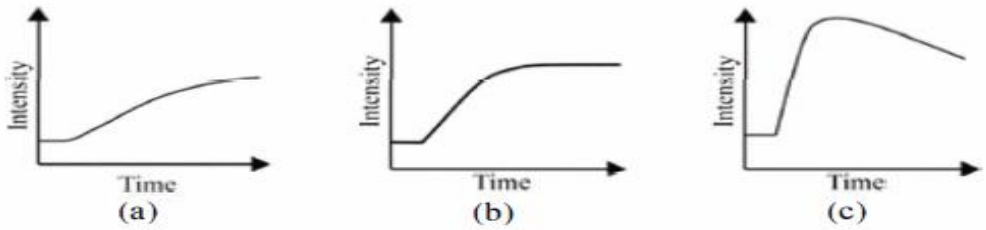


Figure 2. 1 Three types of breast cancer[8]

2.3.Magnetic Resonance Imaging (MRI)

MRI analysis is based on the creation of body images displayed as thin horizontal portions of the breast tissues that can be studied from several angles by using magnetic and radio waves where ionizing radiation is avoided. During the MRI analysis of each breast, a great number of images are acquired to be assessed by a radiologist. Similar

to mammography, which uses X-ray machines equipped for specifically imaging breasts, MRI uses special equipment with higher qualities than the images obtained from the equipment used for MRI scanning of head or chest to create breast images. In the researches related to breast cancer detection, several techniques based on MRI have been developed. A novel and auspicious technique introduced is dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI). It is expected to be particularly suitable for screening dense breasts. DCE-MRI processes involve a contrast agent, generally Gadolinium DTPA, which is useful in the improvement of tissue discrimination. The reason for using this paramagnetic compound is its existence at intravascular and extracellular fluid space and its ability to increase the luminance of these Gadolinium-enhanced tissues remarkably, because of which vascular tissues such as lesions can be detected with ease. Within a typical DCEMRI process, the breast is imaged both before and after the injection of a contrast agent to observe the effects of the contrast agent on the tissues[9].

In a recent study, a computer-aided diagnosis (CAD) system, which has enhanced the sensitivity, precision, and rapidity of MRI studies, has been developed for detecting tumors. Besides the images taken before the injection, at every minute after the injection MR images are taken for six minutes with the T1 mode, the FLASH 3D traverse setting of the MRI machine. The principal task carried out by the contrast agent during this time duration is changing the relaxation times of tissues as a result of which the radiologists can determine the existence and site of a tumor or a lesion.

2.4.Breast Cancer Detection

Another research related to MRI in breast cancer detection introduced a temporal feature analysis. In accordance with the assertions mentioned previously, malignant lesions are distinguished with their rapid contrast enhancement followed by fast washout while benign lesions are identified by their monotonic contrast enhancement. Thus, it has been claimed that the automatic detection of the potential lesions on MRIs can be determined by the examination of the kinetic curves. In this method, firstly, suspicious areas are detected depending on the contrast enhancement properties of the

tissues. Then an angiogenesis map is formed after analyzing the kinetic curves. After that, pixels are displayed in different colors in accordance with the categories they belong [10].

Besides these MRI-based techniques used in the detection of the breast cancer, MRI data also serves as an efficient tool in the development of realistic numerical breast models by allowing the exhibition of the heterogeneous and dispersive nature of the breast tissues. The experiments made to determine the dielectric properties of the breast tissues have helped scientists to conceive the varying nature of dielectric values across the breast. Nevertheless, the complexity of the dispersion of the tissues throughout the breast still poses a difficult problem.

Isa et al. presented a mass detection and segmentation system in mammogram images. The system in the beginning separates the breast from the background then improves the contrast of tissues. The basis of the work is applying constraint region growing based on local statistical texture analysis in order to detect the mass and segment it out. The results showed that the proposed technique has 94.59% of sensitivity and the number of false positive is 3.90 per image[11]. Guruprasad et al. [12] proposed a system for estimating the percentage of cells from the cell characterization which image has high accuracy (96.4286%) carcinoma (cancer) automatic segmentation.

2.5.Classification using ANN and SVM

System classifier tools by using Artificial Neural Network and statistical approaches have been developed for the analysis of pathological images. As do the human visual system, this system performs segmentation and classification [13-15], to recognize objects, which; perceived depth; different tissues or curved surfaces defines a surface tilt with texture information and luminance. In [12], an attempt of soft tissue characterization Artificial Neural Network (ANN) using tissue-primitive features and segmentation tools were made to provide the classifier approach. This is only half of the Artificial Neural Network control structure, which is a supervised approach that

identifies where the ground level; then the algorithm itself will scan the entire image, and performs segmentation and classification in an unsupervised mode.

After training, microcalcifications detection is formulated as one the classification problem, then SVM is applied to improve the detection algorithm. SVM is a promising method for abnormality detection in medical imaging applications. It is used for detecting microcalcifications and their locations in the mammogram image [16].

In [17], a complex Wavelet decomposition method is implemented. Results of the actual use of the wavelet transform of the wavelet complexes give significantly better results than the ones obtained. Even though, they develop quite satisfactory classification results, they only used the size and shape distributions such as nuclear morphometric features.

Rivera et al. proposed a "content-aware" algorithm that improves the dark images, detects details in textured areas, makes the edges sharp and protects the softness of flat areas. This method produces an 'ad hoc' transformation for each image, adjusting the mapping functions to each image features to generate the enhancement, then has been analyzed the contrast of the image's border and regions and gathered all the information with common properties. The functions were extracted from these groups model relation within the image. The results were then adaptively mixed to boost the details in the image. The disadvantage of this method is that it is incapable to improve the darkness into the image area while "image fusion" techniques overcome this limitation and enhances all regions of an image such as mammogram images [18].The first three signals to multiple groups provide information on different frequencies and then following a two-step process:

1. Using a new method to estimate the noise to increase the signal-to-noise ratio of the local.
2. Modeling and then rebuilding the combined image by inverse transform [19].

The performance of the ROC was used to assess the benefits of a contrast. "Dynamic spline wavelets" were also used[20].In [21], four different clustering algorithms for segmentation of nuclei were proposed. Methods of medical decision support system were applied and the actual routine medical data received from a hospital has been tested in terms of classification accuracy. Approximately, the accuracy was 96-100% and the approach was found to be rapid and relatively easy to implement.

2.6.Breast Cancer Prediction using Computer Aided Detection

Five image enhancement algorithms were used in order to understand the impact of the increase in the work and adjust the image processing parameter with a computer-aided detection system for detection of microcalcifications on a mammogram evaluated.

Computer-aided diagnosis (CAD) optimized more pixels to the detected object contrast and the minimum size that can satisfactorily show calcification of the micro-targeted. LRM and wavelet-based linear stretch in which LRM achieved the highest rate of performance (A (Z) = 0.932) and wavelet-based linear elasticity had the rate of performance (A (Z) = 0.926) [22].

Jha et al. suggested a nonlinear non-dynamic stochastic resonance-based technique to enhance the low contrast and dark images. They treated a low contrast as a sub-threshold signal, then they used a noise-enhanced signal processing to improve the image contrast. The proposed technique has applied uniquely with outer noise to equalize the effect of inner noise of a dark image due to insufficient lighting randomly and repeatedly. The noise was added to the image. Changing the noise intensity, noise-induced resonance was obtained at particular optimum noise intensity. The proposed technique can be applied for four distribution types: Gaussian, uniform, poisson, and gamma. Quantitative evaluation of their performances had been done whence of contrast enhancement factor, color enhancement, and perceptual quality measure [23].

Garg et al. proposed a method based on the main idea of the histogram equalization enhancement techniques. They compared and analyzed the performance of the techniques contrast limited adaptive histogram equalization (CLAHE), equal area dualistic sub-image histogram equalization (DSIHE), dynamic histogram equalization (DHE) in terms of peak signal to noise ratio (PSNR), absolute mean brightness error (AMBE) and contrast. The result of this work proved that DSIHE technique in terms of AMBE gave the best results among the rest of the techniques that were used [24].

In [25] they used square support vector machine and minimum redundancy maximum relevance for diagnosis of breast cancer from breast microscopic images. The aim was to determine the microscopic light received cancerous lesions. Here, a total of 180 breasts namely 3x60 are sent as microscopic images taken in Pathology Laboratory University School of Medicine. Some features are quite diverse turning 92 (23x4) of the characteristic of breast microscopic imaging angles.

Zhiming et al. implemented FAHE by adopting three techniques to enhance the quickness of AHE through, firstly acquiring the local histogram by a sliding window with a reiterate approach. Secondly, computing a half of the histogram cumulatively; last, keeping the block size (W2 equal to the product of integral power of two and gray level number) and all the operation of multiplication and division has been changed with the fast bitwise shift [26].

Pizer et al. proved by their survey study that the advanced techniques which enhance digital images' contrast such as Global Contrast Enhancement(GCE), ACH and CLAHE have shown success to grant marks point to the significance of an exact visual model in the techniques' evolution [27].As a result of the foregoing, the conclusion that can be drawn is that although histogram equalization is easy to use, fast and some of its techniques like AHE has proved its efficiency in the images, [16, 28], it has some of the disadvantages that may impede getting the desired result in some cases: if the histogram covers almost the full gray scale, if there are gray values that physically far apart from each other in the image.

In [29] they apply a texture feature in breast cancer microscopic images of the CEA method. For this purpose, a method well-known sequence for 1D horizontal CEA (horizontal and vertical) has been extended to 2D image data using the sequence of images along different directions. Results of CEA images calculated by the method of histological texture features microscopic cell counting accuracy indicate that breast cancer is capable of classification.

Wu et al. proposed a new method using compound "MRFs" based on a general boundary model. The main goal of this method is using the concept of the relationship between labeling and boundaries to enhance the performance of segmentation. The experiments and comparisons of proposed model with other existing MRF methods have proved that the proposed model showed good and satisfactory segmentation results in high noise regions [30].

Dominguez et al. proposed a method with the following scheme:
(1) Preprocessing the original image in order to increase the signal-to-noise ratio to detect the abnormality,
(2) Segmenting the prospective abnormal region, and (3) Lessening the false-positive rate.

Region segmentation is applied after using morphological operations, local scaling and wavelet decomposition and reconstruction via transforming the image multiple threshold levels to a binary image. A set of features is computed from each of the segmented areas.

In this work, 57 mammogram images were tested, contained breast masses from MIAS data set, including speculated, circumscribed, and ill-defined masses. The result of this method reached 80% sensitivity rate at 2.3 as a false-positives per image[31].

Maitra et al. use "binary homogeneous improvement algorithms (BHEA), followed by edge detection algorithm (EDA), breast border detection algorithm (BBDA), muscle detection algorithm (PMDA) and anatomical segmentation of breast ROI (ASB) were

proposed using the latest region growing algorithm (SRGA), which is a modified algorithm and a growing method for each division area (coloring objects boundaries) on a mammogram. In other words, every single pixel area bounded by a line as edge detection process followed by the anatomical segmentation is obtained. The precision is around 99.87% [32].

2.7.Breast Cancer Prediction using Generalized Regression Neural Network

Breast cancer prediction can be classified into different types if they are combined to form a group to identify a particular class [33]. Several strategies that can be used to achieve different categories such as sum, product, min, max, average and majority voting [34]. The use of GRNN is that it is simple and requires no training, although this is offset by the fact that they do not adaptive. Stacked generalization, however, is a way to combine classifiers for fusion-decision on the need for further training [35]. This combination of classification, however, requires careful selection of basic classification to achieve an acceptable rate of false classification[36]. Support Vector Machines (SVM) classifier provides very promising results in a variety of biological activities, including tiny gene expression[32].It has been shown that generalized regression neural network (GRNN) is better than SVM in classifying genetic data[37].

2.8.Statistical Methods in Cancer Prediction

Conventional statistical methods used in breast cancer prediction are regression methods and DA. These methods are mainly employed for classification of patients into different outcome groups using a set of discriminatory factors. A history of these techniques, their principles, advantages, and disadvantages are detailed below.

2.8.1. Regression Models

Regression is the most commonly used the statistical method in medical studies. It is employed to model and assess the relationship between a single and a set of input and

output variables. In regression analysis, inputs are described as predictors or independent variables, while the outputs are termed as predicted outcome or dependent variables.

There are different regression models depending on the type and number of input and output variables and their relationship. Linear regression, models the input-output relation with a straight line, while, in a curvilinear regression model, any function forming a curved line can be used for input-output modeling. Multivariate regression analysis can be employed in the case of developing a model for multiple inputs with a single output.

After developing the regression model for a set of input and output variables, the model is assessed to study how well the model fits the data and whether it requires any modifications. In addition to predictive modeling, regression analysis can be utilized as an exploratory tool to understand the nature of the relationship between multiple factors. This helps to verify any existing relationship between variables and to clarify the description of this relationship.

Linear regression is the simplest form of regression analysis. As mentioned before, it describes the linear relation between a dependent and independent variables. Prior to linear regression, it is common to use another measure of linear relation, called correlation coefficient, to measure the linear dependence of two variables. Correlation coefficient does not consider any causal relationship between the variables and it merely quantifies the strength of the linear relation between them.

In the following sections, the correlation coefficient is explained followed by linear regression analysis. Linear regression is then explained in detail including its model building, underlying assumptions and the interpretation of the fitted model.

2.8.2. Correlation

The correlation coefficient is a measure of linear dependence between two variables. It is commonly employed prior to linear regression to explore the existence of any

linear relationship between the dependent and independent variables. The most commonly used method for quantifying the amount of linear dependency between two continuous variables, *A* and *B*, is Pearson's Correlation Coefficient (PCC). PCC represents the strength of linear association between two variables *A* and *B* by normalizing their covariance with respect to their standard deviation *SA* and *SB* as:

$$PCC_{A,B} = \frac{cov(A,B)}{S_A S_B} = \frac{E\langle (A-\mu_A)(B-\mu_B)\rangle}{S_A S_B}$$

where μA and μB the expected value of random variables A and B PCC to measure the linear dependence between variables assigned a number between -1 and 1. A positive value indicates a positive linear relationship while a negative line and 0 indicates negative implications and no relationship between the variables.

The main characteristic of the PCC is its invariance to a linear transformation of the variables *A* and *B*. This implies that for example, with variables being described as A = 5B+3, PCC between the two variables represent a perfect correlation of *1*.

2.8.3. Linear Regression

Linear regression describes the relationship between an independent variable X and a dependent variable T given a set of data points (xi,ti). This relationship is expressed as a straight line of the form:

$$t_i = \beta_0 + \beta_1 x_i + e_i$$

where β0 and βi are the regression coefficients describing the intersect and the line slope respectively. The term e_i is the error defined as the difference between the data point (xi,ti) and the regression line. Linear regression analysis is performed to estimate the unknown coefficients β0 and β1. The least squares method is commonly used for estimating coefficients to fit a regression line to the data.

2.8.4. Logistic Regression

To overcome the limitation of linear regression in handling dichotomous classification problems, logistic regression was proposed in the early 1970s and become widely available to researchers through statistical software in early 1980s[38].

In medical diagnosis, where the disease outcome is described as present or absent, the outcome is a binary variable taking values of 1 or 0 similar to the logistic regression classification outcome. Hence, logistic regression has been routinely applied and became a standard method in the field of medical diagnosis as a technique, which is simple to apply through several statistical packages. In breast cancer, logistic regression has been frequently used to predict dichotomous outcomes of patient survival, diagnosis and prognosis [38, 39].

As a linear multivariate analysis technique, logistic regression describes the relation between one or more independent variables and a binary outcome. Logistic regression employs the same principles as linear regression for data analysis. Like linear regression, a model is developed to fit the input-output data by estimating a set of coefficients. However, because the output can take any value from a continuous range of outcomes in linear regression, it is difficult to assign a dichotomous output to (2. 2). For this reason, logistic distribution is used for describing the input-output relationship that is defined as the conditional mean of the output T given the input x [40]:

$$\pi(x) = E(T|x) = \frac{e^{\beta_0+\beta_1 x}}{1 + e^{\beta_0+\beta_1 x}}$$

2.9.Discriminant Analysis

The purpose of DA is to predict the class membership of a data point using a set of predictive variables. The simplest form of DA, linear discriminant analysis (LDA), was developed in 1936 by Fisher for data analysis and classification [41]. Similar to LR, LDA is suitable to be applied for linear classification problems with dichotomous

outputs. For this reason, it can be a suitable method in medical diagnosis for identifying differences between patients with and without ALN metastasis and classifying them into one of these groups.

LDA classifies data through maximizing data separability by projecting it into a new space where groups are maximally separated. Fisher LDA finds the transformation vector is such that the distances between the means of transformed inputs are maximized while the variances of the projected data in each group are minimized. The main limitation of LDA in medical applications is that it only deals with continuous input variables and makes more assumptions about the input data compared to other statistical methods such as LR. LDA assumptions about independent variables are multivariate normality and equal covariance. If these conditions are satisfied, LDA gives better or similar results compared to LR [42]. However, medical datasets are unlikely to lend themselves to these properties. Furthermore, it has been proved that LR performs better than or similar to LDA in most applications [42]. Therefore, the application of LDA in medical studies is limited.

Some variations of DA have been proposed to obviate the limitations of LDA such as nonparametric LDA. In this method, the limitation of Gaussian assumption is removed by employing the k-nearest neighbors rule for computing the between-class scatter matrix[43]. However, this method still has limitations when the between-class scatter matrix is singular.

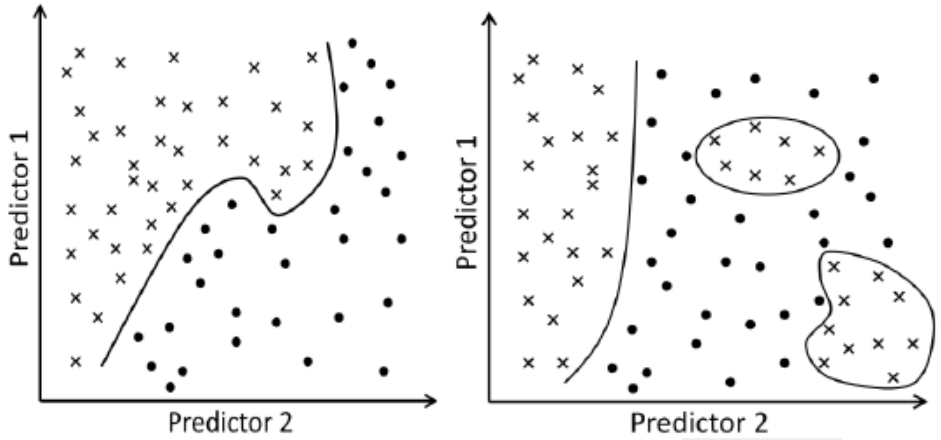
2.10. Problems with Statistical Methods

Using more than one predictor is common in cancer studies as it is unlikely that one factor can provide discrimination between patients with and without a specific outcome. For this reason, multiple predictors are employed to classify the cancer outcome which gives rise to a multidimensional input space. The relationship between multiple predictors is typically a nonlinear and complex one, which renders the linear statistical methods insufficient for these classification problems. In addition, many

statistical methods make prior assumptions about the data such as multidimensional normality that is not generally true in cancer studies.

An example of a nonlinear classification problem with two predictors (i.e. 2-dimensional input space) and a binary output variable is shown in Figure 2.2 (a) and 2.1 (b). In these figures, problems with nonlinear decision boundary and closed curve boundary are presented [44]. In both cases, a linear classifier such as LR or LDA would fail to separate the data. In order to employ linear statistical methods for problems with nonlinear decision boundaries, data can be transformed nonlinearly into a new space where it is more separable and methods such as LR or LDA can be used. Then again, there is no guideline on how to choose an optimal nonlinear transformation and hence, it is unlikely to find a transformation that projects data into a space where it is completely separable.

ANNs can solve this problem by automatically determining the data transformation and finding an optimal decision boundary. MLP is a specific ANN architecture which resembles a form of nonlinear DA that can overcome LDA limitations by transforming the inputs in a nonlinear fashion. Like DA methods, ANNs utilize training data with the known outcome to build a model with which a new datum can be classified. However, they make no or little assumptions about the data and are able to learn nonlinear patterns from training data [45]. For this reason, they are chosen as a desirable classification technique in many real-world problems [46, 47]. In addition, ANNs yield higher accuracy in cancer diagnosis and prognosis compared to conventional statistical methods such as LR and Bayesian classifier [47-49]. ANNs are the main platform used in this study and are fully discussed in the later sections of this chapter.



(a) (b)

Figure 2. 2 Nonlinear classification problems with two predictors and a binary output with (a) nonlinear decision boundary and (b) closed curve boundary [45]

2.11. Machine Learning Techniques in Cancer Diagnosis

Machine learning is a branch of artificial intelligence that embraces a wide range of methods including DT, rule induction methods, Bayesian learning, genetic algorithms and ANNs [50]. These methods employ various statistical and optimization techniques to provide decision making by learning complex patterns in a population from a sample set of data. Because of this characteristic, machine-learning techniques have been widely applied in the field of medical diagnosis and prognosis.

Machine learning methods generally achieve 15-20% improvement in accuracy for predicting cancer survival and recurrence besides providing more insight into the nature of cancer progression. Among various machine learning techniques, DTs are among the most widely used methods of cancer diagnosis [51]. DTs are successfully applied to ovarian and gastric cancers [52, 53]. However, DTs are usually used in conjunction with another classification method to achieve better classification results [54, 55]. A brief history of DT in the medical literature, its structure and learning algorithm, in addition to its advantages and disadvantages are explained in the following section.

2.12. Decision Tree (DT)

DT learning is an inductive inference algorithm that can approximate discrete outcomes using a set of training data. The learned function by DT is represented by a set of branches and nodes that are structured as a graph. The graph starts from a root node that is attributed to one of the input variables followed by leaf nodes attributed to each of the remaining input variables. The outcome of the tree is a category into which the input is classified. An example of the structure of a DT for a simple classification problem is presented in Figure 2.3 [50]. In this example, the aim is to decide whether to play tennis or not. The decision depends on the weather forecast and is if:

Outlook = Overcast
OR Outlook = Sunny AND Humidity = High
OR Outlook = Rain AND Wind = Weak
The output would be 1 otherwise *0*.

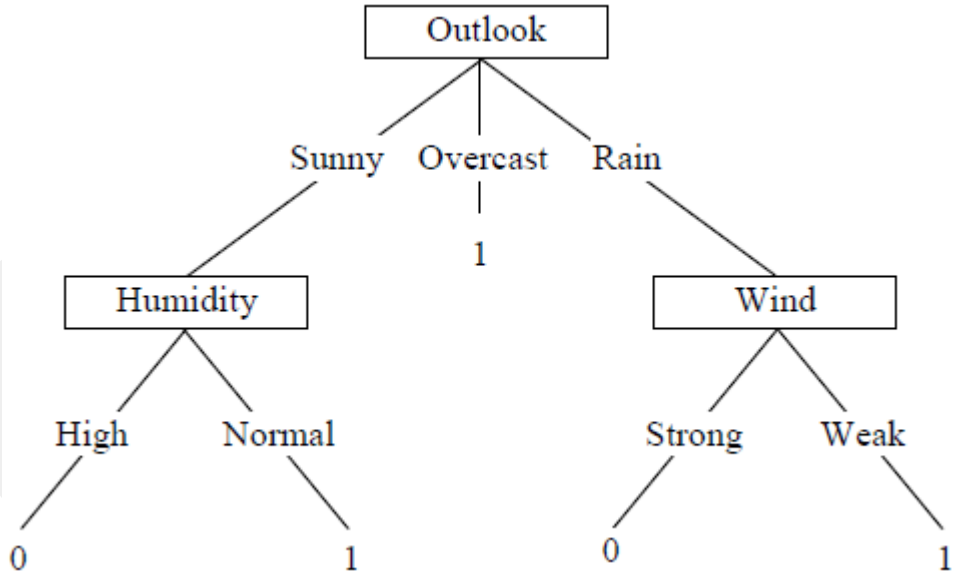


Figure 2. 3 A decision tree structure to decide whether to play tennis or not based on weather forecast[50].

The simplest form of learning in DT consists of a greedy search that starts from the root node and only goes down without backtracking to the previous nodes [50]. The root node is selected using a predefined measure. The training data are then assigned

to the succeeding leaf nodes based on their values of the root node variable. The leaf nodes' variables are then selected using the same measure and a similar process is repeated until a decision is reached or no further division is possible. In ID3 decision trees, the measure for choosing the best attribute in each step is information gain [56]. This measure quantifies how well each variable classifies training data into output categories. Information gain is defined using a well-known measure in information theory called *entropy*. The entropy of the dataset X with respect to a two-category output class is defined as:

$$Entropy(X) = \sum_{c=1,2} -P_c log_2 P_c$$

Where Pc is the proportion of the data points in X belonging to the class. Entropy can take values between0 to 1 depending on the proportion of each class in the data. If all the data belongs only to one category, the entropy is equal to 0 while it is 1 when there are equal numbers of each category. An unequal number of output categories result in an entropy value between 0to 1. The information gain of a variable is then defined as:

$$IG(X,V) = Entropy(X) - \sum_{v \in Values(V)} \frac{|X_v|}{|X|} Entropy(X_v)$$

where V is the input variable and V values (v) are its possible values. Xv is the subset of X for which variable V has values v. DTs maintain several advantages that have made them attractive in medical diagnosis. The results provided by DTs are easy to interpret in terms of if-then rules. In addition, DTs are considered as robust classifiers with quick learning [50]. However, DTs do not perform as well as ANNs in complex classification problems [57].

One limitation in DTs is that each attribute is required to have a small number of discrete values. To address this problem, continuous variables are partitioned into discrete intervals. A new dummy variable is then introduced that takes discrete values

attributed to each of the intervals. The main disadvantage of implementing continuous inputs in DT is the loss of information that occurs due to partitioning.

Another problem that arises with partitioning the continuous variables is selecting the appropriate intervals. Furthermore, DTs are highly prone to overfitting. DTs may present good predictions for small training data with missing values. In effect, a large tree may separate the training data perfectly with no misclassification error. However, such model may over fit the data by having few training samples at the end of each branch which means low generalization ability for new data.

2.13. Support Vector Machines

Support vector machines (SVM) are another machine learning technique which can be classified as Kernel-based methods [56]. SVMs are supervised learning methods based on the statistical learning theory and the Vapnik-Chervonenkis dimension [57]. For a linear classification problem with dichotomous outcomes, SVM finds the decision boundary in the form of a hyperplane such that it is as far as possible from the closest members of both classes. In nonlinear SVMs, the data are recast into a higher dimensional space by using a nonlinear kernel function such that they are linearly separable in the new space. It has been shown that SVM performs as good as ANNs such as the PNN in breast cancer detection [58]. However, using a nonlinear SVM means that an appropriate kernel function and its parameters should be chosen such that the nonlinearly separable data can be mapped into a higher dimensional space where it is linearly separable. Unfortunately, there are no specific rules for choosing such function and it is commonly chosen using trial and error from a set of predefined kernel functions. This makes the success of SVMs dependent on finding an appropriate transfer function, which might not be possible for highly nonlinear and complex data.

2.14. An overview of Artificial Neural Networks

ANNs are a branch of artificial intelligence that are able to learn complicated nonlinear patterns from a set of data. ANNs are parallel computational units, which have become

a valuable classification tool in recent years. They first originated from the idea of simulating human brain abilities in decision making and parallel processing by combining mathematical modeling and engineering design. While the basic structure and characteristics of ANNs still resemble the human brain, their functioning and the way they make decisions have become far different from biological neural networks over time.

Different types of ANNs can be categorized by two main criteria[59]. The first criterion is how the network is encoded, i.e. how the network stores knowledge from the data. Using this measure, ANNs are categorized as supervised and unsupervised. The second criterion is the way the networks are decoded, i.e. the way the network processes new data once it has acquired knowledge from the old data. This criterion classifies ANNs into feed forward and feedback. Figure 2.4 illustrates this classification.

In a supervised network, both the input and output are presented to the network. The network weights are then adjusted by computing an error from comparing the network output and the desired output. The optimum weights are obtained by optimizing the error function. Afterward, presenting new inputs to the network and comparing the answer with the unseen output test the ability of the network in classifying new data. Some ANNs have the ability to learn without teachers. The learning process when only inputs are presented to the network is termed as unsupervised learning. This is achieved by using rules for self-adjustment as the new inputs are presented to the network. The self-adjustment is performed based on predefined rules [60].

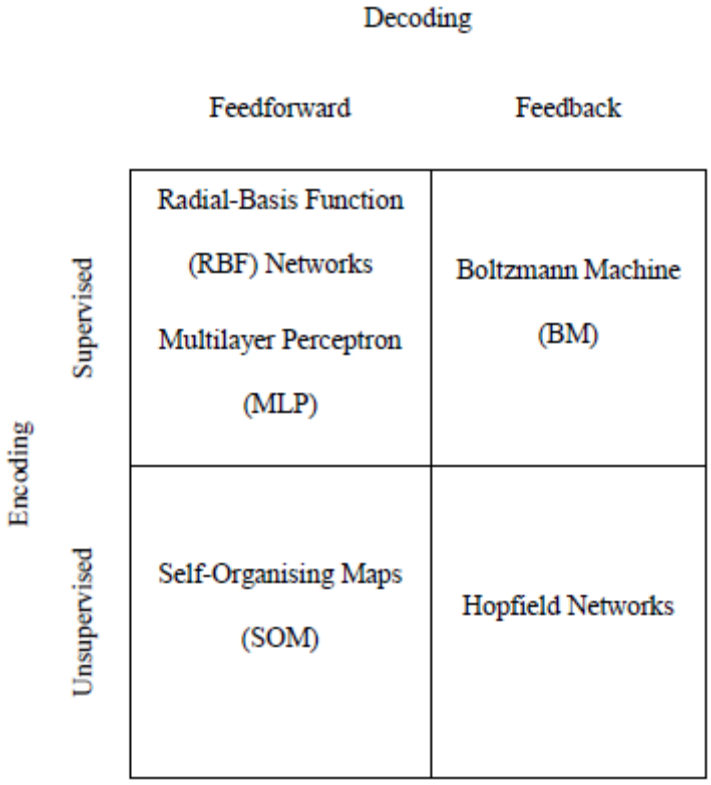


Figure 2. 4 Different types of ANNs can be categorized by two main criteria: how they are encoded, (supervised and unsupervised) and how they are decoded (feedforward and feedback)[60].

Excessive and continuously growing literature is available on ANNs [47]. The theory of ANNs is derived from different disciplines including mathematics, statistics, biology, engineering, computer science and neuroscience. Each of these disciplines contributes to the capability of ANNs as intelligent systems to be employed in a wide range of applications. A comprehensive explanation of the mentioned disciplines forming the foundations of ANNs is out of the scope of this thesis. Therefore, the focus is only on the paradigms that contribute to the ANNs applied in the field of medical diagnosis and more specifically, to breast cancer prediction. In the following sections, the fundamentals of the ANNs are briefly explained to prepare the grounds for further explicating the ANN models used in this project.

2.15. Neurons

As mentioned above, the structure and function of ANNs resemble the human brain, in that they include a massively parallel architecture, but in a relatively small scale. Whilst human brain is made up of 100 billion neurons in average, ANN is based on layers of computing nodes which include only a few computing units.

Biological NNs consist of controlling units called neurons, which have the ability to learn and work in parallel. ANNs emerged in 1943 for the first time by representing a simple form of biological neurons by elements that could perform computation[58]. Biological neurons have different specialization and functioning. A simplified structure of a biological neuron consists of four main parts. Each neuron has a cell body called soma, which can receive input from nearly 10000 other units through an assembly of subtle structures called dendrites. After processing the received signal, the neuron sends out an electrical potential through a long, thin structure called axon, which divides into many branches. Afterward, the electrical output is transferred to the dendrites of other neurons via junctions called synapse.

2.16. Artificial Neural Network Definition and General Features

Computers have become an indispensable part of the modern world and computer systems today can decide on both events, to learn about the relationship between both events. Computers equipped with the feature of solving problems that cannot be expressed mathematically and cannot be solved by using methods based on experience, and lead the development of "Artificial Intelligence". Scientists have been forced to work on the brain which has outstanding features while working on artificial intelligence and inspired by the structure of the brain tried to extraction the mathematical model. In order to model the brain with the idea that all behavior should be modeled accurately the physical components of a variety of artificial cell and network models have been developed. Thus, today's computers as a separate discipline of algorithmic calculation method “neural network” have emerged. In the most general sense of the ANN, mimic many nerve cells in the human brain that are the result of simple processors connected together in different levels of impact regarded as a complex system. Computers are very successful in their mathematical and algorithmic

calculations. Neural networks can be used in non-linear equations and generalization capabilities due to a very complex and large-scale solution to problems that are easily produced. Due to the fact that they are adaptable, fast and easy in analyzing and designing due to their ability to learn made them one of the indispensable elements of our age.

Machines can learn using neural networks as they can work with minimal mistakes unlike the mistakes done by their human counterparts. Since it is an imitation of biological neural networks, an understanding of the structure of biological neural networks will facilitate the understanding of the neural network. There are approximately 10 billion neurons in the human brain and the nerve cells that are estimated to make up 60 trillion connections with each other. Nerve messages from the senses are transmitted to the next cell evaluated by nerve cells. Thus, the message carried by the signal is transmitted to the central and nervous system. Nerve cells generate response signals to evaluate the central nervous system signals. These signals are transmitted in response by the organ to the formation of the nervous system. Thus, the signals from the sensory organs response will be transmitted to the body through the nervous system.

At this point, the general structure of the ANN will be useful to sort out before entering into details about the features that make it important. The most basic capabilities of ANN are in this context[45].

2.16.1. Non-linear Function of the Network

The neuron itself is not linear an essential component ANN. Accordingly, the neural network formed by the interconnection of neurons is not linear and also the non-linear characteristic is due to the nature of ANN is distributed across the network in parallel. In the absence of the desired linear mappings due to non-linear sub-units on the network structure in the fulfillment of the function is mathematically possible. Therefore, the function of structural flexibility allows taking place correctly.

2.16.2. Input-Output Association

ANN guided especially be mentioned later in the study has a structure that allows learning to take place. So that using the elements in the ANN training set to give the desired results for the specified entry revises free elements in its structure. To make fewer errors, each iteration resets weight and connections at each stage.

2.16.3. Becoming Acceptable Easy Adapter

When the threshold value is changed conditions in the ANN or by simple operations such as adjusting the weight applied to the input can be adapted to the new environment.

2.16.4. The Parallelism and Function of Structural Disarray of the System

Neurons within the network structures actually working simultaneously constitute a complex function. Working simultaneously in multiple neurons does not directly affect the success of the network in the event of outages of any neurons. Depending on the application in order to get the proper output connections between neurons can be canceled automatically. This is indicative of the fact that in parallel and distributed structure of all the neurons in achieving the expected results of the ANN.

2.16.5. Generalisation Ability

The generalisation ability of a network during training determines the mapping of numerical data used. The network can produce meaningful answers for unused inputs during training if the generalisation ability of the network is high.

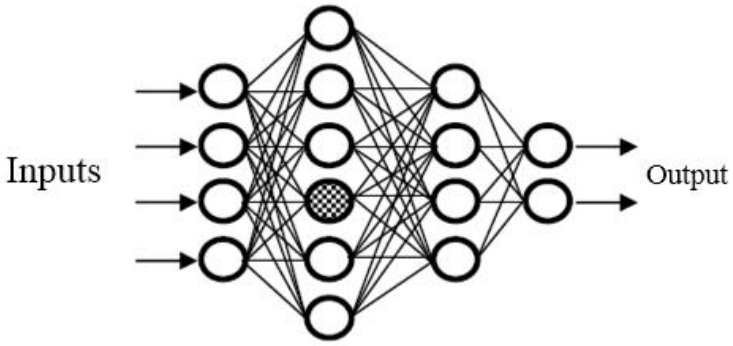


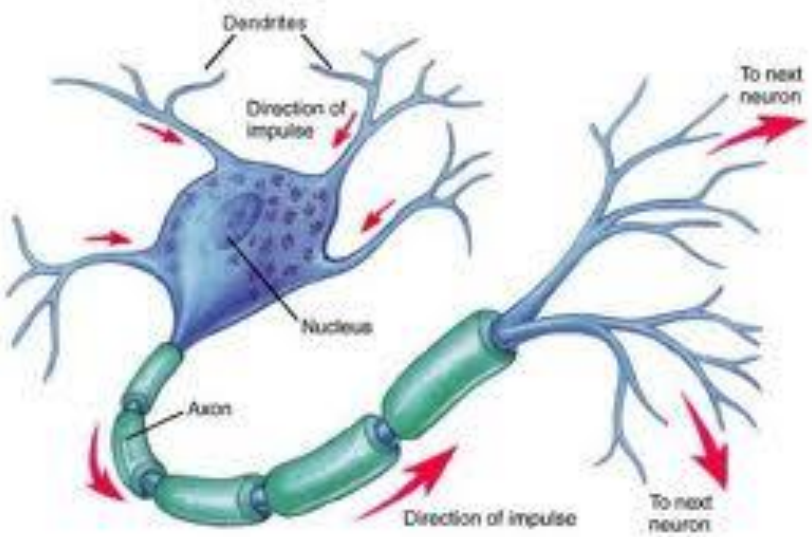
Figure 2. 5 Neuron structure of the network in the event of breakdown [60].

2.16.6. Very Large Scale Application Compliance

In particular, the number of neurons in the hidden layer of the multilayer structure can be increased. The structure of the ANN can be very complex. For this reason, it can be used to solve nonlinear and complex problems.

2.17. General Structure of Neural Network

ANN has four main parts formed of dendrites, axons, the core and connections as shown in Figure 2.6 [61].



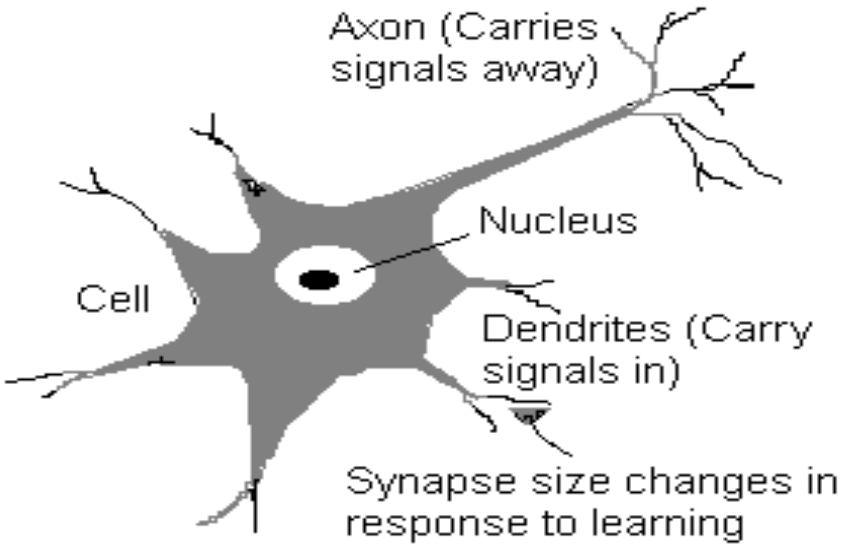


Figure 2. 6 The sections making up the neural network [60].

Dendrites: Dendrites that convey the core of the electrochemical excitation received from other neurons are root-like structures. Electrical stimulation of neurons and dendrites in consecutive sequence is through the synapses between them.

Core: Core is the central processing unit core of the nerve cell, which collects and transmits signals from the dendrites to the axon.

Axon: The axon branches out from the body, which is the cytoplasmic fraction. It has an important role in message transmission.

Connection: It is responsible for transmission signals to other neurons. The amplification of the signal must be at or above the threshold in order to send a signal. A signal below the threshold value below is not transmitted to other neurons.

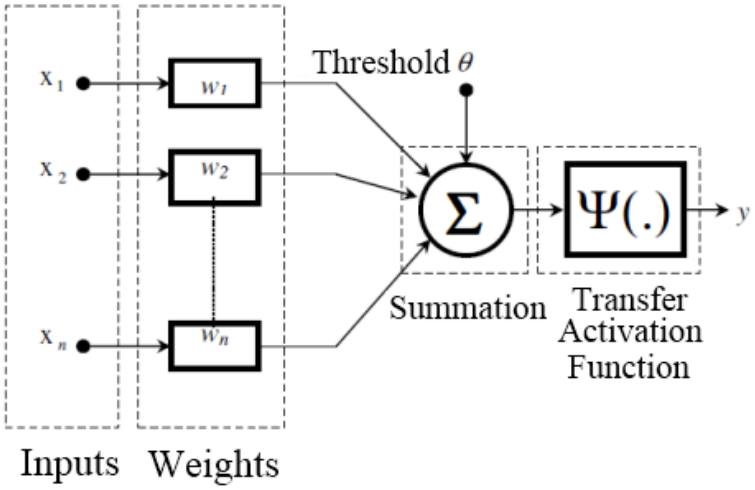


Figure 2. 7 The structure of an artificial neuron [60]

The neuron model shown in simplified form in Figure 2.7 can also be considered as a threshold volume. As a threshold neuron unit receives signals from the synapse, it collects all the signals that are generated by multiplying the appropriate weight. The output depends on whether the result of the sum function value is above or below the threshold activation. The ANN block diagram is shown in Figure 2.8.

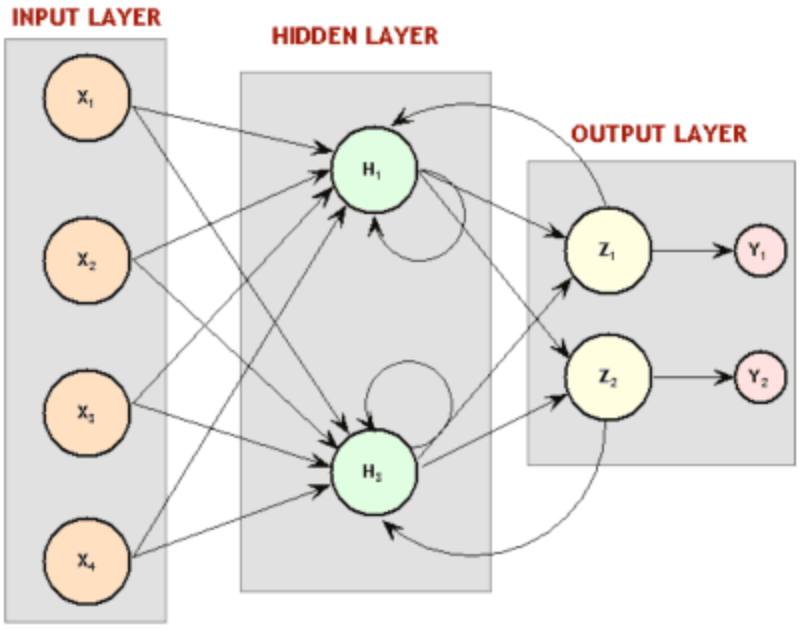


Figure 2. 8 ANN general block diagram [60]

To set the threshold value of the signal entering the activation function at this point bias (θ) is called a fixed input value which is applied to the input neurons. This allows the threshold to be changed in order to receive the expected outputs. ANN is a basic, simple structure and it is versatile. Processing elements are called nodes. Also, there are links between them; each connection (link) is involved in the transmission of signals. There is an unlimited amount of processing element inputs and a single output connection. The output of the processing element may be of any desired mathematical type. To indicate the function of the output of the processing element: x input value, y output value, Ψ transfer function, f (x) collecting function, w_{kj} k is the connection weights and θ_k for sequential neurons, including the neuron's sequential threshold for k can be expressed as follows [62].

$$f(x) = w_{k1}x_1 + w_{k2}x_2 + ... + w_{kn}x_{kn} + \theta = \sum_{j=0}^{n} w_{kj}x_j; \quad \theta = w_{k0}x_0 \tag{1}$$

$$y = \psi(f(x)) = \psi\left(\sum_{j=1}^{n} w_{kj}x_j\right) \tag{2}$$

Expressed in the form of a neuron matrix it is as follows:

$$f_k(x) = \begin{bmatrix} w_{k0} & w_{k2} & w_{k3} ... w_{kn} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ . \\ x_n \end{bmatrix} = w_k^T x \tag{3}$$

Inputs are information that enters the cell from other cells or from the external environment. The information enters the cell via links on weights. Weights will determine the effect on the corresponding input cell. The input values are weighted after entering the processing element; in other words, the impact on the system of each input signal can be replaced by the weight assigned to it. Multiplication is used at this point. There are different weight values of all connections to the transmission between neurons in the ANN input. Thus, with the weight impact of all inputs, each processor element has a weight of its own entry. These weights have the same function with the

varying biological effects of synapses in the neurons. In both cases, some entries will be more effective in producing a neural response. The aggregation function is a function that calculates the inputs coming to a cell and is usually calculated as the sum of the product of the entries related to weights.

2.18. Transfer Functions

2.18.1. Hard-Limiter Transfer Function

The graph of the hard limiter transfer function is given in Figure 2.9 below. In this graph, n is the input of the function and a is the output of the function. The formula of this function is a=f (n). In this function; if the input value is greater than zero, the output value is one; if the input value is smaller than zero, the output value is zero. Generally, this function is used in classification applications.

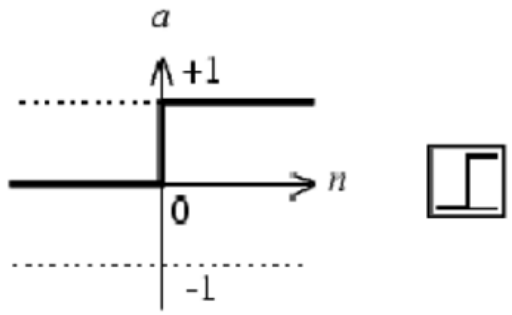


Figure 2. 9 Hard Limit Transfer Function [60]

2.18.2. Linear Transfer Function

The graph of the linear transfer function is given in Figure 2.10. As it can be seen from the figure, the input is given to the output without any change. Here a is equal to n. This function is commonly used in linear filter problems.

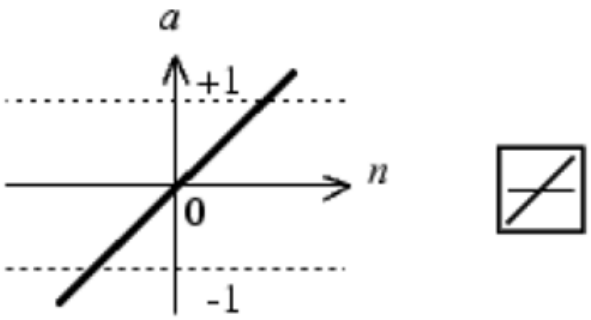


Figure 2. 10 Linear Transfer Function [60]

2.18.3. Log-Sigmoid Transfer Function

The sigmoid is a nonlinear logarithmic function. The output values are between zero and one, independent of what interval the input values are. It is a differential function, so it can be used with back-propagation algorithms. It can be used in the solution of nonlinear problems. The graph of the log-sigmoid transfer function is shown in Figure 2.11.

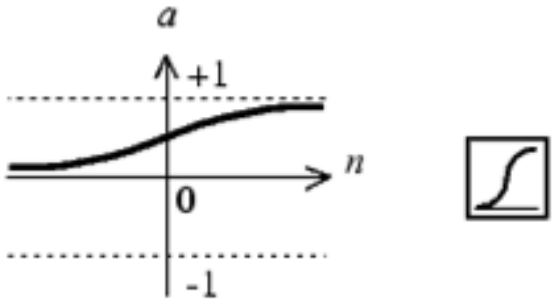


Figure 2. 11 Log-sigmoid transfer function [60].

The mathematical expression of the function is as equation (4) below:

$$a = \log sig(n) = \frac{1}{(1+10^{(-n)})} \qquad (4)$$

2.18.4. Tan-Sigmoid Transfer Function

The graph of the tangent sigmoid transfer function is given in Figure 2.12. The mathematical expression of tan-sigmoid transfer function is as equation 5 below:

$$a = \tan sig(n) = \frac{2}{\left(1+10^{(-2\times n)}\right)} - 1 \qquad (5)$$

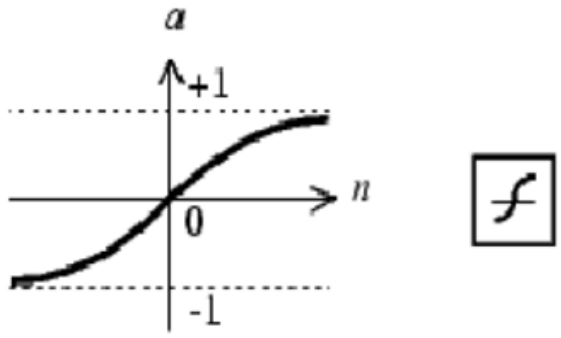


Figure 2.12 Tan-sigmoid transfer function [60].

2.18.5. Learning in Neural Networks

The network connection must ascertain the proper weights and the behaviour of the system in order to organize itself. Information in a neural network is kept in the weights of the neural connections within the network. Therefore, it is important, how to determine the weight. The knowledge that the weight value is stored in the entire network within a node does not mean anything to a single node. Weights across the entire network should have an appropriate value. However, there is a formula originally developed for determining the values of the optimum weights in a neural network. Network processor elements determine optimal weight values using a set of rules over time. This processing is called training. Accordingly, the weight values of a network to be trained must be dynamically varied within a certain measure[63].

Learning events in a neural network take place in two stages. In the first stage, the weight value is taken randomly and the network has to be generated by the output of the network. According to the accuracy of the output value of the second stage, the weights of the network are changed. The objective here is to find a weight value, which will be useful to get an accurate output for these samples. Finding weight values to

produce the correct output of the network indicates that it has the ability to make generalisations about the events represented by the example in the network.

Basically, learning methods are collected in three groups as supervised, unsupervised and reinforcement learning [64].

2.18.6. Supervised Learning

In this method, there is an outside teacher in the neural network. Trainers should produce results relevant for neural network input to the neural network system. So the artificial neural network input / output samples comprise of two presentations.

These two represent the features you need to learn the network. This is part of the network input and generates an output current connection with the information represented by the weight. This output is compared with the output which should be transferred over a network error and intermediate weights are modified to reduce this error.

2.18.7. Unsupervised Learning

There is not any tutor that helps in learning in unsupervised learning. The network gets examples and classifies them according to certain criteria. These criteria can be known in advance. The network itself constitutes their learning criteria.

2.18.8. Reinforcement learning

This method is close to the supervised learning. The reinforcement learning algorithm doesn't need to know the desired output. It does not provide an ANN output to the target output. It uses an accuracy measure to obtain the corresponding output.

2.18.9. Back-propagation Algorithm

The back-propagation algorithm is the most widely used learning algorithm for updating the parameters of the neural network. Today, from the problems with speech recognition solutions such as artificial neural networks to the problems of nonlinear systems, it is used with success in many fields [65].

Due to attempts to reduce the error in the reverse direction, today many versions of the back propagation algorithm have been developed. But the back-propagation algorithm is usually expressed by the generalised delta learning algorithm.

Calculation of the back-propagation algorithm consists of two parts:

1. Advanced calculations
2. Back calculation

2.18.10. Advanced Computing

Advanced methods of calculation start with the administration of the network and from the input layer of each sample in the training data set. Input data is sent to the intermediate layer between the hidden layer and the input layer without any change. Here, the collection function is applied to each of the neural cells in the hidden layer. The input gathered in the hidden layer of neural cells, along with the result of the threshold value of the collection function is calculated. As in biological neural cells, (normalised to the net input of activation functions possessed by every neural cell to generate an electrical signal) it produces an output value for the cell. This creates output value of the input value of the neural cells in the next layer which continues until the output of the computation network is determined. Exit output value is found in the layers and advanced calculation of the network is completed.

2.19. ROC Analysis

The basic idea behind the medical tests is to estimate the likelihood of illness of patients based on their test results. ROC analysis determines the true accuracy of medical diagnostic tests that have been used. The terms used in this analysis, sensitivity to patients (true positive) who test positive (patient) and the number of patients compared to healthy individuals (false positive); where the test was negative (normal) to the number of healthy individuals [16].

ROC analysis is the standard strategy used to decide the sensitivity and specificity of diagnostic methods. To this end, the relationship between the ROC curve sensitivity and specificity of diagnosis is used to define it. The curve between the limits of 0 and

1, and the proximity to the border and above the Y coordinate shows a successful test; whereas the curve slope under 45 degrees shows a failed test [16].

The area under the ROC curve is shown in Figure 2.13. Area 1 represents a perfect test; an area of 0.5 represents a worthless test. A rough guide for the classification accuracy of a diagnostic test system is to use traditional academic points:

90-100 = excellent (A)
80-90 = good (B)
70-80 = fair (C)
60-70 = poor (D)
50-60 = fail (F)

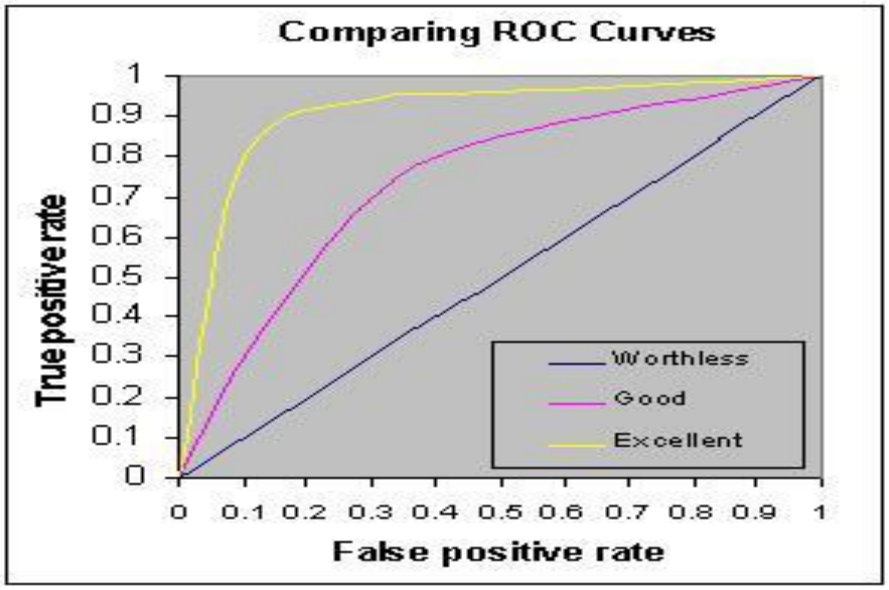


Figure 2.13 Show Different Roc Curves [16]

CHAPTER 3
METHODOLOGY

3.1.Wisconsin Prognostic Breast Cancer (WPBC) Database

The WPBC dataset was prepared by Dr. Wolberg. This data is widely used for classification and regression. This dataset has the cases from patients with both recurrent and non-recurrent cancer types[68]. The dataset comprises of 198 cases (151 non-recurring- 47 recurrence), where each case denotes information following up on one breast tumour case. In-patients were followed at the University of Wisconsin Hospital, during the period from 1984 to 1995 and include just those cases showing intrusive breast tumour and no proof of inaccessible metastases at the time of examination. Every occurrence has 35 attributes, where the initial three credits relate to a unique ID number and to the prediction status (recurring/non-recurring) taken from the recurrence time (time to recur - TTR) or disease-free survival (DFS) time separately. The dataset also includes 30 features, which are computed from a digitised image of fine needle aspirate (FNA) of breast masses while the last two attributes are the diameter of the removed tumour (in centimeters) and the number of positive axillary lymph nodes perceived at the time of the surgical procedure[69].

The WPBC dataset includes the following features: ID of the patient, field of prediction (R = recur, N = non-recur), and time (R => recurrence time, N => disease-free time).The following real-value features are calculated for each cell nucleus: radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry and fractal dimension. The thirty-fourth feature is tumor size and the thirty -fifth is the lymph node status. The diagnosis and prediction have the same features yet the prediction has two additional features as follows: tumour size is the diameter of the erased tumour in centimeters. Tumour size is divided into four classes: T-1 is from 0 - 2 cm. T-2 is from 2 - 5 cm. T-3 is greater than 5cm. T-4 is a tumour of any size that has broken through (ulcerated) skin, or is attached to the chest wall. Lymph node status is the number of positive auxiliary lymph nodes perceived at the time of the surgical procedure. The lymph nodes in the armpit (the axillary lymph nodes) are

the first place breast cancer is likely to spread. Axillary lymph nodes near a breast with cancer are shown in Figure 3.1[70].

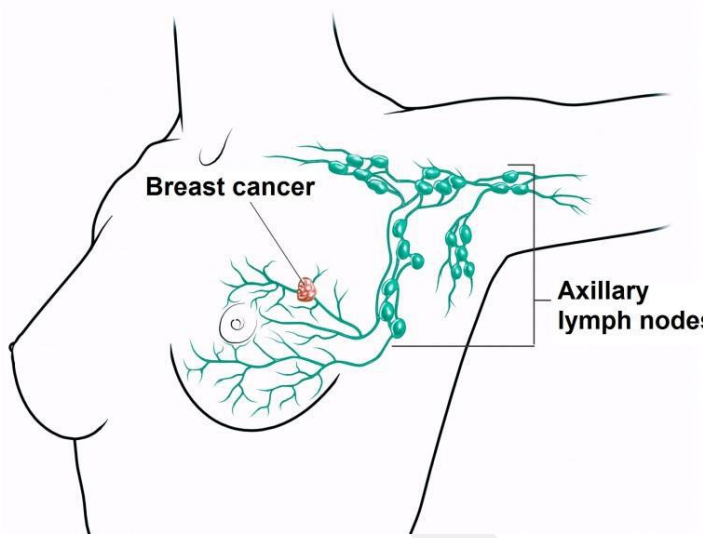


Figure3.1 shows Axillary lymph nodes near a breast with cancer[70]

3.2.Normalisation Method

In this thesis, the method described in[71] is used for normalisation. Here in the first step, all input data are normalised into the range of -1 and 1, and then the data is used in the training algorithm. After getting the output, the normalised data is applied again tothe output data. But this time, the inverse of normalisation is used.

The normalised value of e_i for variable E in the 1st row is calculated as:

$$Normalized\ (e_i) = \frac{e_i - E_{min}}{E_{max} - E_{min}} \qquad (4.1)$$

where

E_{min} = the minimum value for variable E

E_{max} = the maximum value for variable E

The normalisation method then calculates min and max values for each row, and maps the values of each row accordingly. However, when compared with a single column vector, it has inconsistently calculated min and max values for the column and mapped its values accordingly. The function that is used in this thesis is consistent in calculating the min and max values for rows.

3.3.Classifier Methods

3.3.1. Generalised Regression Neural Network

In this thesis, the MATLAB programming languages are used for simulating the results. MATLAB is easy-to-use academic software. For this reason it is used for this work.

A generalised regression neural network (GRNN) is often used to approximate an unknown function. It has a bearing layer and a special linear layer. The architecture of GRNN is shown below. It is similar to the radial base network, but has a second layer which is a little different. The spread is the only variable need to be determined in GRNN. A smaller spread has the potential to provide a better fit to the data, but the function approximation will be less smooth. Newgrnn is used for training the data.

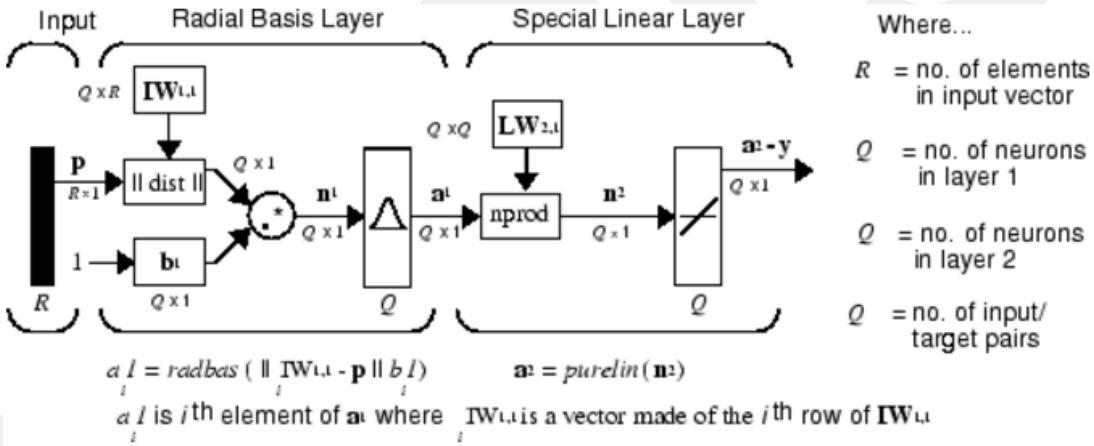


Figure 3.2 Architecture of generalised regression neural network[71]

In this thesis, the different spreads are tested, starting from 0.1 up to 1 in steps of 0.1 and the best spread was obtained at 0.7.

3.3.2. Support Vector Machine

A support vector machine (SVM) classifies samples by creating a hyper-plane. In other words, the algorithm produces an optimal hyper-plane that classifies samples. An example of SVM is shown in the following figure.

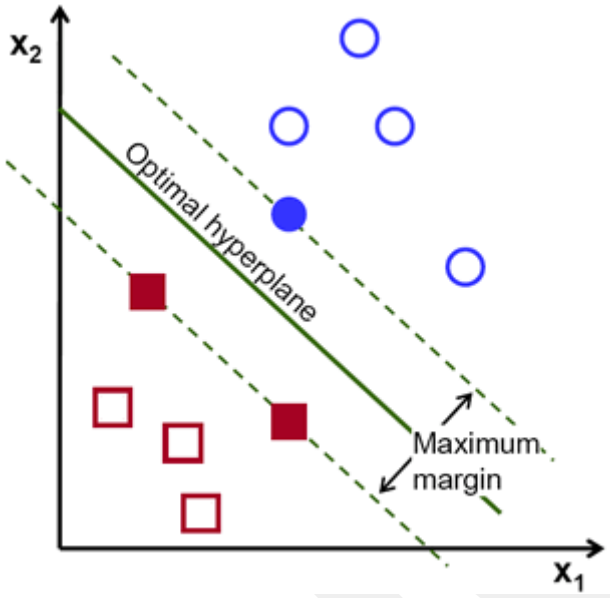


Figure 3.3 SVM for classification of data[72]

In SVM, the class of an input vector $\mathbf{x}$ can be decided by evaluating the sign of $y(\mathbf{x})$.

$$y(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b$$

If $y(\mathbf{x}) > 0$ we assign $\mathbf{x}$ to class +1 and if $y(\mathbf{x}) < 0$, we assign it to class -1. Here $\phi(\mathbf{x})$ can map $\mathbf{x}$ to a space of higher, possibly infinite, dimensions.

Given a data set comprising N input vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and their corresponding labels $t_1, \ldots, t_n$, where $t_n \in \{-1, +1\}$, we would like to find $\mathbf{w}$ and b are so that it explains the training data: $y(\mathbf{x}_n) \geq 1$ when $t_n = +1$ and $y(\mathbf{x}_n) \leq 1$ when $t_n = -1$. This can be rewritten in a single constant:

$$t_n(\mathbf{w}^T \phi(\mathbf{x}) + b) \geq 1, \ \ n = 1, \ldots, N$$

In addition, $\mathbf{w}$ and b are chosen so that the distance between the decision boundary $\mathbf{w}^T \phi(\mathbf{x}) + b = 0$ (a line in the 2-d case, a plane in the 3-d case, a hyper-plane in the n-d case) and the closest points are maximised. This distance is called the margin. Geometrically, the margin is found to be $2/\|\mathbf{w}\|^2$ and so the maximum margin problem can be equivalently expressed as the minimisation problem:

$argmin_{\mathbf{w},b}\frac{1}{2}\|\mathbf{w}\|^2$

In this thesis, the linear kernel functions are used for SVM. We used svmtrain for the training of data and svmclassify for classification of data.

3.3.3. Multilayer Perceptron

The MLP is a feed-forward ANN in which the data is entered at the input layer and propagated in one direction from the hidden layer(s) to the output layer. Hidden and output layers are composed of single units called perceptrons. Each perceptron in the MLP receives a set of inputs which are first weighted and then added together. The resultant value is used to trigger an activation function that will map the combined inputs to an appropriate output response. Train and sim functions are used for training data and simulating data.

In this thesis, the Multilayer perceptron with back propagation algorithm, with a Levenberq Marquart training algorithm, is used for classification of recurrent and non-recurrent data. A different number of neurons are used in the one-hidden-layer MLP and in the two-hidden-layer MLP. In this work, 20 scenarios were tested for the one-hidden-layer MLP. A total of 360 scenarios were used for the two-hidden-layer MLP.

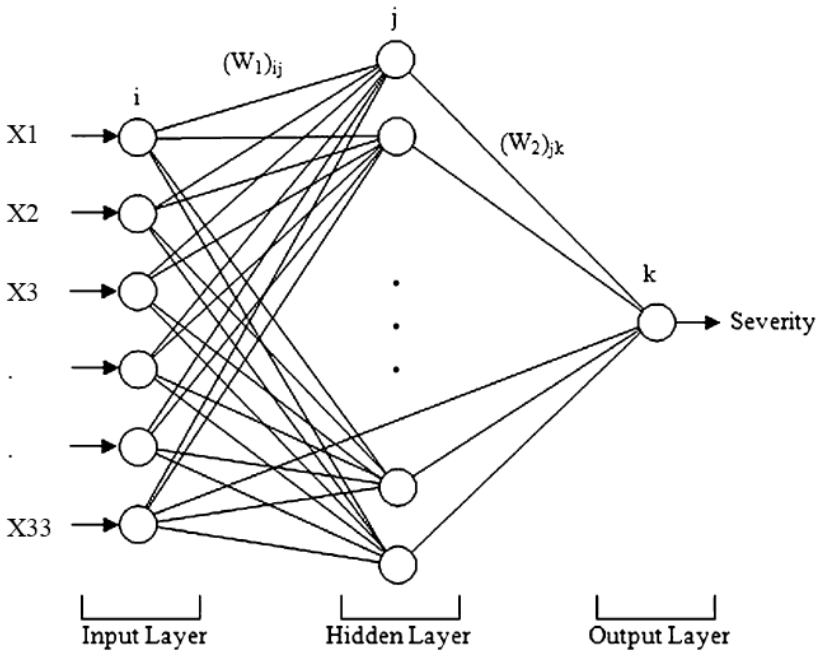


Figure 3.4 Structure of an ANN[60]

3.4.Leave-One-Out Cross-Validation

Leave-One-Out Cross-Validation (LOOCV) is a special case of k-fold cross-validation where k equals the number of instances in the data. In other words, in each iteration closely all the data except for a single observation is used for training. The model is tested on that single observation. An accurate estimate obtained using LOOCV is known to be almost unbiased, but it has high variance, leading to unreliable estimates [3]. It is still widely used when the available data are very rare, especially in Bioinformatics where only dozens of data samples are available.

In leave-one-out for testing all possible ways to divide the original dataset into training and testing sets are used. If you have 5 samples (from 1 to 5), you will do the following steps:

1. The first sample will be for testing, the rest (second, third, fourth, fifth) will be for training.
2. The second sample will be for testing, the rest will be for training.
3. The third sample will be for testing, the rest will be for training.
4. The fourth sample will be for testing, the rest will be for training.

5. The fifth sample will be for testing, the rest will be for training.
6. Mean of accuracy of the previous 5 steps is calculated.

As usual, the true error is estimated as the average error rate on test examples.

$$E = \frac{1}{N}\sum_{i=1}^{N} E_i$$

E is the error and N is a number of samples and is equal to a number of iterations.
LOOCV has been used successfully for model selection for these reasons:

1. It does not waste data.
2. There is no bias in performance estimation.

3.5.Performance Measurement

In order to determine the effectiveness of the classification algorithm used, a measurement is needed. Commonly used measurements include classification accuracy, F-Measure, precision, recall, receiver operating characteristic (ROC) curves and area under curve (AUC). These measurements can be calculated by the classification results commonly tabulated in a matrix format called a Confusion Matrix. A confusion matrix summaries the outcome of the algorithm in a matrix format. For a classical binary classification problem, the classifier labels the items as either positive or negative. In our binary example, the confusion matrix will have four outcomes: True positives (TP) are positive items correctly classified as positive. True negatives (TN) are negative items correctly identified as negatives. False positives (FP) are negative items classified as positive. False negatives (FN) are positives items classified as negative. Table 3.1 illustrates a sample confusion matrix.

Table 3. 1Confusion Matrix

Confusion Matrix		Classified As:	
		Negative	Positive
Actual Class	Negative	TN	FP
	Positive	FN	TP

The following performance measures use the values of the confusion matrix in their calculation.

Classification Accuracy: The simplest performance measure is accuracy. The overall effectiveness of the algorithm is calculated by dividing the correct labelling against all classifications. $accuracy = (TP + TN) / (TP + FP + TN + FN)$.The accuracy that is determined may not be an adequate performance measure when the number of negative cases is much greater than the number of positive cases.

CHAPTER 4
EXPERIMENTAL RESULTS

4.1.Recurrent and Non-Recurrent Cancer

There are different data mining techniques that can be used for the prediction of breast cancer recurrence. In this thesis, we analysed breast cancer data using three classification techniques to predict the recurrence of cancer and then compared the results.

4.2.Summary of Proposed Methods

The summary of the proposed methods is illustrated in the following figure.

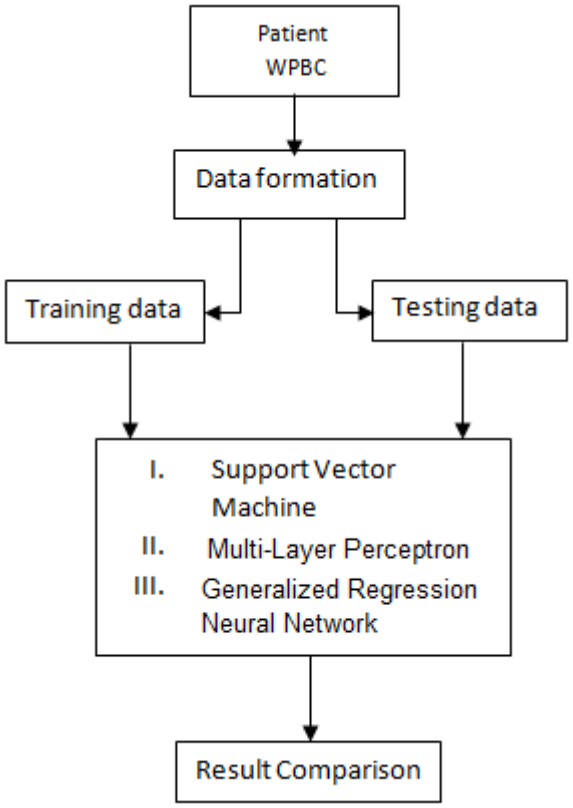


Figure 4. 1Summary of proposed methods

The result of GRNN is shown in table 4.1

Table 4. 1 Result for GRNN

spread	Accuracy
0.1	71.134021
0.2	70.618557
0.3	72.680412
0.4	76.288660
0.5	78.865979
0.6	78.350515
0.7	79.896907
0.8	77.835052
0.9	78.865979
1 “default”	78.350515

The best accuracy is 79.8969 when spread set is at 0.7.

The confusion matrix of GRNN is illustrated in figure 4.2. As seen in this figure the accuracy is 79.9%. Also, in this figure about 145 samples are correctly classified as non-recurring and 10 samples are correctly classified as a recurrence. About 36 samples are mistakenly classified as non-recurring and 3 samples are mistakenly classified as a recurrence.

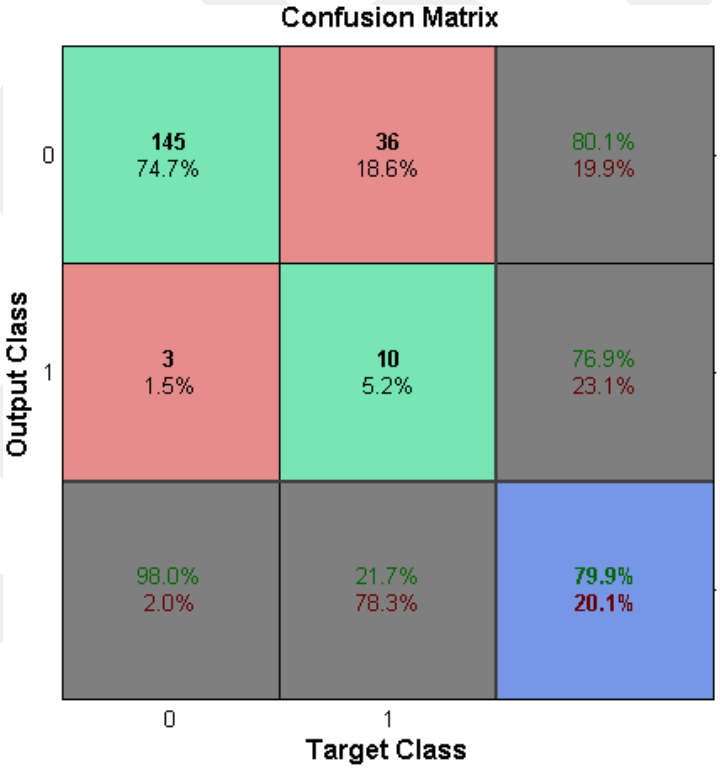


Figure 4. 2The confusion matrix for GRNN

The ROC analysis of GRNN is shown in figure 4.3. As shown in this figure the area under the curve is 59.86%. This curve shows the true positive rate versus false positive rate.

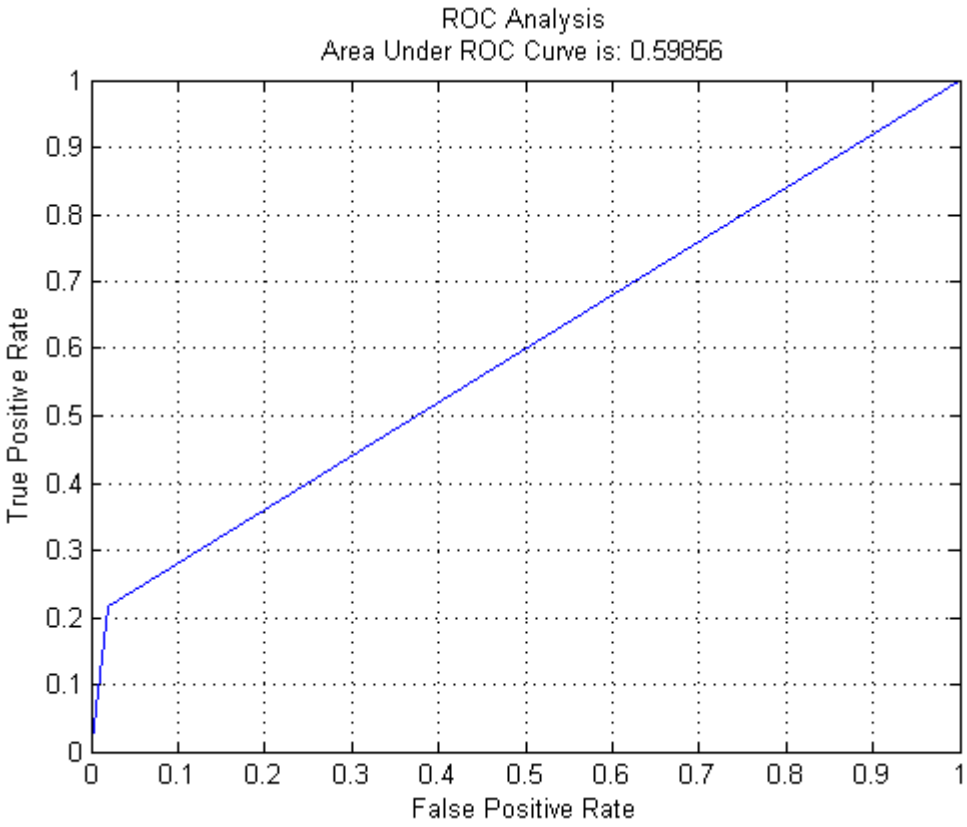


Figure 4. 3 ROC analysis of GRNN

To evaluate a classifier, you have to apply it in a number of cases where one has knowledge of the "true" class of the respective objects, at least in retrospect. An example of such a case is a medical laboratory test, which seeks to determine whether a person has a particular disease. Later it is determined by more complex tests, whether the person actually suffers from this disease. The test classifies the persons in the categories of "sick" and "healthy". Since this is a yes / no question, we also say that the test is positive (classification "sick") or negative (classification "healthy"). In order to assess how well suited the laboratory test for the diagnosis of the disease, its actual state of health is now compared with the result of the tests in each patient. There are four possible cases which may occur.

True positive: The patient is sick, and the test makes the classification correctly.
False negative: The patient is sick, but the test wrongly classifies as healthy.
False positive: The patient is healthy, but the test wrongly classifies as ill.
True negative: The patient is healthy, and the test makes the classification correctly.

In the first and last case, the diagnosis is correct. In the other two cases, there is an error. The four cases are also named differently in different contexts. In the context of detection theory, a true positive case is also known as a hit, a false negative case as a miss and a true negative case as a correct rejection.

It is now counted how often each of the four possible combinations of test results (determined class) and state of health (actual class) has occurred. These frequencies are entered into a so-called truth matrix (also called confusion matrix):

	Person is recurrent (rp + fn)	a person is non-recurrent (fp + rn)
Test positive (rp + fp)	true positive (RP)	false positive (FP)
Test negative (fn + rn)	false negative (fn)	true negative (RN)

Sensitivity or true positive rate (TPR)

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN}$$

Specificity (SPC) or true negative rate (TNR)

$$SPC = \frac{TN}{N} = \frac{TN}{FP + TN}$$

False positive rate (FPR)

$$FPR = \frac{FP}{N} = \frac{FP}{FP + TN} = 1 - SPC$$

Accuracy (ACC)

$$ACC = \frac{TP + TN}{P + N}$$

4.3.Sensitivity and False Negative Rate

Sensitivity (also true-positive rate or hit rate; English sensitivity, false positive rate, recall or hit rate) indicates the proportion of correctly classified as positive properties of the totality of the actual positive properties. For example, corresponding to sensitivity at a medical diagnosis the proportion of actually sick (recurrent), in whom the disease has also been detected. The sensitivity corresponds to the estimated conditional probability

$$\frac{r_p}{r_p + f_n}$$

According tothe result of false-negative rate (false negative rate or miss rate) the proportion of falsely classified as negative objects that are positive in reality, so in the example actually sick (recurrent), but which are diagnosed as healthy.

The false negative rate is the estimated conditional probability.

$$\frac{f_n}{r_p + f_n}$$

4.4.Specificity and False Positive Rate

The specificity (also true-negative rate or distinguishing characteristic; English specificity, false negative rate, or correct rejection rate) indicates the proportion correctly classified as negative listings on the totality of the actual negative objects. For example, the specificity of a medical diagnosis to the proportion of healthy individuals, where no disease is present.

The specificity corresponds to the estimated conditional probability.

$$\frac{r_n}{r_n + f_p}$$

The false-positive rate (also failure rate; English fallout or false positive rate) is the proportion of falsely classified as positive objects that are negative in reality. For

example, someone healthy might be diagnosed sick even though they are not. This is the false-positive rate.

The false positive rate is the estimated conditional probability.

$$\frac{f_p}{r_n + f_p}$$

Since both measures relate to the case, in reality, the negative category is present (second column of truth matrix), the specificity and the false positive rate from 1 to 100%is added.

Table 4. 2 Result for MLP when one node exists in the first hidden layer

First hidden layer	Second hidden layer	Accuracy
1	1	76.804124
	2	75.773196
	3	74.226804
	4	74.742268
	5	74.742268
	6	72.164948
	7	73.711340
	8	72.680412
	9	74.226804
	10	73.195876
	11	75.773196
	12	73.195876
	13	75.773196
	14	73.195876
	15	73.711340
	16	75.257732
	17	73.195876
	18	77.835052
	19	73.711340
	20	72.680412
Average		74.3299
Standard Deviation		1.4921

In this table, the first hidden layer used 1 neuron and tested about 20 neurons in the second layer. The highest accuracy is got for 1 neuron in the second hidden layer. This accuracy is 76.804124.

Table 4. 3 Result for MLP when tow nodes exist in the first hidden layer

First hidden layer	Second hidden layer	Accuracy
2	1	72.164948
	2	74.226804
	3	72.680412
	4	72.164948
	5	71.134021
	6	74.226804
	7	72.680412
	8	74.226804
	9	72.680412
	10	71.134021
	11	71.134021
	12	73.195876
	13	70.618557
	14	70.618557
	15	70.618557
	16	71.649485
	17	75.773196
	18	69.072165
	19	69.587629
	20	72.680412
Average		72.1134
Standard Deviation		1.6965

In this table, the first hidden layer used 2 neurons and tested about 20 neurons in the second layer. The highest accuracy is for 6and 8 neurons in the second hidden layer. This accuracy is 74.226804.

Table 4. 4 Result for MLP when three nodes exist in the first hidden layer

First hidden layer	Second hidden layer	Accuracy
3	1	77.835051
	2	73.195876
	3	73.711340
	4	72.164948
	5	67.010309
	6	69.072165
	7	68.556701
	8	67.010309
	9	71.649485
	10	68.556701
	11	69.072165

	12	69.587629
	13	72.164948
	14	72.164948
	15	70.103093
	16	72.680412
	17	68.556701
	18	74.742268
	19	68.041237
	20	69.587629
Average		70.7732
Standard Deviation		2.8089

In this table, the first hidden layer used 3 neurons and tested about 20 neurons in the second layer. The highest accuracy is for 1 neuron in the second hidden layer. This accuracy is 77.835051.

Table 4. 5 Result for MLP when four nodes exist in the first hidden layer

First hidden layer	**Second hidden layer**	**Accuracy**
4	1	74.742268
	2	71.134021
	3	71.134021
	4	74.742268
	5	69.587629
	6	72.680412
	7	71.134021
	8	68.041237
	9	70.103093
	10	67.010309
	11	70.618557
	12	71.134021
	13	64.432990
	14	66.494845
	15	71.649485
	16	68.556701
	17	68.556701
	18	68.556701
	19	65.979381
	20	72.164948
Average		69.9227
Standard Deviation		2.7494

In this table, the first hidden layer used 4 neurons and tested about 20 neurons in the second layer. The highest accuracy is for 1 and 4 neurons in the second hidden layer. This accuracy is 74.742268.

Table 4. 6 Result for MLP when five nodes exist in the first hidden layer

First hidden layer	Second hidden layer	Accuracy
5	1	72.680412
	2	70.103093
	3	71.134021
	4	68.556701
	5	69.587629
	6	62.886598
	7	68.041237
	8	69.072165
	9	69.587629
	10	71.587629
	11	70.103093
	12	63.917526
	13	70.618557
	14	71.134021
	15	67.010309
	16	69.072165
	17	63.917526
	18	69.072165
	19	67.525773
	20	67.525773
Average		68.6567
Standard Deviation		2.6279

In this table, the first hidden layer used 5 neurons and tested about 20 neurons in the second layer. The highest accuracy is for 1 neuron in the second hidden layer. This accuracy is 72.680412.

Table 4. 7 Result for MLP when six nodes exist in the first hidden layer

First hidden layer	Second hidden layer	Accuracy
6	1	74.226804
	2	73.195876
	3	71.649485
	4	65.979381
	5	71.649485

	6	64.432990
	7	72.164948
	8	64.432990
	9	64.432990
	10	65.979381
	11	65.979381
	12	66.494845
	13	69.072165
	14	63.402062
	15	68.556701
	16	72.164948
	17	67.525773
	18	69.072165
	19	71.134021
	20	69.072165
Average		68.5309
Standard Deviation		3.3174

In this table, the first hidden layer used 6 neurons and tested about 20 neurons in the second layer. The highest accuracy is for 1 neuron in the second hidden layer. This accuracy is 74.226804.

Table 4. 8 Result for MLP when seven nodes exist in the first hidden layer

First hidden layer	**Second hidden layer**	**Accuracy**
7	1	71.649485
	2	69.587629
	3	69.587629
	4	68.041237
	5	63.402062
	6	67.010309
	7	63.402062
	8	69.979381
	9	65.979381
	10	66.494845
	11	69.072165
	12	67.010309
	13	72.164948
	14	61.855670
	15	64.948454
	16	69.525773
	17	71.556701
	18	67.525773
	19	68.556701
	20	66.494845

Average	67.6923
Standard Deviation	2.8499

In this table, the first hidden layer used 7 neurons and tested about 20 neurons in the second layer. The highest accuracy is for 1 neuron in the second hidden layer. This accuracy is 71.649485.

Table 4. 9 Result for MLP when eight nodes exist in the first hidden layer

First hidden layer	Second hidden layer	Accuracy
8	1	71.432990
	2	76.288660
	3	72.041237
	4	69.587629
	5	68.041237
	6	64.948454
	7	63.402062
	8	67.010309
	9	64.948454
	10	65.463918
	11	69.587629
	12	71.134021
	13	64.432990
	14	64.948454
	15	67.010309
	16	63.402062
	17	68.556701
	18	59.278351
	19	66.494845
	20	62.886598
Average		67.0448
Standard Deviation		3.9038

In this table, the first hidden layer used 8 neurons and tested about 20 neurons in the second layer. The highest accuracy is for 20 neurons in the second hidden layer. This accuracy is 84.6154.

Table 4. 10 Result for MLP when nine nodes exist in the first hidden layer

First hidden layer	Second hidden layer	Accuracy
9	1	72.164948
	2	73.195876
	3	68.556701

	4	68.556701
	5	61.340206
	6	65.463918
	7	65.979381
	8	70.618557
	9	72.164948
	10	68.556701
	11	71.649485
	12	65.979381
	13	65.463918
	14	68.041237
	15	62.886598
	16	66.494845
	17	71.649485
	18	65.979381
	19	72.164948
	20	70.103093
Average		68.3505
Standard Deviation		3.3506

In this table, the first hidden layer used 9 neurons and tested about 20 neurons in the second layer. The highest accuracy is for 3 and 11 neurons in the second hidden layer. This accuracy is 84.6154.

Table 4. 11 Result for MLP when ten nodes exist in the first hidden layer

First hidden layer	**Second hidden layer**	**Accuracy**
10	1	69.587629
	2	72.680412
	3	67.525773
	4	69.072165
	5	68.041237
	6	69.072165
	7	64.948454
	8	68.041237
	9	65.463918
	10	72.680412
	11	69.072165
	12	67.010309
	13	63.917526
	14	67.525773
	15	60.824742
	16	65.463918
	17	68.556701
	18	69.072165

	19	60.824742
	20	64.948454
Average		67.2165
Standard Deviation		3.1749

In this table, the first hidden layer used 10 neurons and tested about 20 neurons in the second layer. The highest accuracy is for 2 and 10 neurons in the second hidden layer. This accuracy is 72.680412.

Table 4. 12 Result for MLP when eleven nodes exist in the first hidden layer

First hidden layer	**Second hidden layer**	**Accuracy**
11	1	77.835051
	2	68.041237
	3	72.680412
	4	67.010309
	5	67.010309
	6	68.041237
	7	69.587629
	8	72.164948
	9	67.010309
	10	62.371134
	11	70.618556
	12	68.556701
	13	64.432990
	14	61.855670
	15	66.494845
	16	63.917526
	17	68.556701
	18	61.855670
	19	71.134021
	20	65.979381
Average		67.7577
Standard Deviation		3.9769

In this table, the first hidden layer used 11 neurons and tested about 20 neurons for the second layer. The highest accuracy is for 1 neuron in the second hidden layer. This accuracy is 77.835051.

Table 4. 13 Result for MLP when twelve nodes exist in the first hidden layer

First hidden layer	Second hidden layer	Accuracy
12	1	73.711340
	2	70.103093
	3	66.494845
	4	73.195876
	5	68.556701
	6	63.402062
	7	69.587629
	8	72.164948
	9	67.010309
	10	62.371134
	11	70.618556
	12	68.556701
	13	64.432990
	14	61.855670
	15	67.494845
	16	63.917526
	17	68.556701
	18	62.855670
	19	71.134021
	20	66.979381
Average		67.6500
Standard Deviation		3.6277

In this table, the first hidden layer used 12 neurons and tested about 20 neurons in the second layer. The highest accuracy is for 1 neuron in the second hidden layer. This accuracy is 73.711340.

Table 4. 14 Result for MLP when thirteen nodes exist in the first hidden layer

First hidden layer	Second hidden layer	Accuracy
13	1	71.164948
	2	73.195876
	3	68.556701
	4	61.556701
	5	68.340206
	6	65.463918
	7	65.979381
	8	70.618557
	9	72.164948
	10	68.556701
	11	71.649485
	12	65.979381
	13	64.463918

	14	68.041237
	15	62.886598
	16	66.494845
	17	76.649485
	18	65.979381
	19	72.164948
	20	73.103093
Average		68.6505
Standard Deviation		3.8806

In this table, the first hidden layer used 13 neurons and tested about 20 neurons in the second layer. The highest accuracy is for 17 neurons in the second hidden layer. This accuracy is 76.649485.

Table 4. 15 Result for MLP when fourteen nodes exist in the first hidden layer

First hidden layer	**Second hidden layer**	**Accuracy**
14	1	73.432990
	2	74.288660
	3	73.041237
	4	68.587629
	5	68.041237
	6	64.948454
	7	63.402062
	8	63.010309
	9	69.948454
	10	65.463918
	11	69.587629
	12	71.134021
	13	66.432990
	14	64.948454
	15	62.010309
	16	63.402062
	17	68.556701
	18	66.278351
	19	66.494845
	20	62.886598
Average		67.2948
Standard Deviation		3.7224

In this table, the first hidden layer used 14 neurons and tested about 20 neurons in the second layer. The highest accuracy is for 2 neurons in the second hidden layer. This accuracy is 74.288660.

Table 4. 16 Result for MLP when fifteen nodes exist in the first hidden layer

First hidden layer	Second hidden layer	Accuracy
15	1	71.587629
	2	73.680412
	3	68.525773
	4	69.072165
	5	68.041237
	6	69.072165
	7	64.948454
	8	68.041237
	9	65.463918
	10	73.680412
	11	68.072165
	12	67.010309
	13	63.917526
	14	67.525773
	15	63.824742
	16	66.463918
	17	69.556701
	18	69.072165
	19	60.824742
	20	64.948454
Average		67.6665
Standard Deviation		3.2081

In this table, the first hidden layer used 15 neurons and tested about 20 neurons in the second layer. The highest accuracy is for 2 and 10 neurons in the second hidden layer. This accuracy is 73.680412.

Table 4. 17 Result for MLP when sixteen nodes exist in the first hidden layer

First hidden layer	Second hidden layer	Accuracy
16	1	75.288660
	2	69.041237
	3	76.680412
	4	68.010309
	5	69.010309
	6	68.041237
	7	69.587629
	8	73.164948
	9	68.010309
	10	65.371134
	11	72.618556
	12	69.556701
	13	64.432990

	14	61.855670
	15	66.494845
	16	63.917526
	17	68.556701
	18	65.855670
	19	72.134021
	20	65.979381
Average		68.6804
Standard Deviation		3.8160

In this table, the first hidden layer used 16 neurons and tested about 20 neurons in the second layer. The highest accuracy is for 3 neurons in the second hidden layer. This accuracy is 76.680412.

Table 4. 18 Result for MLP when seventeen nodes exist in the first hidden layer

First hidden layer	**Second hidden layer**	**Accuracy**
17	1	71.164948
	2	72.195876
	3	69.556701
	4	69.556701
	5	63.340206
	6	65.463918
	7	65.979381
	8	70.618557
	9	72.164948
	10	68.556701
	11	71.649485
	12	65.979381
	13	65.463918
	14	69.041237
	15	63.886598
	16	67.494845
	17	70.649485
	18	66.979381
	19	71.164948
	20	71.103093
Average		68.6005
Standard Deviation		2.8266

In this table, the first hidden layer used 17 neurons and tested about 20 neurons in the second layer. The highest accuracy is for 2 neurons in the second hidden layer. This accuracy is 72.195876.

Table 4. 19 Result for MLP when eighteen nodes exist in the first hidden layer

First hidden layer	Second hidden layer	Accuracy
18	1	72.432990
	2	73.288660
	3	74.041237
	4	69.587629
	5	68.041237
	6	65.948454
	7	64.402062
	8	63.010309
	9	68.948454
	10	65.463918
	11	69.587629
	12	71.134021
	13	64.432990
	14	64.948454
	15	67.010309
	16	64.402062
	17	68.556701
	18	59.278351
	19	66.494845
	20	69.886598
Average		67.5448
Standard Deviation		3.7161

In this table, the first hidden layer used 18 neurons and tested about 20 neurons in the second layer. The highest accuracy is for 3 neurons in the second hidden layer. This accuracy is 74.041237.

Table 4. 20 Result for MLP for different number nodes in the first hidden layer

Scenario	Number of Neurons in One Hidden Layer	Accuracy
1	1	73.195876
2	2	77.835051
3	3	75.257732
4	4	70.618557
5	5	77.835051
6	6	77.319587
7	7	72.680412
8	8	74.742268
9	9	77.319587

10	10	77.319587
11	11	73.195876
12	12	76.804123
13	13	74.742268
14	14	75.257732
15	15	76.288659
16	16	75.257732
17	17	74.742268
18	18	71.6494845
19	19	73.195876
20	20	70.618557
Average		74.7938
Standard Deviation		2.3107

In this table, the best accuracy is for 2 and 5 neurons and this value is 77.835051.

The confusion matrix of MLP, which include the following structures neurons 2 and 5 in the first hidden layer and also 3 and 11 neurons in the first hidden layer and one neuron in the second hidden layer is illustrated in figure 4.4. As seen in this figure, the accuracy is 77.8%. Also, in this figure about 146 samples are correctly classified as non-recurring and 5 samples are correctly classified as a recurrence. About 41 samples are mistakenly classified as non-recurring and 2 samples are mistakenly classified as a recurrence.

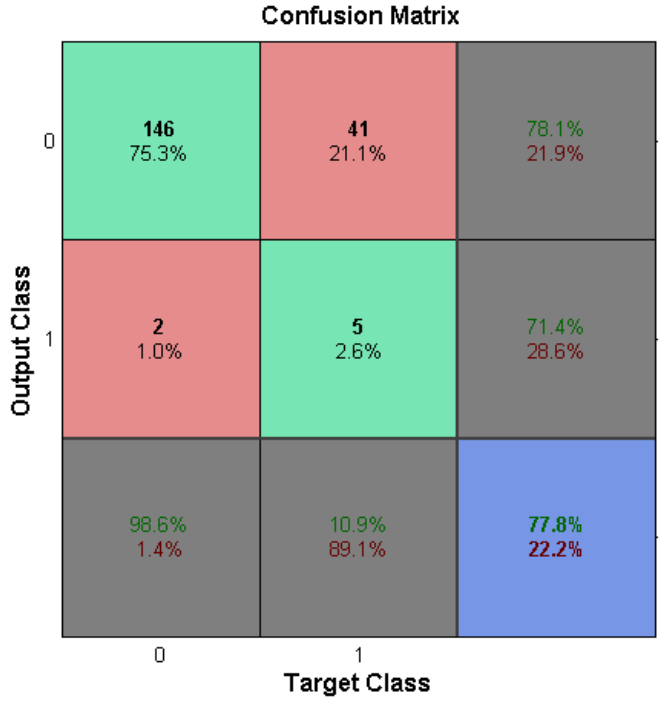


Figure 4. 4 The confusion matrix for MLP

The ROC analysis of MLP is shown in figure 4.5. As shown in this figure the area under the curve is 54.75%. This curve shows the true positive rate versus false positive rate.

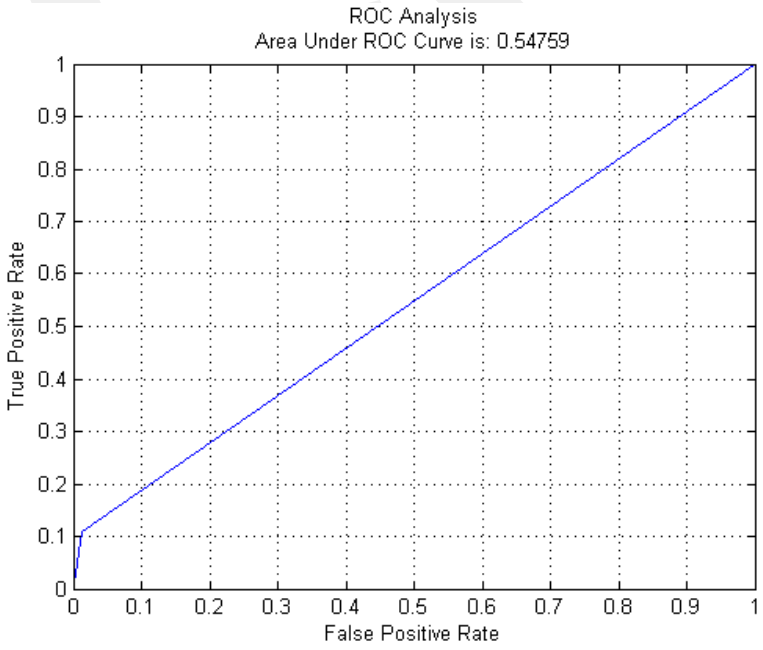


Figure 4. 5 ROC analysis of MLP.

4.5.Summary of MLP Results

We got the best results of MLP in the following scenarios:

- (3-1-1) and (11-1-1) when we use two hidden layers
- (2-1) and (5-1) when we use one hidden layer.

4.6.Support Vector Machine Results

In this thesis, we used linear support vector machine for recurrent and non-recurrent prediction. The accuracy is 75.773196%. The confusion matrix of SVM is illustrated in figure 4.6. As seen in this figure the accuracy is 75.8%. Also, in this figure about 115 samples are correctly classified as non-recurring and 32 samples are correctly classified as a recurrence. About 14 samples are mistakenly classified as non-recurring and 33 samples are mistakenly classified as a recurrence.

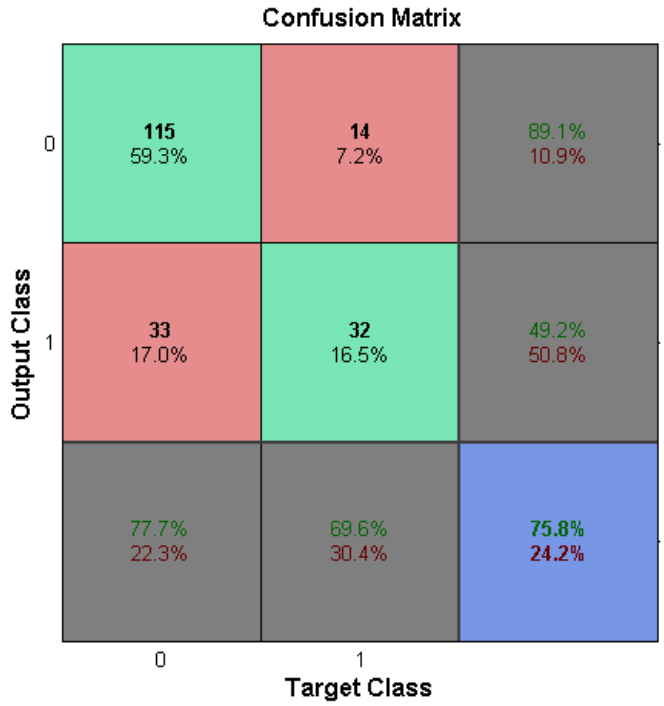


Figure 4. 6 The confusion matrix for SVM

The ROC analysis of SVM is shown in figure 4.7. As shown in this figure the area under the curve is 73.63%. This curve shows the true positive rate versus false positive rate.

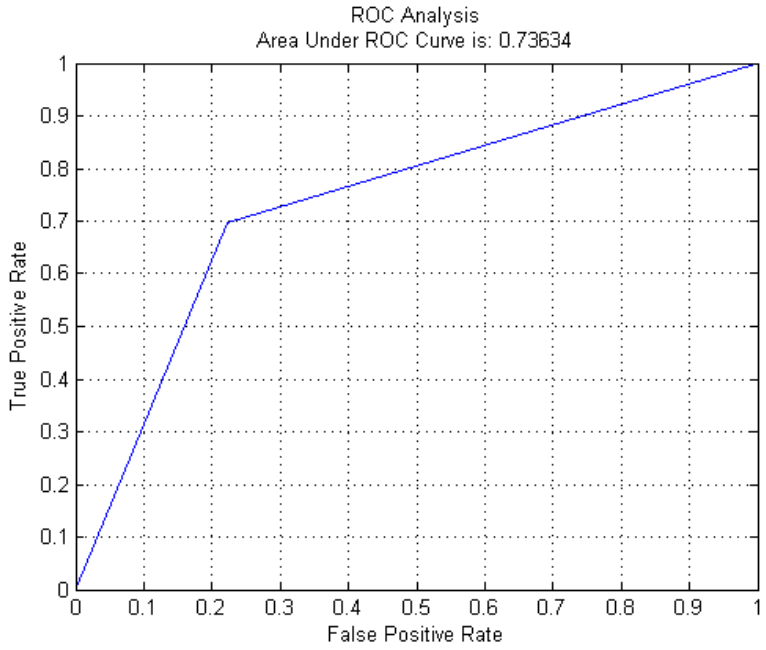


Figure 4. 7 ROC analysis of SVM.

4.7.Statistical Quality Criteria of Classification

By calculating different parameters for assessing, relative abundances of the classifier can now be calculated from the values of truth matrix. This can also be interpreted as estimates of the conditional probability of the occurrence of the corresponding event. The dimensions differ according to the population, to which the relative frequencies relate. Thus, for instance, only all the cases should be considered, in which the positive or negative category actually exists (sum of the entries in a column of truth matrix), or considering the set of all objects that are classified as positive or negative (sum of the entries in a row of the truth matrix). This choice has a serious impact on the calculated values, especially when one of the two classes occurs much more frequently in total than the other.

CHAPTER 5
DISCUSSION AND CONCLUSIONS

5.1.Discussion

In this thesis, the aim was to classify breast cancer data and to estimate the classification accuracy of the designed classifier. Feature selection has not been a direct aim of the study as it considered all feature combinations to investigate each feature potential in output classification. Therefore, the research focused on two stages of pattern classification system design; namely, classifier design and classification error estimation considering the limitations of the available dataset – small sample size (WPDC) and high complexity.

The Stages involved in designing a pattern classifier are as follows:

1. Feature selection (use all features)
2. Classifier Design (Considering the small sample size and high data complexity).
3. Classification Error Estimation (Considering a small sample size).

The comparison between our methods is illustrated in the followingtable.

Table 5. 1 Comparison between our methods

MLP	GRNN	SVM
77.8 %	79.9%	75.8 %

As shown in this table the best result is obtained when GRNN is used and the accuracy value is 79.9%. The GRNN is a powerful algorithm for classification and used for the regression method for classification and does not require an iterative procedure and training weights as in MLP. Nevertheless, in ANN, the used method is the Levenberq Marquardt method and in this algorithm, there are no regression analyses of data.

Besides, in Support Vector Machine there are no regression analyses and this is the disadvantage of the classification method.

In addition to these, we noticed that the SVM is not more efficient than the MLP in learning and test sets. Because SVM has so many parameters which need to be adjusted to learn the data, whereas MLP is easy to learn but needs a lot of time to train the data.

For MLP, the best result is for 2 and 5 neurons in the first hidden layer, and this accuracy value is 77.8%. Also, for two layers we got the same results in the two following scenarios:

1. Three neurons in first hidden layer, and one neuron in second hidden layer (3-1-1).
2. Eleven neurons in first hidden layer, and one neuron in second hidden layer (11-1-1).

The methods employed to achieve these aims and the simulation results were explained in the previous chapters. The MLP, GRNN, and SVM have been designed and developed during the period of this study based on all the available markers and the criteria for achieving high accuracy. In previous researches [66, 67]indicate the lack of agreement on a combination of predictive markers for accurate small sample size and complex nonlinear input-output relationship are common problems in predicting the metastasis in breast cancer. Since many biomarkers, obtained in a non-invasive manner, have been identified to be related to the state of nodal involvement in breast cancer, it is advantageous to design a classifier, which can classify patients using the non-invasively acquired data. For this purpose, the designed classifier should be able to capture the complex nonlinear input-output relation while avoiding over-fitting to the available data. This necessitates designing the classifier with all available cases, and then to evaluate the generalisation ability of the designed classifier using a resampling method.

In this study, the MLP, GRNN, and SVM have been employed as a platform to predict the state of nodal involvement in breast cancer patients based on some Biomolecular

markers as well as their different combinations. An MLP neural network was also designed for this application as a benchmark method to be compared against the ANN results.

Table 5.2.Shows the comparison between the proposed method in the current study and other ones in the literature.

Table 5. 2 Comparison between proposed method and other literature

Multilayer Perceptron[66]	95.74%
Generalised Regression Neural Network (GRNN) [66]	98.8%
Support Vector Machine (SVM) [73]	97.37%
Proposed Method for MLP	77.8%
Proposed Method for GRNN	79.9%
Proposed Method for SVM	75.8%

There may be several reasons for the differences between the results of the current study and the ones in the literature. These reasons will be discussed in the following sections.

3.1.1. Sampling

Sampling the data is a process in which two choices have to be made. The first choice is regarding the sampling size while the second choice is regarding the sampling rate. The sampling size is an important factor in the forecasting ability of the prediction models. In [66]they illustrated that a large sample (699 samples) outperformed a smaller sample (222 samples) in the training of an ANN for the purpose of exchange rate prediction. In [66] they used the database containing 699 samples with 683 complete data and 16 samples with missing attributes. In this thesis, 194 samples with 188 complete data and 4 samples with missing attributes were used.

3.1.2. Normalisation

In [66] they didn't use any normalisation. In this thesis, we normalized the data which was defined in section 3.2. This may also have an effect on the performance of the classifier.

3.1.3. Cross-validation Type

Dividing the data into several subsets by appointing specific data for training purposes and for test purposes, is required for building a GRNN and MLP. Usually, the data are divided into two subsets, one set for training, one set for testing. The training set is used for developing the GRNN and MLP, the test set is used to evaluate the prediction ability of the GRNN and MLP. The validation set is usually part of the test set and is used to avoid the over-fitting problem or to determine the stopping point of the training process. In[66] they used cross-validation 10- fold and. However, in this work, we used leave-one-outcross-validation in order to estimate the performance of the classifier.

3.1.4. Comparison on GRNN

The performance of a GRNN is substantially lower compared to the one in[66]. In [66] they used a different database and also in their GRNN they used a linear algorithm for the training of data. In our proposed GRNN, the used dataset has less samples than the literature. In [66], half of the database was used for training where 222 samples of the training data belong to the benign class and 120 samples belong to the malignant class. On the other hand, in this study, 194samples were used for training and testing. In simulation, in each iteration one of the data was left for testing data and the remaining data were used for training. This technique assessed an unbiased performance estimation. Therefore, the accuracy of proposed GRNN is less than the accuracy in the literature. The accuracy values are presented in Table 5.3. As seen in this table the sensitivity of GRNN is high.

Table 5. 3 Classification accuracy for leave one out of data for GRNN

Dataset	**Sensitivity**	**Specificity**	**Accuracy**
WPBC	98.0%	21.7%	79.9%

3.1.5. Comparison between GRNN, MLP, and SVM

The accuracy was 79.9% in GRNN, 77.8% in MLP, and 75.8% in SVM. All these classifiers usedthe leave-one-out-cross-validation for 194 samples. GRNN has the advantage that it doesn't use a lot of iteration for convergence and doesn't need a lot of parameters. It just needs to spread. Also, SVM is less accurate than the GRNN and MLP because the reason is due to the large (compared to NN) number of parameters required for configuration:

- Choice of kernel (such as linear and polynomial and RBF etc.)
- Selection of kernel parameters (such as RBF sigma, constraint box, poly order)
- Selection of the value of the margin parameter

The accuracy values for MLP are presented in Table 5.4. As seen in this table the sensitivity of MLP is very high.

Table 5. 4 Classification Accuracy for leave one out of Data for MLP

Dataset	**Sensitivity**	**Specificity**	**Accuracy**
WPBC	98.6%	10.9%	77.8%

In MLP, network weights (the MLP's fitting parameters, adjusted during training) are adjusted so that the sum-of-square error between the network output and the actual value (target) is minimized. To find the best architecture of the neuron number is computationally expensive. In this study, the neuron architecture is tested about 380 times.

Training an SVM, by contrast, means an explicit determination of the decision boundaries directly from the training data. This is of course required as the predicate step to the optimisation problem required to build an SVM model: minimising the aggregate distance between the maximum-margin hyper-plane and the support vectors. In practice, though, it is harder to configure the algorithm to train an SVM. The accuracy values for SVM are presented in Table 5.5. As seen in this table the sensitivity of SVM is less than the GRNN and MLP.

Table 5. 5Classification Accuracy for leave one out of Data for SVM

Dataset	**Sensitivity**	**Specificity**	**Accuracy**
WPBC	77.7%	69.6%	75.8%

5.2.Conclusions

In this thesis, the Wisconsin Breast Cancer Prognosis data set is used with 194 samples of real clinical data. Breast cancer prediction is implemented by using MLP, GRNN, and SVM. The best result is obtained when GRNN is used. The experimental results have been presented in the previous chapter in the three main sections of the study: 1) Multilayer Perceptron (MLP), 2) Generalized Recurrent Neural Network (GRNN), and 3) Support Vector Machine (SVM). Chapter 4 covered the experiments planned to investigate the reliability of different error estimation methods for the designed ANNs in this study. The results of the MLP, GRNN, and SVM including their accuracy, sensitivity, specificity and apparent error have been computed and compared. These results confirm the superiority of the MLP, GRNN and SVM in obtaining more accurate results. Moreover, the sensitivity and specificity of the GRNN are higher or close to that of the MLP. Nevertheless, the apparent errors obtained from the SVM are only slightly different from the GRNN and MLP. This can prove that the degree of generalisation of the GRNN is comparable with that of the MLP and SVM.

The main steps in the design of a pattern classifier system are feature selection, classifier design, and classification error estimation. All these steps should be designed

considering the size, dimensionality and complexity of the available data in order to achieve an optimum pattern classification system, which yields maximum accuracy.

5.3.Future Work

Future plans will be to focus on working image processing, and towards developing segmentation algorithms used to be able to detect and extract the cancerous regions, even if there is more than one suspected region in one image. Features space of the masses and calcifications is extensive which opens the door to improving a classifier to determine the type of lesion as a benign or a malignant tumour, so the expected result would be malignant mass, benign calcification or normal. Some suggestions for later work may be as follows. The performance of the segmentation algorithms should be improved through developing mechanisms to choose the threshold depending on the statistical information of the image histogram. The more features of the abnormality should be extracted using advanced algorithms. Besides, the most appropriate features should be selected. A high sensitivity classifier should be designed to accurately determine abnormal area as a benign or malignant tumour.

REFERENCES

[1] L. Jones, M. Bazargan, A. Lucas-Wright, J. V. Vadgama, R. Vargas, J. Smith, *et al.*, "Comparing perceived and test-based knowledge of cancer risk and prevention among Hispanic and African Americans: an example of community participatory research," *Ethnicity & disease,* vol. 23, p. 210, 2013.

[2] "http://www.cancercouncil.com.au/wp-content/," 2014.

[3] H.-S. Juon, S. Choi, A. Klassen, and D. Roter, "Impact of breast cancer screening intervention on Korean-American women in Maryland," *Cancer detection and prevention,* vol. 30, pp. 297-305, 2006.

[4] W. N. Street, O. L. Mangasarian, and W. H. Wolberg, "An inductive learning approach to prognostic prediction," in *ICML*, 1995, pp. 522-530.

[5] N. C. Dolan, A. M. Lee, and M. M. McDermott, "Age-related differences in breast carcinoma knowledge, beliefs, and perceived risk among women visiting an academic general medicine practice," *Cancer*, vol. 80, pp. 413-420, 1997.

[6] V. Özmen, "Breast Cancer in Turkey: Clinical and Histopathological Characteristics (Analysis of 13.240 Patients)."

[7] P. http://www.hopkinsmedicine.org/healthlibrary/test_procedures/gynecology/breast_magnetic_resonance_imaging_mri_92.

[8] K. Sim, F. Chia, S. Chong, C. Tso, S. F. Abbas, and S. Omar, "Spatial based breast MR images lesion detection and colourisation," in *2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, 2010.

[9] C. Varini, T. W. Nattkemper, A. Degenhard, and A. Wismuller, "Breast MRI data analysis by LLE," in *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, 2004, pp. 2449-2454.

[10] H. Wang, Z. Huo, and J. Zhang, "Computerized classification method for differentiating between benign and malignant lesions on breast MR images," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, 2006, pp. 6950-6952.

[11] M. ISA, N. ASHIDI, and S. S. TING, "Automatic segmentation and detection of mass in digital mammograms," 2012.

[12] G. Bhat, V. G. Biradar, H. Sarojadevi, and N. Nalini, "Artificial Neural Network based Cancer Cell Classification," *Computer Engineering and Intelligent Systems*, vol. 3, pp. 7-16, 2012.

[13] R. C. Gonzalez, *Digital image processing*: Pearson Education India, 2009.

[14] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *Systems, Man and Cybernetics, IEEE Transactions on,* vol. 8, pp. 460-473, 1978.

[15] K. R. Castleman, "Digital image processing," *Englewood Cliffs, NJ,* 1996.

[16] K. Thangavel, M. Karnan, R. Sivakumar, and A. K. Mohideen, "Automatic detection of microcalcification in mammograms–A review," *International Journal on Graphics Vision and Image Processing,* vol. 5, pp. 31-61, 2005.

[17] S. I. Niwas, P. Palanisamy, K. Sujathan, and E. Bengtsson, "Analysis of nuclei textures of fine needle aspirated cytology images for breast cancer diagnosis using complex Daubechies wavelets," *Signal Processing,* vol. 93, pp. 2828-2837, 2013.

[18] A. R. Rivera, B. Ryu, and O. Chae, "Content-aware dark image enhancement through channel division," *Image Processing, IEEE Transactions on,* vol. 21, pp. 3967-3980, 2012.

[19] Z. Wang, C. Clavijo, E. Roessl, U. van Stevendaal, T. Koehler, N. Hauser, *et al.*, "Image fusion scheme for differential phase contrast mammography," *Journal of Instrumentation,* vol. 8, p. C07011, 2013.

[20] R. Mekle, A. F. Laine, S. Smith, C. Singer, T. Koenigsberg, and M. Brown, "Evaluation of a multiscale enhancement protocol for digital mammography," in *International Symposium on Optical Science and Technology*, 2000, pp. 1038-1049.

[21] M. Kowal, P. Filipczuk, A. Obuchowicz, J. Korbicz, and R. Monczak, "Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images," *Computers in biology and medicine,* vol. 43, pp. 1563-1572, 2013.

[22] A. Papadopoulos, D. I. Fotiadis, and L. Costaridou, "Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques," *Computers in biology and medicine,* vol. 38, pp. 1045-1055, 2008.

[23] R. K. Jha, R. Chouhan, and P. Biswas, "Noise-induced contrast enhancement of dark images using non-dynamic stochastic resonance," in *Communications (NCC), 2012 National Conference on*, 2012, pp. 1-5.

[24] R. Garg, B. Mittal, and S. Garg, "Histogram equalization techniques for image enhancement," *International Journal of Electronics & Communication Technology,* vol. 2, pp. 107-111, 2011.

[25] J. Luts, F. Ojeda, R. Van de Plas, B. De Moor, S. Van Huffel, and J. A. Suykens, "A tutorial on support vector machine-based methods for classification problems in chemometrics," *Analytica Chimica Acta,* vol. 665, pp. 129-145, 2010.

[26] W. Zhiming and T. Jianhua, "A fast implementation of adaptive histogram equalization," in *Signal Processing, 2006 8th International Conference on*, 2006.

[27] S. M. Pizer, B. M. Hemminger, and R. E. Johnston, "Display, Including Enhancement, of Two-Dimensional Images," *Image-Processing Techniques for Tumor Detection,* pp. 101-130, 1998.

[28] S. Pizer, "An automatic intensity mapping for the display of CT scans and other images," *Proc. 7th Int. Mtg. on Information Process. in Medical Imaging,* pp. 276-309, 1981.

[29] A. Phinyomark, S. Jitaree, P. Phukpattaranont, and P. Boonyapiphat, "Texture analysis of breast cancer cells in microscopic images using critical exponent analysis method," *Procedia Engineering,* vol. 32, pp. 232-238, 2012.

[30] J. Wu and A. Chung, "A segmentation method using compound Markov random fields based on a general boundary model," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, 2005, pp. II-1182-5.

[31] A. R. Dominguez and A. K. Nandi, "Enhanced multi-level thresholding segmentation and rank based region selection for detection of masses in mammograms," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, 2007, pp. I-449-I-452.

[32] I. K. Maitra, S. Nag, and S. Bandyopadhyay, "Automated digital mammogram segmentation for detection of abnormal masses using binary homogeneity enhancement algorithm," *Indian. J. Com. Sci. Engg*, pp. 416-427, 2011.

[33] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *ICML*, 1996, pp. 148-156.

[34] M. Fairhurst and A. F. R. Rahman, "Generalised approach to the recognition of structurally similar handwritten characters using multiple expert classifiers," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 144, pp. 15-22, 1997.

[35] S. Džeroski and B. Ženko, "Is combining classifiers with stacking better than selecting the best one?," *Machine learning*, vol. 54, pp. 255-273, 2004.

[36] L. I. Kuncheva, "A theoretical study on six classifier fusion strategies," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, pp. 281-286, 2002.

[37] M. S. B. Sehga, I. Gondal, and L. Dooley, "Statistical neural networks and support vector machine for the classification of genetic mutations in ovarian cancer," in *Computational Intelligence in Bioinformatics and Computational Biology, 2004. CIBCB'04. Proceedings of the 2004 IEEE Symposium on*, 2004, pp. 140-146.

[38] A. F. Cabrera, "Logistic regression analysis in higher education: An applied perspective," *Higher education: Handbook of theory and research*, vol. 10, pp. 225-256, 1994.

[39] E. Biganzoli, P. Boracchi, L. Mariani, and E. Marubini, "Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach," *Statistics in Medicine*, vol. 17, pp. 1169-1186, 1998.

[40] D. W. Hosmer Jr and S. Lemeshow, *Applied logistic regression*: John Wiley & Sons, 2004.

[41] K. J. Van Zee, D.-M. E. Manasseh, J. L. Bevilacqua, S. K. Boolbol, J. V. Fey, L. K. Tan, *et al.*, "A nomogram for predicting the likelihood of additional nodal metastases in breast cancer patients with a positive sentinel node biopsy," *Annals of surgical oncology*, vol. 10, pp. 1140-1151, 2003.

[42] S. J. Press and S. Wilson, "Choosing between logistic regression and discriminant analysis," *Journal of the American Statistical Association*, vol. 73, pp. 699-705, 1978.

[43] K. Fukunaga and J. Mantock, "Nonparametric discriminant analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, pp. 671-678, 1983.

[44] S. S. Cross, R. F. Harrison, and R. L. Kennedy, "Introduction to neural networks," *The Lancet*, vol. 346, pp. 1075-1079, 1995.

[45] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*: John Wiley & Sons, 2012.

[46] F. Murtagh, "Multilayer perceptrons for classification and regression," *Neurocomputing*, vol. 2, pp. 183-197, 1991.

[47] G. P. Zhang, "Neural networks for classification: a survey," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 30, pp. 451-462, 2000.

[48] E. J. Gamito and E. D. Crawford, "Artificial neural networks for predictive modeling in prostate cancer," *Current oncology reports,* vol. 6, pp. 216-221, 2004.
[49] J. A. Gómez-Ruiz, J. M. Jerez-Aragonés, J. Muñoz-Pérez, and E. Alba-Conejo, "A neural network based model for prognosis of early breast cancer," *Applied Intelligence,* vol. 20, pp. 231-238, 2004.
[50] T. M. Mitchell, "Machine learning. WCB," ed: McGraw-Hill Boston, MA:, 1997.
[51] J. A. Cruz and D. S. Wishart, "Applications of machine learning in cancer prediction and prognosis," *Cancer informatics,* vol. 2, p. 59, 2006.
[52] Y. Su, J. Shen, H. Qian, H. Ma, J. Ji, H. Ma, *et al.*, "Diagnosis of gastric cancer using decision tree classification of mass spectral data," *Cancer science,* vol. 98, pp. 37-43, 2007.
[53] A. Vlahou, J. O. Schorge, B. W. Gregory, and R. L. Coleman, "Diagnosis of ovarian cancer using decision tree classification of mass spectral data," *BioMed Research International,* vol. 2003, pp. 308-314, 2003.
[54] E. Papageorgiou, C. Stylios, and P. Groumpos, "A combined Fuzzy Cognitive Map and decision trees model for medical decision making," in *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, 2006, pp. 6117-6120.
[55] M. U. Khan, J. P. Choi, H. Shin, and M. Kim, "Predicting breast cancer survivability using fuzzy decision trees for personalized healthcare," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, 2008, pp. 5148-5151.
[56] J. R. Quinlan, "Induction of decision trees," *Machine learning,* vol. 1, pp. 81-106, 1986.
[57] J. Bray, J. Sludden, M. Griffin, M. Cole, M. Verrill, D. Jamieson, *et al.*, "Influence of pharmacogenetics on response and toxicity in breast cancer patients treated with doxorubicin and cyclophosphamide," *British journal of cancer,* vol. 102, pp. 1003-1009, 2010.
[58] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of mathematical biology,* vol. 52, pp. 99-115, 1990.
[59] B. Kosko, "Unsupervised learning in noise," *Neural Networks, IEEE Transactions on,* vol. 1, pp. 44-57, 1990.
[60] S. S. Haykin, S. S. Haykin, S. S. Haykin, and S. S. Haykin, *Neural networks and learning machines* vol. 3: Pearson Education Upper Saddle River, 2009.
[61] H. C. Lee, R. Ramotowski, and R. Gaensslen, *Advances in fingerprint technology*: CRC press, 2001.
[62] S. Haykin and N. Network, "A comprehensive foundation," *Neural Networks,* vol. 2, 2004.
[63] B. D. Ripley, *Pattern recognition and neural networks*: Cambridge university press, 1996.
[64] C. M. Bishop, *Neural networks for pattern recognition*: Oxford university press, 1995.
[65] R. N. Naguib and G. V. Sherbet, *Artificial neural networks in cancer diagnosis, prognosis, and patient management*: CRC press, 2001.
[66] T. Kıyan and T. Yıldırım, "Breast cancer diagnosis using statistical neural networks," 2003.

[67] M. Senthilarasi, S. Md Mansoor Roomi, and M. Prasanna, "Shape based approach for detecting Musa Species in fruit industry," in *Advanced Computing (ICoAC), 2014 Sixth International Conference on*, 2014, pp. 157-160.

[68] I. Anagnostopoulos, C. Anagnostopoulos, A. Rouskas, G. Kormentzas, and D. Vergados, "The Wisconsin Breast Cancer Problem: Diagnosis and DFS time prognosis using probabilistic and generalised regression neural classifiers," *Draft Version of paper to appeared in the Oncology Reports, special issue Computational Analysis and Decision Support Systems in Oncology, last quarter*, 2005.

[69] R. A. Khan, N. Ahmad, and N. Minallah, "Classification and Regression Analysis of the Prognostic Breast Cancer using Generation Optimizing Algorithms," *International Journal of Computer Applications*, vol. 68, 2013.

[70] G. I. Salama, M. Abdelhalim, and M. A.-e. Zeid, "Breast cancer diagnosis on three different datasets using multi-classifiers," *Breast Cancer (WDBC),* vol. 32, p. 2, 2012.

[71] I. Saritas, I. A. Ozkan, and I. U. Sert, "Prognosis of prostate cancer by artificial neural networks," *Expert Systems with Applications*, vol. 37, pp. 6646-6650, 2010.

[72] http://docs.opencv.org/2.4/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html.

[73] K. Menaka and S. Karpagavalli, "Breast Cancer Classification using Support Vector Machine and Genetic Programming," *International Journal of Innovative Research in Computer and Communication Engineering* vol. 1, pp. 1410-1417, 2013.

APPENDICES A

CURRICULUM VITAE

PERSONAL INFORMATION
Surname, Name: Aymen Alhasadi
Date and Place of Birth: 1979/04/29
Marital Status:Married
Phone: +905392293070,
Email: afh_amon@yahoo.com

EDUCATION:

Degree	Institution	Year of Graduation
M.Sc.	Atilim University	2015-2016
High Diploma	Higher Institute For Comprehensive Professions, Derna	2002-2003
Secondary School	Alintfithia School	1996-1997

Certifications:

1) **OZONE certified in Scientific Research Methods ,2016**
2) **OZONE certified In References Management (END NODE SOFTWARE),2016**
3) **ICDL (**International Computer Driving License**), 2010.**
4) **Microsoft certificated in MCSD (Microsoft Certified Solutions Developer), 2007.**
5) **Maintenance Certification of personal computers from the Arabic Union for Technologies), 2005.**

Foreign languages: English and Beginner Turkish

Hobbies: Reading, swimming, and traveling