
GAZİ ÜNİVERSİTESİ
Gazi
1926

EMPLOY THE WAVELET TRANSFORMS AND ARTIFICIAL NEURAL NETWORKS FOR EARLY DETECTION AND CLASSIFICATION OF BREAST CANSER BASED ON STATISTICAL FEATURES OF MEDICAL IMAGES

Karzan Faidhi HAMAD

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTORA OF PHILOSOPHY
IN STATISTICS DEPARTMENT

GAZİ UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

DECEMBER 2024

ETHICAL STATEMENT

I at this moment declare that in this thesis study, I prepared the thesis writing rules of Gazi University Graduate School of Natural and Applied Sciences;

- All data, information, and documents presented in this thesis have been obtained within the scope of academic rules and ethical conduct,
- All information, documents, assessments, and results have been presented by scientific ethical conduct and moral rules,
- All material used in this thesis that is not original to this work has been exhaustively cited and referenced,
- No change has been made in the data used,
- The work presented in this thesis is original,

or else, I admit all loss of rights to be incurred against me.

Karzan Faidhi HAMAD

03/12/2024

EMPLOY THE WAVELET TRANSFORMS AND ARTIFICIAL NEURAL NETWORKS FOR EARLY DETECTION AND CLASSIFICATION OF BREAST CANSER BASED ON STATISTICAL FEATURES OF MEDICAL IMAGES

(Ph.D. Thesis)

Karzan Faidhi HAMAD

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

December 2024

ABSTRACT

This dissertation deals with the application of machine learning to diagnose and classify digital images of breast cancer (benign and malignant), utilizing a dataset of 150 digital images. Based on the Artificial Neural Networks (ANN), Wavelet Transformation (WT), and a set of statistical and geometric measures (Mean, Standard Deviation, Entropy, Solidity, Feret….), which are extracted from the shape (benign and malignant) of digital images, through using (t-test and factor analysis), after extracting the data from all the images. Our methodology involves using Artificial Neural Networks (ANN) to classify variables, both before and after applying the Wavelet transform to the dataset. We successfully identified essential variables that influence the diagnosis of various mass shapes. After applying the Wavelet transform, our study showed a significant improvement in the ANN model's performance, with the classification accuracy increasing from (82.2%) to (93.18%), as measured by (kappa and ROC curves). These findings highlight the significance of our research in the field of breast cancer diagnosis and classification. Furthermore, we identify specific measurements, such as (height, entropy, and Feret), primary in shaping the diagnostic process. These metrics emerge as significant determinants influencing the accurate classification of tumor shapes and contributing to improved breast cancer diagnosis. Implementing our research involves utilizing (R, Spss, ImageJ, and MATLAB), which enable comprehensive data processing and analysis.

Science Code	:	20516
Key Words:	:	Machine Learning, artificial neural networks (ANN), discrete wavelet transform (DWT), medical image, ROCcurve,
Page Number	:	121
Supervisor	:	Prof. Dr. Bülent ÇELİK
Co-supervisor	:	Assist. Prof. Dr. Rizgar Maghded AHMED

TIBBİ GÖRÜNTÜLERİN İSTATİSTİKSEL ÖZELLİKLERİNİ KULLANARAK DALGACIK DÖNÜŞÜMÜ VE YAPAY SİNİR AĞLARI YARDIMIYLA MEME KANSERİNİN ERKEN TEŞHİSİ VE SINIFLANDIRILMASI

(Doktora Tezi)

Karzan Faidhi HAMAD

GAZİ ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

Aralık 2024

ÖZET

Bu çalışmada, makine öğrenmesi ile dijital meme kanseri görüntülerinin (iyi huylu ve kötü huylu) teşhis ve sınıflandırılması amaçlanmıştır. Çalışmada 150 dijital görüntüden oluşan bir veri seti kullanılmıştır. Mamografi görüntülerinde ilgilenilen bölgeye ilişkin istatistik ve geometrik ölçümler (Ortalama, Standart Sapma, Entropi, Solidite, Feret gibi) ImageJ yazılımı kullanılarak elde edilmiştir. ImageJ yazılımdan elde edilen ölçümler açımlayıcı faktör analizi ve bağımsız gruplar t-testi ile incelenerek anlamsız özellikler çıkarılmıştır. Önemli özellikler kullanılarak önce yapay sinir ağları ile daha sonra ayrık dalgacık dönüşümü uygularak iyi huylu ve kötü huylu meme kanserinin sınıflandırılması yapılmıştır. Sınıflandırma doğrulukları karmaşıklık matrisi, Kappa istatistiği ve ROC eğrileri ile değerlendirilmiştir. Ayrık dalgacık dönüşümü uygulanmadan önce yapay sinir ağları ile elde edilen sınıflandırma doğruluğu %82,22 iken, ayrık dalgacık dönüşümü uygulandıktan sonra yapay sinir ağlarının sınıflandırma performansı artmış ve sınıflandırma doğruluğu %93,2'ye ulaşmıştır. Ayrıca meme kanserinin sınıflandırlmasında Image J yazılımından elde edilen height, entropi, feret ve Intden özellikleri en önemli özellikler olduğu sonucuna varılmıştır. Bu tez çalışmasında verileri elde etme ve analizinde Image J, R ve MATLAB yazılımları kullanılmıştır.

Bilim Kodu	:	20516
Anahtar Kelimeler	:	Makine öğrenmesi, yapay sinir ağı, ayrık dalgacık dönüşümü, tıbbi görüntü analizi
Sayfa Adedi	:	121
Danışman	:	Prof. Dr. Bülent ÇELİK
İkinci danışman	:	Assist. Prof. Dr. Rizgar Maghded AHMED

ACKNOWLEDGEMENTS

At the prime praise be to ALLAH, the lord of universe, blessing and peace up on Muhammad (God's peace and pray be upon him). I am grateful to some people, who worked hard with me from the beginning till the completion of the present study. In particular, my supervisors, Prof. Dr. Bülent ÇELIK and Ass. Prof. Dr. Rizgar AHMED, for his assistance, ideas, and feedback during the process of doing this study. Without his guidance and support, this dissertation couldn't be completed on time. I deeply and individually appreciate the Prof. Dr. Meral EBEGİL for his great help during my unsupervised period. I would like to extend my heartfelt gratitude to Prof. Dr. Birdal ŞENOĞLU and Prof. Dr. Fikri GÖKPINAR, who served as supervisory members of my PhD committee. Their guidance and support were invaluable in enhancing my thesis. I am also very thankful for the former Dean of college of Administration and Economics Prof. Dr. Ahlam WALI and the head of Statistics and Informatics department Asst. Prof. Dr. Behkal SEDEEQ; for their continued encouragement, support and providing facilities required for this investigation. Thanks are owed to all my colleagues at the Statistics and Informatics Department and especial thanks given to Dr. Taha ALI, Dr. Dlshad M. SALEH, and Alan RAHIM, for their continuous guidance and support. Finally, but by no means least, I am highly grateful to family for almost unbelievable support. They are the most important people in my world, I cannot forget their virtue throughout my life, and I dedicate this dissertation to them.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

Figure **Page**

Figure **Page**

SYMBOLS AND ABBREVIATIONS

The symbols and abbreviations used in this study are presented below, along with their explanations.

Symbols	**Explanations**
db	Daubechies Wavelet
bior	Biorthogonal wavelet
$\boldsymbol{\delta_{uni}}$	Universal Thresholding
$\boldsymbol{\delta_{sur}}$	Sure Thresholding
$\boldsymbol{Wn^{(ST)}}$	Soft Thresholding
$\boldsymbol{Wn^{(HD)}}$	Hard Thresholding

Abbreviations	**Explanations**
ANN	Artificial Neural Network
AUC	Area under Curve
BCET	Balance Contrast Enhancement Techniques
CAD	Computer Aided Diagnosis
CNN	Convolutional Neural Network
DDSM	Digital Database for Screening Mammograph
DWT	Discrete Wavelet Transform
IDWT	Inverse Discrete Wavelet Transform
IntDen	Integrated Density
KMO	Kaiser-Meyer-Olkin
KNN	K-Nearest Neighbors
LD	Logistic Discriminant
LRA	Logistic Regression Analysis
MRI	Magnetic Resonance Imaging
OCR	Optical Character Recognition
PNN	Probabilistic Neural Network
ROC curve	Receiver Operating Characteristic Curve
ROI	Regions Of Interest

Abbreviations	Explanations
SVM	Support Vector Machine
TIFF	Tag Image File Format

1. INTRODUCTION

The human eye has several nerves connected to the brain that send signals to the brain simply by looking at an image to distinguish its elements. These elements are divided into multiple regions, which differences in color, shape, or texture can distinguish. The development in the computer world has led to an attempt to incorporate this human ability into software and mechanisms that can be obtained via a computer. The ability of a computer to distinguish certain areas within an image has attracted the interest of many researchers, as this field is of great benefit in all specializations. Describing an image, extracting information, and distinguishing patterns are among the wide applications in the field of digital images. Image processing is a vital scientific field, and it has received great success and interest in various applications and different fields, including medical, military, industrial, communications, and space (Jain, 1989).

Digital image processing is often misunderstood as a process of solely decorating and organizing images to produce a different outcome. However, this is far from the truth. The focus of digital image processing is to extract essential features and measurements from the image, which can be used and analysed by computers. This includes statistical and geometric measurements. Explains that digital image processing is far more complex than just modifying the appearance of an image (Ammar, 1992).

When a computer reads an image, it divides the screen into a group of tiny squares called "pixels." These pixels are the elements that make up a digital image. The values for each pixel are saved in such a way that they have their shade or color. This process is known as pixelation. The term pixel comes from the words "picture element" (Gonzalez and Woods, 2018).

One of the important techniques that researchers use in the digital image field is the ability of computers to distinguish the region of interest (ROI) in an image. This technique is widely used for various purposes, such as feature extraction, information retrieval, and pattern recognition.

Generally, techniques of digital image processing are concerned with the following:

- Improvement of pictorial information and its analysis by humans.
- Analyzing images for feature extraction for understanding and interpreting (Al-Asmari, 2010).

Medical imaging plays a vital role in medicine, particularly in non-invasive and clinical studies. Analysis and diagnosis of medical images help radiologists and physicians to make accurate diagnoses.

Breast cancer is still one of the biggest health challenges faced globally and affects millions of women annually. Detecting it accurately and on time is crucial for successful treatment and improved prognosis. In recent years, advancements in medical imaging technologies have revolutionized the early detection and diagnosis of breast cancer. These technologies offer a non-invasive means of examining internal structures with unprecedented detail and precision. However, the sheer volume and complexity of medical images generated present significant challenges for effective analysis and interpretation.

In response to the challenge of early detection and classification of breast cancer from medical images, the integration of Wavelet transforms, and Artificial Neural Networks (ANNs) has emerged as a promising approach. Wavelet transforms offer an efficient means to extract relevant features from medical images, owing to their multi-resolution analysis capabilities. They capture both spatial and frequency information at various scales. Artificial neural networks, inspired by the functioning of the human brain, are well-suited for image classification tasks due to their ability to learn complex patterns and relationships within data.

Breast cancer is a common disease among women, and there have been many studies on it. These studies have focused on laboratory examination and biopsy, but mammogram image analysis needs to be given more attention, particularly in investigating various shapes of mass. This research takes masses, both malignant and benign, and uses digital features, discrete wavelet transformation, statistical measurements, and artificial neural networks to achieve the best classification that distinguishes those masses.

The aim of this thesis is to utilize Wavelet transforms and Artificial Neural Networks to detect and classify breast cancer at an early stage, using statistical features extracted from

medical images. By combining the strengths of these two computational techniques, this research aims to improve the accuracy and speed of breast cancer diagnosis and recommend a classification method that depends on Wavelet transformation.

The thesis is divided into five chapters. The first chapter consists of an introduction with aim of the thesis, second chapter literature review. Chapter three is divided into four sections: Section one explains the Anatomy of the Breast, section two covers Digital Image Processing, section three explains Artificial Neural Network, and Section Four covers Wavelet Transformation. Chapter four focuses on the application of an Artificial Neural Network and Wavelet Transformation of mammogram Images (Malignant and Benign) mass. It uses testing of hypothesis and factor analysis to select significant parameters. Lastly, chapter five presents the main results, conclusions, and recommendations of the entire thesis.

2. LITERATURE REVIEW

Breast cancer is a significant health issue worldwide, and detecting it early is essential for better patient outcomes. In the last few decades, artificial neural networks (ANNs) and image processing have become a valuable tool in breast cancer detection. ANNs have the potential to improve diagnostic accuracy and make clinical workflows more efficient. This literature review gives an overview of essential studies and advancements in the use of ANNs, image processing, and wavelet transformation for breast cancer detection.

Furundzic et al. (1998) In this paper, present an Artificial Neural Network (ANN) based model that effectively identifies women at high risk of acquiring breast cancer. The study proposes the use of ANN on a case-controlled basis, which resulted in excellent outcomes for early detection. The model achieved accurate classifications while reducing adverse outcomes (Furundzic et al., 1998).

Wandre et al., (2020) studied breast masses by using an Artificial Neural Network (ANN) to classify the characteristics (shape and texture) of both types (malignant and benign). As a result, both species were correctly classified at 81% (Wandre et al., 2020).

Abdolmaleki et al. (2001) developed a neural network system for extracting and analyzing quantitative data from time-intensity profiles, which is used to predict biopsy results for various breast lesion disorders. The study compared the performance of radiologists to that of the Artificial Neural Network (ANN) in terms of sensitivity, specificity, and accuracy. The results showed that the ANN was able to diagnose with 87.9% accuracy and classify 107 out of 120 cases correctly, which is a significant improvement compared to radiologists, who had an accuracy rate of 74% (Abdolmaleki et al., 2001).

Abdolmaleki et al. (2004) developed two models to classify breast cancer into benign and malignant types. The first model was a non-algorithmic Artificial Neural Network (ANN), while the second was an algorithmic logistic regression analysis (LRA). The researchers extracted six features from MRIs of 161 patients and used them to recommend biopsy results using both models. Both models were then coded to predict biopsy results. The findings showed a correlation between the features and biopsy results. However, LRA outperformed

ANN with excellent specificity and sensitivity despite some statistically significant variables (Abdolmaleki et al., 2004).

Abdolmaleki et al. (2004) used logistic discriminant (LD) to differentiate breast cancer based on ultrasonic parameters. They found that 67 out of 72 cases were correctly classified, achieving an accuracy of 93% (Abdolmaleki et al., 2004).

Rashed and Awad (2006). A diagnostic system and model were proposed for detecting breast tumors in digital mammograms. The system uses discrete wavelet transformation (DWT) in conjunction with Artificial Neural Network (ANN) classification to achieve early detection of tumors. The proposed diagnostic system has been found to classify mammograms effectively (Rashed and Awad, 2006).

Al-Naami et al. (2009) demonstrated that it is possible to differentiate between (malignant and benign) brain cancer without the need for a biopsy. They achieved this by utilizing a method of image analysis and statistical measures on MRI images of brain cancer. Through statistical analysis, they took measures such as the (mean, range, and box plot...). By comparing the statistical measurements for both benign and malignant tumors using a t-test, they discovered a significant difference between the two types of tumors (Al-Naami et al., 2009).

Ayer et al. (2010) conducted a study to compare the effectiveness of logistic regression (LR) and artificial neural network (ANN) models for mammography-based breast cancer risk estimation. They found that each model had its strengths and that the Logistic Regression Model (LRM) is generalized by Artificial Neural Network (ANN). It was also determined that neither model can replace the other in medical diagnosis, but they can be used together to complement each other (Ayer et al., 2010).

Surendiran and, Vadivel (2010). Presented a paper that demonstrated the extraction of tumour type-specific features from mammography images using multivariate analysis methods. Their study also showed how to differentiate between (malignant and benign) tumours based on Stepwise ANOVA Discriminant analysis, achieving a classification accuracy of 87%, which was higher than that of other existing methods (Surendiran and Vadivel., 2010).

Ragab et al. (2013). the Support Vector Machine (SVM) technique and Artificial Neural Network (ANN) were compared for detecting benign and malignant breast cancers. The study utilized discrete wavelet transform (DWT) as a feature extraction technique for detecting breast, CT, and PCA cancers. The study concluded that using an Artificial Neural Network (ANN) as a classifier is superior to a Support Vector Machine (SVM) in detecting breast cancer (benign and malignant) (Ragab et al., 2013).

Muhammad and Sami (2014) conducted a study on classifying bone images of fractures and tumors by using Discrete Wavelet Transform (DWT) and digital image processing techniques. They employed a set of statistical processes and an Artificial Neural Network (ANN) to achieve this. The study involved analyzing 125 samples, and the neural network was able to classify the images with 100% efficiency (Muhammad and Sami 2014).

Vijayarani et al. (2015). In a study, four types of kidney diseases were diagnosed, namely chronic glomerulonephritis, renal failure, chronic kidney disease, and acute nephritic syndrome. The study aimed to predict kidney disease using two algorithms: Artificial Neural Network (ANN) and Support Vector Machine (SVM). The accuracy and execution time of the two algorithms were compared, and the results showed that Artificial Neural Network (ANN) classification is a better and more appropriate method than Support Vector Machine (SVM) algorithms. Moreover, ANN classification takes less time for classification (Vijayarani et al., 2015).

Beura et al (2015). Breast tumors were classified into normal, malignant, and benign categories using the DDSM and MIAS standard databases. All mammograms were analysed using custom matrix forms for GLCM and 2D-DWT and were subjected to both F-test and T-test. The results showed that the T-test had better accuracy than the F-test. For breast cancer (normal and abnormal) and (benign and malignant) in the MIAS database, accuracy rates of 98% and 94.2% were achieved, respectively. In contrast, for DDSM characteristics, accuracy rates of 98.8% and 97.4% were obtained (Beura et al., 2015).

Sehrawat et al. (2017) proposed a technique for removing noise from images through the use of a Gaussian filter. They then employed support vector machine (SVM) and discrete wavelet transformation (DWT) methods to detect and classify breast cancers (benign and malignant).

Their findings revealed that the SVM algorithm outperformed other techniques in terms of performance and quality. (Sehrawat et al., 2017).

Hamad and Naeem (2018). They proposed a system for accurately detecting breast cancer by using MRI images of breast tumors (normal, benign, and malignant). First, they applied a discrete wavelet transformation (DWT) to remove noise and improve image quality. Then, they developed a probabilistic neural network (PNN) that was able to classify all three types of breast tumors with 90% accuracy (Hamad et al., 2018).

Tahooresi et al. (2018) proposed a hybrid model based on machine learning for early detection of breast cancer. They compared three classification techniques (ANN, SVM, and KNN) using image and blood datasets to determine their effectiveness in diagnosing breast cancer. The study found that SVM, either alone or in combination with blood and imaging data, is the most effective technique for diagnosing cancer, achieving an accuracy rate of 99.8% (Tahooresi et al., 2018).

Sepandi et al. (2018). They conducted a study that used Fine Needle Aspiration Biopsy (FNAB) and imaging techniques to identify cellular diseases. The objective of the study was to develop a model to assist radiologists. The outcomes of applying an Artificial Neural Network (ANN) to a set of breast imaging data showed a sensitivity and specificity of 0.90 and 0.82, respectively. Additionally, the negative and positive predictive values were 0.90 and 0.80, respectively (Sepandi et al., 2018).

In their study, Wadkar et al. (2019) focused on using SVM and ANN algorithms to detect breast tumors. They concentrated on one specific model and evaluated its reliability as well as the duration of the detection process. The outcomes showed that the Artificial Neural Network (ANN) classification rate was 97%, while the support vector machine (SVM) classification rate was 91%. This indicates that the ANN algorithm is much more effective and accurate than the SVM algorithm (Wadkar et al., 2019).

In a study conducted by Jannat et al. in 2019, the researchers utilized wavelet transform (WT) and artificial neural network (ANN) to diagnose heart palpitations that arise due to cardiac rhythm problems. The study aimed to detect heart murmurs resulting from the appearance of mild vibrations in the heart (Jannat et al., 2019).

Das et al. (2020) Introduced a standardized system for detecting breast cancer. They used various techniques to classify problems in medical diagnosis and gathered and classified 75 mammography images into three categories: (normal, malignant, and benign). The main objective was to use a neural network to detect breast cancer by removing noise from each image and obtaining better features for accurate diagnosis using balance contrast enhancement techniques (BCET). To achieve this, they applied discrete wavelet transformation (DWT) to remove image noise and used Artificial Neural networks (ANN) to classify the images. Their findings showed that the system could accurately classify images as (normal, malignant, or benign) in 95% of cases, enabling doctors to detect mammograms earlier (Das et al., 2020).

Mohammed et al. (2020) conducted a study on classifying open mammography images using linear discriminant analysis (LDA). They used 85% of the dataset for training and 15% for testing. The suggested strategy resulted in an 81% accuracy rate. (Mohammed et al., 2020).

Khandezamin et al. (2020) developed a new technique for diagnosing breast cancer, which involved using three sets of data (WPBC, WBCD, and WDBC) from both (benign and malignant) cases. The procedure had two steps:

Step 1: Starting with logistic regression (LR)

Step 2: Using the (GMDH) Neural Network.

The technique had impressive results, with accuracy rates of 99.4% for WBCD, 99.6% for WDBC, and 96.9% for WPBC. (Khandezamin et al., 2020).

Abdul Hameed et al. (2021) used ANN and LR methods to predict breast tumors with (WDBC). ANN outperformed LR in effectiveness, sensitivity, and specificity (Abdul Hameed et al., 2021).

3. RESEARCH DESIGN AND METHODOLOGICAL FRAMEWORK

In this chapter firstly theoretical aspects of breast cancer (Definition of the breast, define breast cancer, Breast cancer statistics, Breast Cancer Diagnosis... etc.) have been explained.

Secondly tackled with theoretical aspects of digital image processing in general (Some basic concepts, A Simple Image Formation Model, Sampling and Quantizatio.etc.).

Thirdly tackled with theoretical aspects of Artificial Neural Networks (An Overview of Artificial Neural Networks Development, Components of the artificial neural network and biological, Types of Activation Functions, ….Etc.) and has dealt with Wavelet Transformation (Wavelet shrinkage, Wavelet Properties, Some of Wavelet Function, Etc.)

3.1. Definition of the Breast

Each breast contains about (15-20) lobes, and each lobe has small lobes ending with sacs that produce milk. All these lobes are connected to channels. These channels reach the nipple in the middle of a dark area in the breast. It is known that a quantity of fat fills the spaces around the lobes and channels, and there are no muscles in the breasts. However, muscles below the chest cover the ribs, and each breast contains blood vessels and lymphatic vessels that lead to lymph nodes. Moreover, these glands are found in groups under the armpit and chest. This is shown in illustration (3.1) for breast components (Carol, 2005).

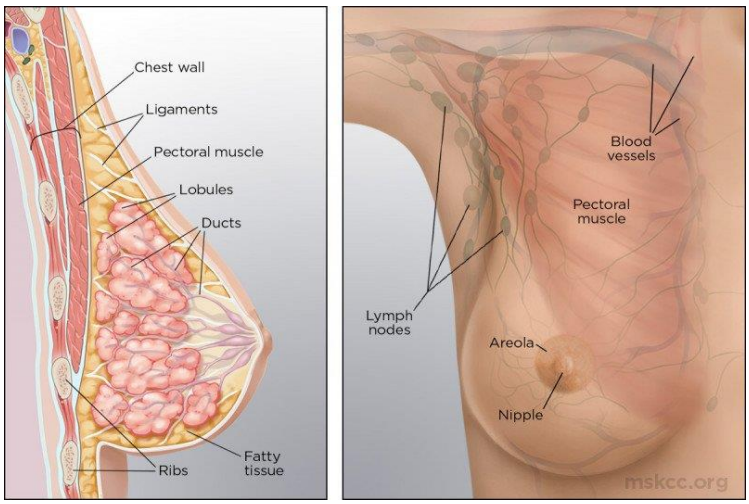

Figure 3.1. Anatomy of the Breast

3.2. Breast Cancer

Define breast cancer

Breast cancer, in general, is a type of disease that causes the affected cells to grow, change, and multiply out of control and gives cancer the name of the part from which it started. Breast cancer means the irregular growth and spread of cells originating in the breast tissue. It multiplies rapidly and can form a piece or a clump of extra tissue. Moreover, tissue masses are called tumours. Tumours are either cancerous (malignant) or noncancerous (benign). Malignant tumours proliferate and destroy healthy tissues of the body. Moreover, it is an abnormal growth of cells lining the milk ducts or lobes of the breast. Breast cancer is one of the common malignant tumours, and the tumour often forms in the milk ducts, sometimes in the lobes, and a small part in the rest of the tissues (Carol, 2005).

Breast cancer statistics

Breast cancer is the most familiar malignant tumour among women and the second cause of death from cancer among females. Based on data from the World Health Organization, breast cancer is the most typical malignant tumour among women (Jemal et al., 2010).

According to data from the World Health Organization, breast cancer is the According to the (WHO), breast cancer is the most common cancer, with more than 2.2 million cases in 2020 recorded. About one in twelve women will develop breast cancer in their lifetime. In addition, it is the number one cause of death from breast cancer among women, with approximately 685,000 women dying from it in 2020 (WHO, 2021).

Statistics in 2006 indicate that every (3 minutes), a woman is diagnosed with breast cancer. Moreover, every (13 minutes) one of them dies worldwide (Cancer.org, 2006).

According to estimates, in 2010, breast cancer affected more than (200,000) women in the United States of America, and about (40,000) women died because of this disease (Rizgar, 2013).

Some factors that are associated with an increased risk of breast cancer include

The causes that lead to breast cancer are not known to most patients. Several factors increase the chances of the disease appearing, including:

1. Age**:** There is an increase in the risk of breast cancer with age. Breast cancer is the leading reason of death among women between the ages of 25-50 (DeVita et al., 2023; Jemal et al., 2010).
2. Genetic factors**:** Breast cancer is heavily influenced by genetics. If a woman has a first-degree relative (such as a sister or mother) who has been diagnosed with breast cancer, the chances of her developing the disease rise by 1.5 to 3 times. Genetic mutations in two essential genes, namely (BRCA1 and BRCA2), significantly increase the risk of acquiring breast cancer. Additionally, these mutations also lead to a higher likelihood of ovarian cancer. These types of tumors often develop at an early age.
3. Menstrual cycle: The number of menstrual cycles a woman experiences affects her breast cancer risk. Moreover, The decrease in the number of menstrual cycles caused by starting menstruation at a later age, ending it at an earlier age, or being pregnant can reduce the risk of breast cancer.
4. Lactation: Women who breastfeed their children have a lower risk incidence of breast cancer than those who do not breastfeed their children, and the risk of breast cancer decreases with the extension of the lactation period (Tryggvadottir et al., 2001).
5. Excess obesity: Excessive obesity is a risk factor for breast cancer. Regular exercise reduces the risk. (Calle et al., 2003; Thune et al., 1997).
6. Radiation**:** Women who receive radiation to the chest area for cancer at a young age have a higher risk of developing breast cancer.

Breast cancer symptoms

1. Lumps under the armpit or neck.
2. Nipple changes.
3. Skin texture change.
4. Skin ulcers.
5. Metastasis symptoms like bone pain.

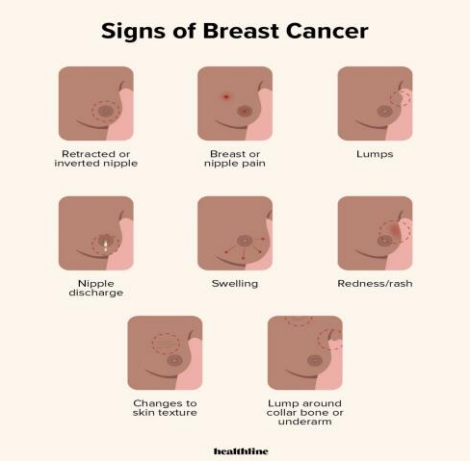

Figure 3.2. Explains the symptoms of breast cancer

Breast cancer diagnosis

When specific symptoms suggest the presence of breast cancer, doctors may perform various tests to confirm the diagnosis. If a mammogram shows an abnormality, several tests can be used to diagnose breast cancer (American Cancer Society, 2022).

Here are some of the most important ones:

a) Diagnostic mammograms

This unique tool is used for detecting and diagnosing early breast tumors. They use X-rays for imaging and can also assist in taking samples from these tumors. In many developed countries, there is a trend to use this type of imaging routinely, particularly for middle-aged women, to aid in the early detection of breast cancer (Dove, 2004).

Figure 3.3. Mammograms machine

b) Ultrasound of the breast

When a lump appears in the breast, ultrasound waves can be used to determine whether it is a solid mass or a fluid-filled sac. This imaging method uses sound waves to capture detailed images of the breast's structure, helping doctors identify the type of lump present (Mayo Clinic, 2022).

Figure 3.4. Ultrasound machine

c) Magnetic resonance imaging (MRI)

This type utilizes a magnetic field to obtain prominent images of the part to be imaged in any direction. It is an effective diagnostic tool in cases where breast cancer is suspected (Al-Nuaimi, 2006).

To diagnose breast cancer in some cases (MRI) is used:

- When the density of the breast tissue is high, which usually occurs in young women. In this case, it isn't easy to check the breasts using mammography. MRI can be utilized to detect any tumours in the breast.

- When checking for a swollen lymph gland under the armpit that contains suspicious cancerous cells, doctors cannot locate the primary breast tumour through other tests.

Figure 3.5. MRI machine

d) Breast biopsy

When a doctor suspects that a patient could have breast cancer, they may perform a breast biopsy. This procedure involves taking a sample of breast tissue under a microscope to check for the existence of cancerous cells. A biopsy is usually done after a physical exam, mammogram, or other imaging tests have shown changes in the breast that may be cancerous. This is the most reliable way for doctors to diagnose breast cancer.

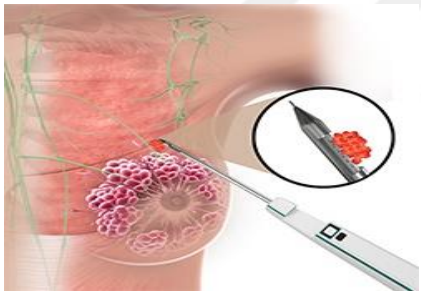

Figure 3.6. Biopsy

Stages of breast cancer

When diagnosing cancer, the specialist and doctor must determine the stage at which the cancer has progressed. This diagnosis is essential to determine whether cancer has spread and affected other organs or has remained confined to a specific area. The diagnosis of the disease and its type depends on this determination.

Breast cancer has different stages, which determine the extent of the cancer's spread. Early stages are localized or invasive, and cancer can be removed, with either the breast preserved or removed entirely. Cancer progresses through five different stages (Rosenthal, 2000):

Stage 0: is the earliest stage, and it refers to non-invasive cancer that doesn't invade neighbouring cells.

Two types of tumors exist at this stage:

1. Ductal carcinoma in situ
2. Lobular carcinoma in situ

Stage I: is an early stage where the cancer hasn't spread beyond the breast and nearby tissues.

Stage II: is also an early stage where the cancer may affect nearby tissues. However, cancer may spread to the lymph nodes under the armpit.

Stage III: is an advanced stage where cancer has flared to the lymph nodes under the armpit and other tissues adjacent to the breast. The size of the tumour increases to 5cm, but it hasn't spread to other parts or organs. The recovery rate at this stage is medium.

Stage IV: is the metastatic stage; breast cancer metastasizes to other organs, including bones, lungs, liver, and brain (Rosenthal, 2000).

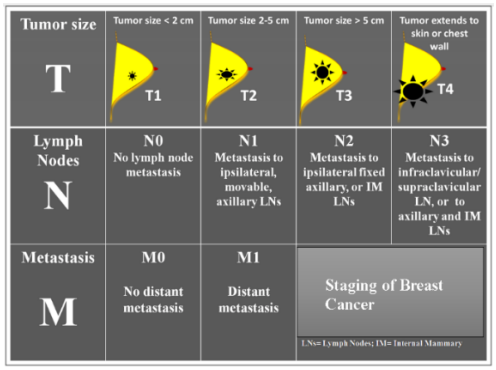

Figure 3.7. Stages of breast cancer

<u>Some factors that can potentially lower the risk of developing breast cancer (Leon,1989)</u>

1- Exercise: Engaging in physical activity for more than 4 hours per week can lower the risk of breast cancer.
2- Early pregnancy: Women who have their first pregnancy before the age of 20 have a lower risk of developing breast cancer.

3- Breastfeeding: Women who breastfeed their children have a higher likelihood of staying healthy and not developing breast cancer.

Methods of treatment

Breast cancer treatment options depend on aspects such as the size and location of the tumour, laboratory research results, the age and overall health of the patient, and whether the lymph nodes under the armpit are affected. The size of the breast and the extent of the tumour's spread are also considered. Treatment for breast cancer generally involves a combination of methods. If the tumour is small, less than 3 cm in size, and detected early, surgery to remove the breast may not be necessary. Instead, the tumour itself can be removed, and the rest of the breast can be treated with radiation therapy. If the tumour is larger or has spread to the lymph nodes, chemotherapy, radiation therapy, and hormonal therapy may also be used. The following methods are commonly used to treat breast cancer (Abeloff et al., 2008; Roses, 1999):

a) Surgery
b) Chemotherapy
c) Radiation Therapy
d) Hormonal Therapy

3.3. Computer Imaging

Dealing with digital images means dealing with a tremendous amount of information, which calls us to develop the actual fields in computer imaging, such as (image compression) or (image segmentation), whether the human being is the receiver or the computer. This difference in the recipient calls us to divide computer imaging into two essential categories. They are:

- Computer Vision
- Image processing

Moreover, as shown in the diagram below:

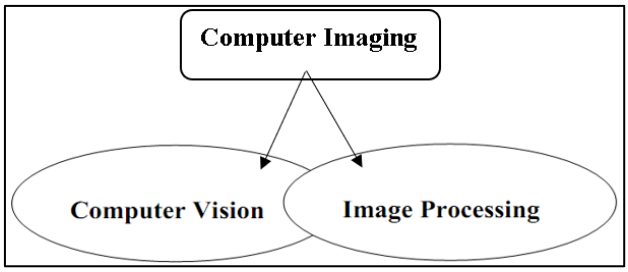

Figure 3.8. Computer imaging

As image processing takes the image as input and produces images as output after making improvements and developments on it, and computer graphics relate to entering information, describing it, and creating images as outputs that are images for the recipient, and (output) are images (input) such as engineering drawings and maps, i.e., human beings.

Pattern recognition and computer vision use images as inputs and give descriptive information in the device or (output) are images and (input) as outputs such as the number and type of cells under the microscope, the computer (Murugaraja and Balamurali, 2014; Scott, 1998).

3.4. Digital Image Processing

3.4.1. Basic concepts in digital image processing

Image

An image is a depiction, photograph, display, or any other form that visualizes an object or a scene. The image in the digital image processing process has another meaning, as it means that it is a $\text{two} - \text{dimensional}$ function $f(x, y)$ where (x, y) are plane coordinates, and the the image at that point. as displayed in the figure (5.9), (Al-Asmari, 2010; Ammar,1992; Gonzales and Woods, 2018).

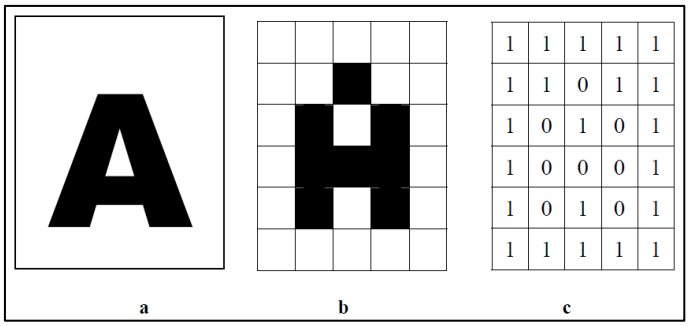

Figure 3.9. Example of a binary digital image

a) Representing the letter, A as an image.
b) The letter A is a group of pixels.
c) The grey level values that represent their corresponding screen points

Pixel

Every tiny space occupied by a number in a digital image is called an image point (which is the smallest unit in a digital image), and this point indicates the spatial coordinates and the size of space occupied by the image point is called the precision spatial component of the image (Resolution Pixel Spatial).

Every tiny space occupied by a number in the digital image is called the pixel point (the smallest unit in the digital image). This point indicates the spatial coordinate. The size of the space occupied by the pixel point is called the pixel spatial resolution from (nanometers - a microscopic unit of measurement) in microscope images to tens of kilometers in satellite images. Every small square of the matrix that forms the digital image contains essential information that determines the value of the grey level at that point; such a matrix is named a picture element (Picture element, Image element, Pixel).

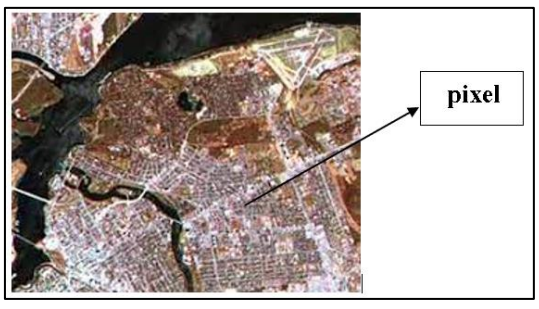

Figure 3.10. Example of a pixel

Band

There may be more scenes related to the image. For example, when dealing with the colours, the scene may be our three compounds' images(R-G-B) (Red - Green - Blue), spectral scanning devices mounted on aircraft and satellites give 3 – 30 images, and all the pictures together make up an image called the image. Moreover, every single image is called a Band.

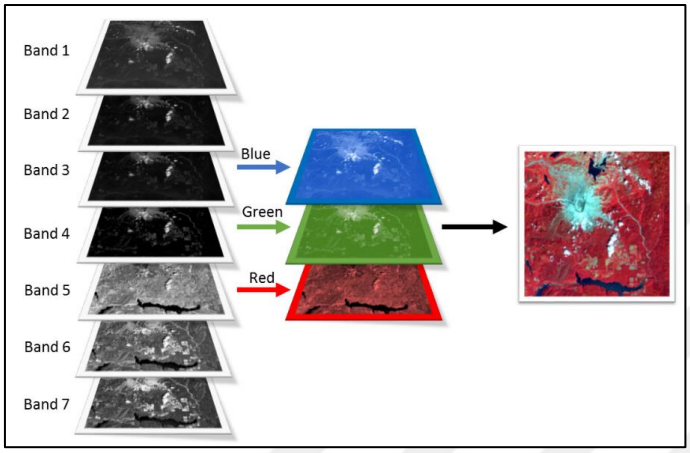

Figure 3.11. Example of a band

Image transformation

Transform theory has played an essential and fundamental role in digital image processing. It is a point of interest in theoretical work and practical applications in image enhancement, restoration, and description.

The term (digital image conversion) means changing the (original) digital data values of the image units with (new) values that help in the interpretation and description of the image (Pearson, 1991).

Luminance

It is the sensitivity to a wide range of changes in lighting. A person is sensitive to 14 logarithmic units of lighting levels (14 levels, per level, is ten degrees more elevated than the previous one), ranging from dark to bright lighting (Levkowitz, 1997). Figure (3.12) illustrates this.

There is a direct relationship between the lustre and illumination of objects. The greater the lustre reflected from the body, the greater the illumination in the form of a logarithmic function. This property is used to improve images taken with dark or bright lighting, as the percentage of the white colour is increased or decreased to obtain an image with more unique features (Gomes, 1997; Scott, 1998).

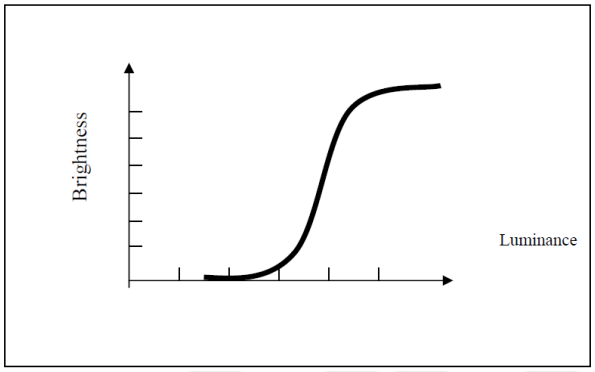

Figure 3.12. Human eye vision algorithm

Contrast

Contrast is defined as the ratio between the object luminance and the background luminance on which the objects are located, where:

$$C = \frac{\text{Object Luminance}}{\text{Background Luminance}} \tag{3.1}$$

Contrast sensitivity relies on the spatial distribution of light and dark areas in the image, and images can be improved using this feature. A filter can enhance the contrast ratio in the image by subtracting a certain percentage from the value of each colour unit to increase the contrast between image points (Scott, 1998).

This feature (contrast) can be taken advantage of by compressing and encoding images by selecting the most frequency data and allocating it to the area most sensitive to the spatial frequency of this impurity (Gomes, 1997).

Texture enhancement

Texture is essential spatial information in digital image processing applications. The contrast in gray gradations per unit area of the image expresses it. The texture is not said to be smooth or homogeneous when the contrast is low. In contrast, the texture is rough when the contrast is large enough to distinguish phenomena. Its formation and characteristics of spatial phenomena include the size of the phenomenon, the shadow, and the image's grayscale (Lillesand and Kiefer, 2000).

3.4.2. A simple image formation model

An image is formed by the reflection of light from objects. As we mentioned earlier, images can be represented by a two-dimensional function $f(x, y)$. The value of this function at any two coordinates (x, y) is a positive value whose physical meaning is determined according to the source of the image. In monochrome images, the function values correspond to grayscale values. It is important to note that when an image is generated from a specific physical process, the image values are proportional to the energy radiated from a natural source, such as electromagnetic waves. Therefore, $f(x, y)$ must take non-zero and finite values, as light is a form of energy, and the function $f(x, y)$ must not be zero or limited, that is

$$0 < f(x, y) < \infty \tag{3.2}$$

These components are called illumination (i) and reflectance (r) and are denoted by $i(x, y)$ and $r(x, y)$. Moreover, the product of two functions combines results in $f(x, y)$:

$$f(x, y) = i(x, y)\, r(x, y) \tag{3.3}$$

Where

$$0 < i(x, y) < \infty \quad and \quad 0 < r(x, y) < 1$$

The grey level value at point (x, y) differs from $(x + 1, y + 1)$, symbolized by (G).

$$G_{max} \geq G \geq G_{min}$$

$$r_{max} \times i_{max} \geq G \geq r_{min} \times i_{min}$$

$$Level\ 255 \geq G \geq 0$$

(Gonzales & Woods, 2018)

3.4.3. Sampling and quantization

For the image to be suitable for computer processing, this image must be represented by data commensurate with the contents of these images, as the images are digitized in each of the spatial, and this is called the digitization of the image samples and amplitudes, and this is called the gray level quantization, as shown in Figure (3.13).

The sampling process examines the density of continuous images in specific locations, and the quantization process determines the digital illumination of each column from black through gray to white. The column is the sample image, and quantization is the image unit because it represents the discrete digital component of the digital image (Al-Asmari, 2010; Gonzales and Woods, 2018).

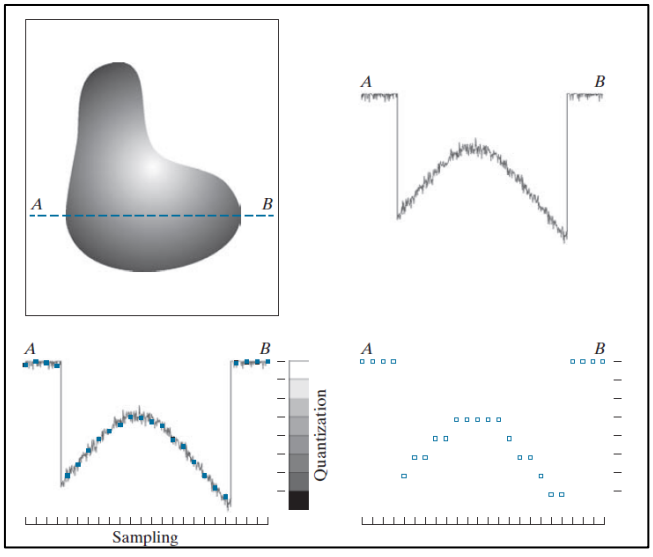

Figure 3.13. Sampling and quantization in image processing

3.4.4. Digital image representation

After the sampling and quantization process, the image can be displayed by a matrix of real numbers. Suppose we assume that the number of rows is (M) and columns are (N), then instead of using the actual values for each of (x, y). We can use consecutive integers (0→M - 1, 0→N - 1). Because the image is now just a matrix and assuming that the first point is (x, y) = (0,0), we can now write the image function as follows: (Ansari and Borse, 2013; Gonzales Woods, 2018).

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) \ldots\ldots & f(0,N-1) \\ f(1,0) & f(1,1) \ldots\ldots & f(1,N-1) \\ \vdots & \vdots & \vdots \\ f(M-1,0) & f(M-1,1) \ldots\ldots & f(M-1,N-1) \end{bmatrix} \quad (3.4)$$

Each matrix element is called (image element, pixel, or picture element). Now, we can also express the gray levels in integer numbers. In that case, the digital image becomes a two-dimensional function, and each of its coordinates and values is an integer number, as:

$M, N > 0$, $\mathrm{G}\,(\text{number of gray levels}) \geq 0$

One of the essential things that we should know is the number of pixels that we must represent the image with and the number of grey levels for each of the units of this image to know the space that this image needs for storage. It is common in the digital image processing system. It is possible to know the gray level that we need in each image unit through the following equation:

$$\mathrm{G} = 2^{\mathrm{K}} \quad (3.5)$$

Where:

G: Gray level of the imaging unit.

k: the number of bits representing the grey level of the imaging unit.

The imaging unit's lowest number can take zero; the highest number depends on how that number is stored, depending on the color system. For example, when we hold each image

element in the form of one bit, that means that it only takes the values (1,0) or black and white or that we store each image element in the form of one byte equal to 8 bits that the highest importance that the image point can take is 255. Other formats are also possible and depend on the architecture of the computer itself, such as (16-bit, 24-bit, 32-bit, or Others) (Gonzales and Woods, 2018).

This can be seen in Table 3.1:

Table 3.1. Digital file data values

2^K	Number of bits	Range of numerical numbers	G
2^6	6	0-63	64
2^7	7	0-127	128
2^8	8	0-255	256

If we have an image that contains (255 * 255) image units and is represented by (128) gray levels, then what is required to store this image is:

$G = 2^K$

$128 = 2^K$

$k = 7 \rightarrow ln(128) = k\ ln\ (128)$

$b = 512 * 512 * 7 = 1835008\ bits$

That is, we need (229376 bytes) to store this image (1 byte = 8 bits)

3.4.5. Threshold based segmentation

Thresholding is a type of image segmentation. We change the pixels to make the image easier to analyze, converting the images from colour or gray to simple black-and-white images (Al-amri, S et al., 2010).

Figure 3.14. Threshold

3.4.6. Types of digital image processing

Two-dimensional images are represented on the computer by zero and one (0,1). Every digital image comprises a group of pixels. Each image is a matrix containing rows and columns of pixels. The more pixels, the more precise the image.

The type of images is:

Binary images

The most basic and simplest image is a binary image, which only has two possible values for each pixel, commonly black and white, or (0) and (1). Because a single binary digit can represent each pixel in a binary image, they are also known as 1-bit/pixel images. When generic forms or outlines are needed to complete the task, computer vision applications are where these images are most frequently used. For example, grasp an object using a robotic gripper or Optical Character Recognition (OCR). Grayscale images are frequently transformed into binary images using a threshold value that turns white images one (1) and black images zero (0) (Umbaugh, 2010).

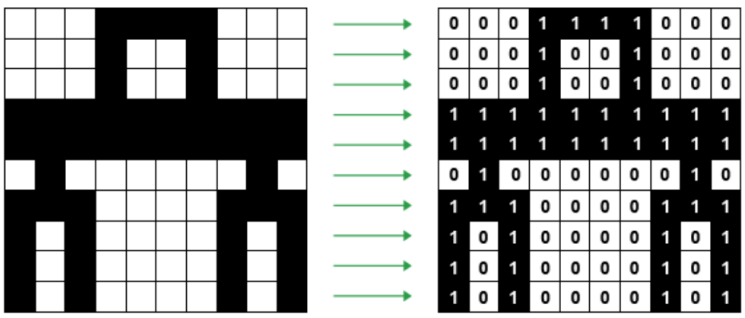

Figure 3.15. Binary images

Gray level images

These images appeared after the emergence of binary images, which are also monochrome. However, they consist of grayscale that only gives information about the illumination intensity. There is no colour information in it, and the number of cells used (Bits) to represent each image unit shows the number of levels of illumination intensity and the common images used (8 Bits). It can display (256) colour gradations, and although these images are not coloured, they are still used in many applications. Binary images are frequently grayscale representations utilizing a designated threshold value. In grayscale imagery, each pixel receives a value of (1) white if it surpasses the threshold value. Conversely, pixels falling below the threshold are assigned a value of (0), representing black. (Abdul Karim, 2010; Kumar and Verma, 2010).

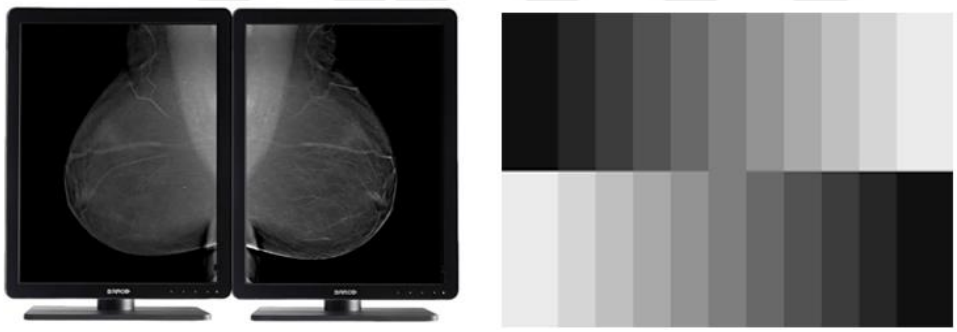

Figure 3.16. Grayscale image

Colour images

Three-band monochrome picture data can simulate a colour image, with each band representing a different colour. Each spectral band's brightness information is preserved in the digital image information. After the appearance of color images with time, the color display increased. The number of color gradations that can be displayed for each color increased. However, the representation of the images remained the same, as each image unit consists of the collection of the three primary colors (red - green - blue) and combining the lighting intensity of these three colours. The desired color (RGB) is obtained. For example, the imaging unit (x, y) gives red if its value is (100, 60, 20) and if its value is (15, 90, 30), it will give the colour green. The nominal unit is blue if its value is (7, 25,230), just as the small team is displayed in black if its value is (0,0,0), and it is shown in white if its value is

(255, 255, 255). Therefore, this type of image is known as (a 24-bit color image), and Figures (3.17) and (3.18) illustrate this.

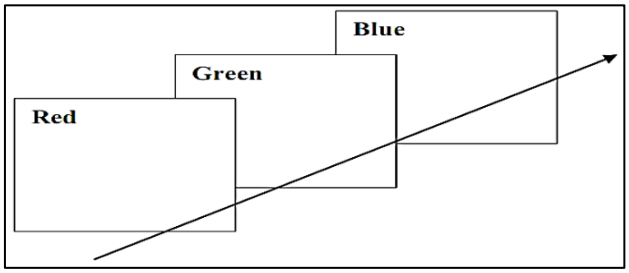

Figure 3.17. RGB vector of image unit at the location (x, y)

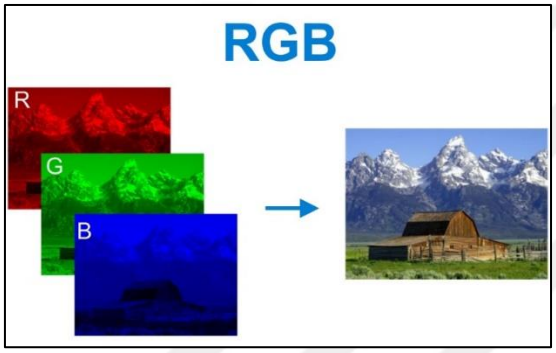

Figure 3.18. Color image

Multispectral images

These images contain information outside the limits of ordinary human vision, including infrared, ultraviolet, X-rays, radar data, and others.

This type of image cannot be sensed typically because the data represented in it cannot be seen directly from the human vision system. The sources of these images are satellite systems, various types of airborne radars, medical diagnostic imaging systems, infrared imaging systems, and others.

As shown in Figure (3.19), which is a satellite image.

Figure 3.19. Multispectral images

3.4.7. Basic image processing methods

Figure (3.20) illustrates the basic operations that take place on digital images, which are not limited to those whose input and output are images but extend to those whose input and output are images and characteristics extracted from the image (Gonzales and Woods, 2018; Acharya and Ajoy, 2005).

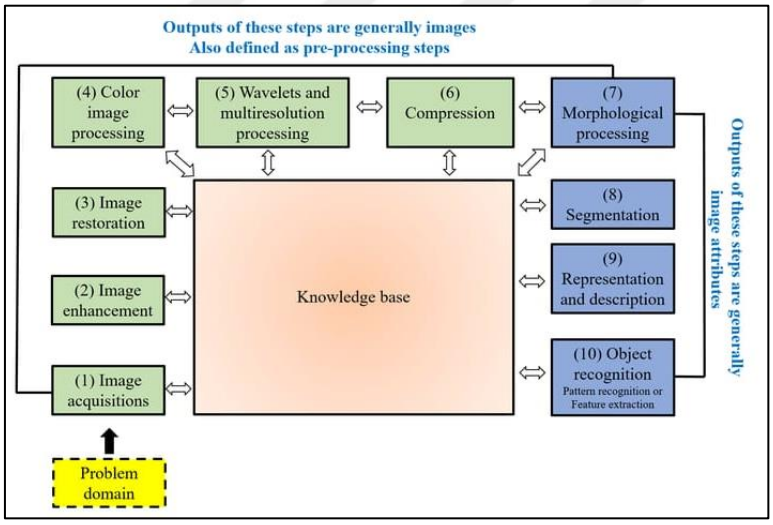

Figure 3.20. Fundamentals of digital image processing

- Acquisition: The first step in image processing is to acquire the digital image with a digital camera.
- Pre-processing: After obtaining the digital image, the next step is pre-processing. Pre-processing without using the resulting digital image.

The first step is pre-processing the image to obtain a more successful result. These processes are basically:

- ✓ Image enhancement,
- ✓ Image restoration
- ✓ Wavelets
- ✓ Image compression

- The acquisition and pre-processing of the image are called low-level image processing.
- Segmentation: After pre-processing, the segmentation phase is started. Segmentation is the most complex application of image processing. It is the process of separating the object and the background in an image or regions of interest within the image with different properties. Partitioning produces basic information on its shape by detecting the boundaries and areas of an object in an image. If we are interested in the shapes of objects, segmentation gives us information about the edges, vertices, and boundaries of that object. On the other hand, if the internal features of the objects in the image, such as surface coating, area, colors, and skeleton, are concerned, regional partitioning should be done. As with character or pattern recognition in general, for solving complex problems, it may be necessary to use both partitioning methods (boundaries and areas) together.
- Inference: The basic information obtained from the image and the details of interest are brought to the fore. In other words, it separates the feature areas being searched from the background and each other.

3.4.8. Feature extraction

Defining a set of image features will most efficiently or understandably denote the information essential for analysis and Classification (Tian, 2013).

The purpose of feature extraction is classification, which is more essential than analysis and classification. Machine learning and classification tools work directly on the image but on the features contained in the image in image processing applications. This is possibly done by:

1. Eradicating excess in the image data.
2. Eliminate differences in data during classification, especially the data value of images with little or no value.
3. Restructuring image data to improve classification efficiency.
4. Extract information from image data such as (text, size, shape....).
5. Extracting secret information in images.

3.5. Statistical Measures in Digital Imaging Processing

3.5.1. Some statistical measures to describe shape of mass

Mean of image

An image can be described using the moments of its gray levels or a region of interest (ROI). The brightness points of an image have a mean value and intensity range between 0 and 255. When the mean value is near 0, the image is close to black, and when the mean value is near 255, the image is bright. (Ferreira and Rasband, 2012; Gonzales and Woods, 2018)

Let Z be a random variable representing the intensity of the gray levels of the dashed image.

$p(z_i), i = 0,1,2, \ldots . L-1$, represents the corresponding histogram.

L: the number of distinct intensity levels.

The nth moment of Z about the mean is:

$$m_n(z) = \sum_{i=0}^{L-1} (z_i - m)^n \, p(z_i) \tag{3.6}$$

Where *m* is the mean value of Z (the average intensity):

$$m = \sum_{i=0}^{L-1} z_i \, p(z_i) \tag{3.7}$$

Variance of image

It is a phrase for the degree of variance of the image. It is the measure of how many pixel values are in an image. The more significant the variance, the better the image's detail and grey level. Moreover, images with low variance indicate (the presence of slight differences in the gray levels and tiny changes in the image).

The pixel values of an image change as one colour changes to another. The variance can be used to analyze distorted images and calculated using the difference in intensity of an image. The second moment of the gray levels of the image:

The second moment [the variance $\sigma^2 = m_2(z)$].

This measure can be used to find the Smoothness of the image according to the following formula:

$$R(z) = 1 - \frac{1}{1 + \sigma^2(z)} \tag{3.8}$$

Histogram of image

A graph shows how many gray values are in the image. It is essential in terms of giving information about the image. In a dark image, the gray values are at the down end of the histogram. In a bright image, the gray values are at the upper end. The gray values are evenly distributed across the range in an image with good contrast.

Skewness of image

A measure of the imbalance of the distribution of the gray values about the mean inside the region of interest (ROI).

$$m_3\,(z) = \sum_{i=0}^{L-1} (z_i - m)^3\, p(z_i) \tag{3.9}$$

Kurtosis of image

A measure of the "peakedness" distribution of the gray values around the mean within the ROI. The Kurtosis of the histogram of gray levels is used to compare the deviation of the peak of the frequency distribution curve from the peak of the normal curve (flattened, tapered, or moderate), which is the fourth moment about the arithmetic mean, as follows:

$$m_4\,(z) = \sum_{i=1}^{L-1} (z_i - m)^4\, p(z_i) \tag{3.10}$$

Entropy of image

Entropy measures the degree of randomness; its value ranges from (0) to ($2\log_2 K$). The lowest randomness condition occurs when the probability of one of the random variables is equal to the correct one (1), which is predefined and is (E = 0). and the most excellent randomness condition occurs when all events are of equal probability (E = $2\log_2 K$)

$$E = -\sum_{i=1}^{k}\sum_{j=1}^{k} p_{ij}\, log_2\, p_{ij} \tag{3.11}$$

G: Co-occurrence matrix

Integrated density of image

It is the totality of the gray values of all pixels inside the region of interest (ROI).

Median of image

The median gray value of the pixels within the region of interest (ROI) is the gray value in the middle of the mean of the two middle gray values when all gray values are sorted by their numerical value.

3.5.2. Some measurements of shape to describe (ROI) in digital images

Implementation of numerical measurements is one of the methods used in pattern recognition in digital image processing (Al-Hinawy, 2004; Eric, 2011; Ferreira and Rasband, 2012; Gonzales and Woods, 2018; Russ, 2006).

Among these measurements are unique shape measurements, which describe the tumour's shape using some numerical measurements. Generally, in mammogram images, the shapes, as shown in Figure (3.21), are classified as (Round, oval, lobulated, irregular, and Architectural Distortion).

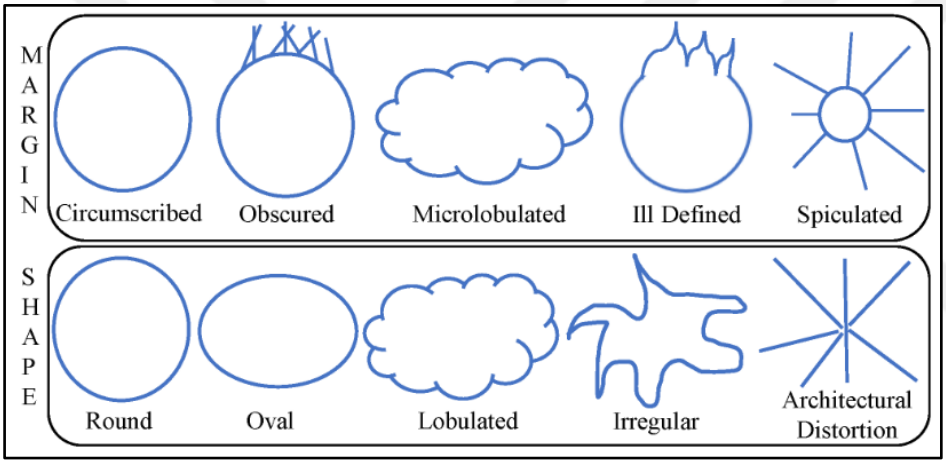

Figure 3.21. Shape to describe (ROI) (Kopans, 1989)

Some critical measurements can be illustrated and explained in describing the shape of benign and malignant tumour masses as they are:

Area

It is the number of pixels for the selected area, and it can be calculated for the circle in the following formula:

$$a = \pi r^2 \; and \; \pi = \frac{22}{7} \tag{3.12}$$

Perimeter

It is the length of the perimeter of the mass and can be calculated for a circle with the following formula:

$$p = 2\pi r \tag{3.13}$$

Width and height:

It is the mass's specific width and length.

Major and minor axes:

An ellipse's major and minor axes are the ellipse's diameters (lines via the center). The major axis is the most extended diameter, and the minor axis is the briefest if the length is similar.

Aspect ratio

It is the ratio of the length to the width of the primary and secondary axes of the specified mass according to the following formula:

$$AR = \frac{[\text{Major Axis}]}{[\text{Minor Axis}]} \tag{3.14}$$

Max. Feret (max. Diameter)

The longest line connecting two points on the perimeter of the mass can be joined so that the resultant line passes via the middle of gravity of the mass.

Min. Feret (min diameter)

The minor line connecting two points on the perimeter of the mass can be joined so that the resultant line passes via the middle of gravity of the mass.

As shown in Figure (3.22)

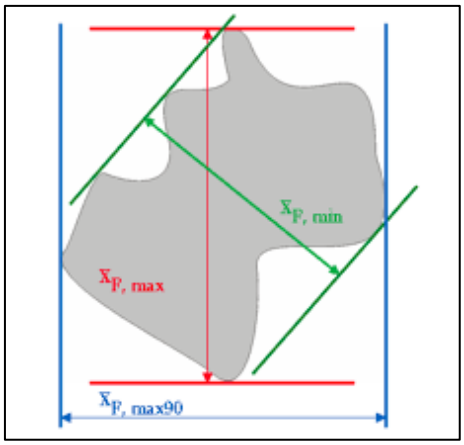

Figure 3.22. Max & Min Feret

Feret

It is the size of a mass along a given direction. Alternatively, the distance between two parallel planes bounds the mass along a given direction.

$$\frac{D_{F.max}}{D_{F.min}} \quad D_F: Feret\ diameter \tag{3.15}$$

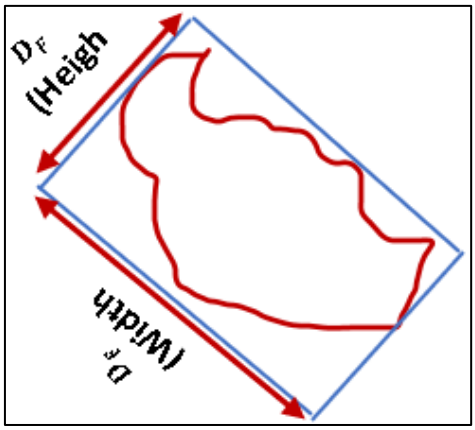

Figure 3.23. Max & Min - 2

Circularity

The expression gives the circularity ratio.

$$R_c = \frac{4\,\pi\,area}{p^2} \tag{3.16}$$

Solidity

It is the overall measure of the concavity of a particle, defined as the ratio between the image's area and the convex area. As the equation explains, the particle tends to solidity whenever the image area approaches the convex area.

$$s = \frac{[\text{Area}]}{[\text{Convex area}]} \tag{3.17}$$

3.6. Artificial Neural Network

Studies in the modern era began to revolve around trying to understand the complex mental processes that the human mind performs while thinking. Then, these cognitive processes were translated into computer simulations, with the computer's increasing ability to solve complex problems.

Attempts have emerged by researchers in the field of artificial intelligence, as they have faced difficulties in creating a machine or device capable of learning and acquiring knowledge that helps solve problems and improve computer performance so that it is somewhat similar to the human level or better than it. It may significantly impact our daily lives, Current and future.

Therefore, some scientists have sought to try to simulate the nervous system, especially the nerve cell (which is the basic unit of the nervous system). Billions of nerve cells are spread throughout the human body and branch into branches that work to transmit feelings of sensation and reactions to and from the brain and to study how the cell is processed. Neural biological information, storing and retrieving it, and then taking advantage of it to build an interconnected structure of a software structure similar in performance to human nerve cells. They arrived at a science called Artificial Neural Networks, which falls under the science of artificial intelligence, so they turn computers into smart devices that can like how humans acquire knowledge.

Artificial neural networks (ANNs) are nonparametric models capable of addressing non-linearity. Used in many fields, including signal processing, control, pattern recognition, medicine, speech recognition, business, finance, and others (Rabunal and Dorado, 2006).

3.6.1. An overview of artificial neural networks development

The development of artificial neural networks has a long history. Many researchers have contributed to its development, and here we focus briefly on those with a prominent role in developing artificial neural networks.

The general concept of the first artificial neural network is credited to the two scientists (McCulloch and Pitts, 1943). They developed the first mathematical model for a neuron (simple cell) with a two-state (binary) and a fixed (threshold) to represent logical functions with equal weights. They described the model as a unit calculated for a linear sample with multiple inputs and one output. They simplified the structure and functional performance of the cells. The brain has multiple inputs and one output, with only two possible states: active and inactive. Some of the two scientists' proposals were retained and preserved until now without modification, and other proposals were developed (Rabunal and Dorado, 2006).

The researcher (Heb,1949) developed the first rule for learning artificial neural networks and called it "the Hebbian Learning Rule". He hypothesized that if two neurons were activated, the connection's strength would increase, meaning several cells were next to each other, but two were transmitting information. Densely, it strengthens their relationship, and processing operations become faster with repeated stimulation by the same data (Fausett, 1994).

Rosenblatt (1959) Developed the first artificial neural network, called the (perceptron) cell. It consisted of a single layer of cells and a series of connections (binary weights) modified to suit the problem. The learning process was limited to linear models. Linear models were used for estimation, classification, or forecasting purposes. Enthusiasm and interest in single-layer neural networks diminished for many years, especially after the evidence presented by (Minsky and Paper, 1969) in their famous study on evaluating neural networks, which showed the limitations of dealing with single-layer neural networks, which caused many researchers to move away from this. And people lost confidence in neural networks (Hagan et al., 1996).

In his doctoral thesis, (Paul Werbos, 1974) added hidden layers to neural networks for the first time, giving sensory perception the ability to calculate functions and relationships for linear and nonlinear models and linking the input and hidden layers with weights that are modified using (backpropagation). However, this development did not spread widely until after the mid-eighties of the last century when the feedback algorithm was rediscovered through three independent studies (Lecun, 1985; Parker, 1985; Rumelhard, Hinton and Williams, 1986). Moreover, trained multilayer neural networks on it to learn and expand its uses to include many fields (Hagan et al., 1996).

In the period (1982-1988), Hopfield presented several research on a new type of neural network known as the Hopfield network, which played a role in accelerating the capabilities of artificial neural networks and their development. This neural network is connected internally, and each neural unit is connected with all other departments (Fully Connected Network) and uses a symmetric weight matrix. Hopfield added the idea of feedback networks (Feedback Neural Network) (Shankar, 2008).

(Kohonen, 1989) developed a network called the (Kohonen Network) or (Self-organizing Maps). The basic idea behind this network is the structure of interconnected processing units. That competes for the input signal and compares it with the vectors stored within these units. Each separately until the winning element is reached, which is the element that contains a vector similar or close to the input vector data (Rabunal and Dorado, 2006).

3.6.2. Components of the artificial neural network and biological

The hypotheses and algorithms of artificial neural networks are nothing but an attempt to simulate a more profound and more precise understanding of the work of biological neural networks in terms of their ability to receive, process, and send electrochemical signals through neurons. The human brain consists of millions of neurons that communicate with each other in a very complex way to form the neural network. Each neuron has three sections (dendritic branches, cell body, and axon). Dendrites receive incoming electrical signals. To the cell from other cells and transmit them, signals are collected in the cell body, some of which excite the cell and stimulate it when the inside exceeds the value of the limit plate of the neuron. So, it discharges an electrical charge from the cell body to the axon of the cell,

which in turn passes it to another cell connected to it by Dendritic branches as in Figure (3.23) (Hagan et al, 1996; Poznyak et al,2001).

The connection between neurons is done through synapses, which convert electrical signals using chemical processors called neurotransmitters. Therefore, communication in the nervous system is an electrochemical process, and the nodes adjust the value of incoming signals (Hagan et al, 1996; Poznyak et al,2001).

Based on the characteristics of the biological cell, the basic features of artificial neural networks were formed, as they consist of the inputs represented by the vector $\underline{X}$ that receives the signal and then multiplied by the weight represented by the vector $\underline{W}$. The set of inputs corresponds to dendrites in a living neuron, and the weights correspond to synapses. The value of the weights for the input signals expresses the strength of the connection between the previous layer and the subsequent layer. The weighted inputs are collected and then passed to the activation function, which consists of the summator. To sum the weighted inputs, which is expressed by the value (Net), which is equal to $\text{Net} = \sum_{i=1}^{n} x_i w_i$ Which corresponds to the body of the nerve cell to determine its level of activation effect to produce the output signal that is the input to other cells linked to it, as in Figure (3.23) (Hagan et al, 1996; Poznyak et al, 2001).

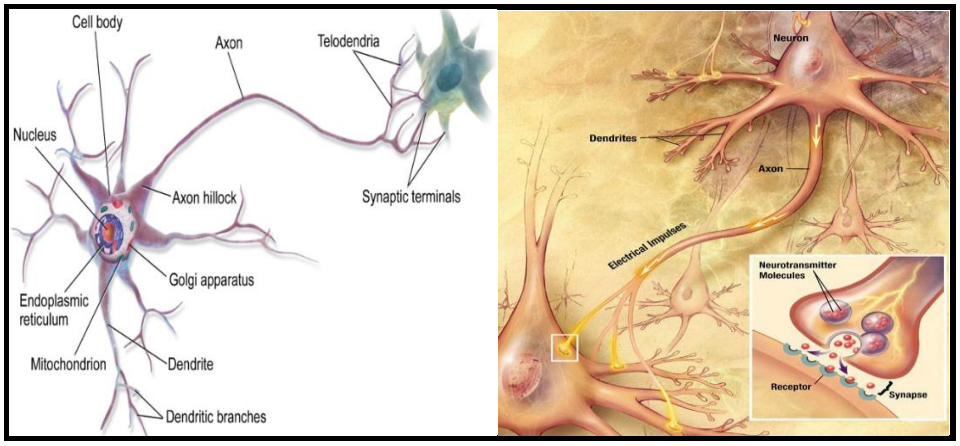

Figure 3.24. Biological neuron

Based on Figure (3.23), the following points can be inferred:

1- Dendrites: Each neuron has entry points that receive in the form of electrical impulses from other neurons in the network.
2- Cell Body (Soma): The system processes the information received from the dendrites and decides on what action to take.
3- Axon terminals: outputs in the form of neurons transmit electrical impulses to communicate with other neurons (Graupe, 2013; Omar and Rizgar, 2020).

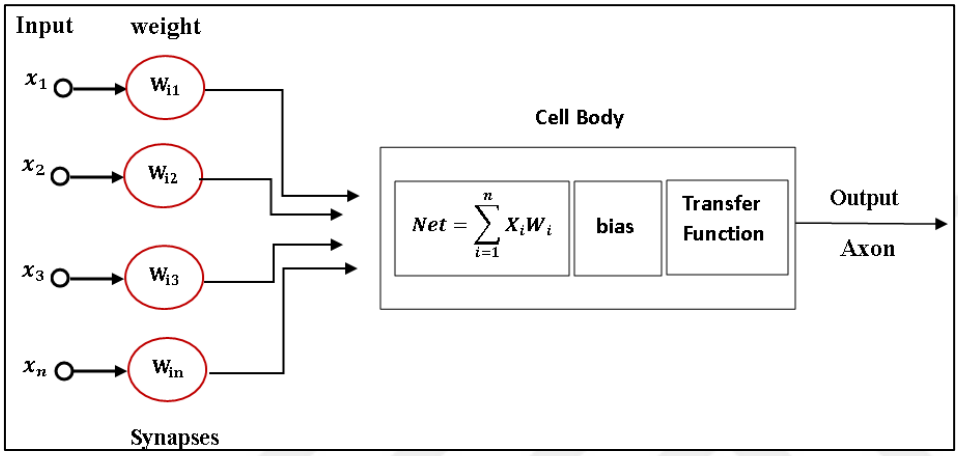

Figure 3.25. Components of an artificial neuron cell

Figure (3.25) illustrates that the neural network is composed of three distinct segments, which are as follows:

- The Input Layer consists of the values of x.
- The hidden layer contains the values of w and the resulting processes.
- The output Layer of the model includes the value of y.

The practical use of artificial neural networks lies in their ability to modify the weights of the nodes connecting the neurons to produce a desired output.

$$N_{Pj}^{L+1} = \sum_{i=1}^{n^L} W_{ij}^{L}\, Out_i^{L} + Bias_j^{L+1} \tag{3.18}$$

$$Out_{pi}^{L+1} = f\left(N_{Pj}^{L+1}\right) = \frac{1}{1 + e^{-BN_{Pj}^{L+1}}} \tag{3.19}$$

3.6.3. Types of activation functions

There are specific and well-known activation functions in training algorithms, and one of the common activation functions used in training algorithms is:

Threshold function or step function

It is in the form of a curve resembling a staircase in two cases:

A. Binary case: This function determines the output of the contract, i.e. the output., where the output becomes equal to one if it is greater or equal to zero, and the output becomes equal to zero if it is less than zero, as in the figure

$$f(x) = \begin{cases} 1 & if\ x \geq 0 \\ 0 & if\ x < 0 \end{cases} \tag{3.20}$$

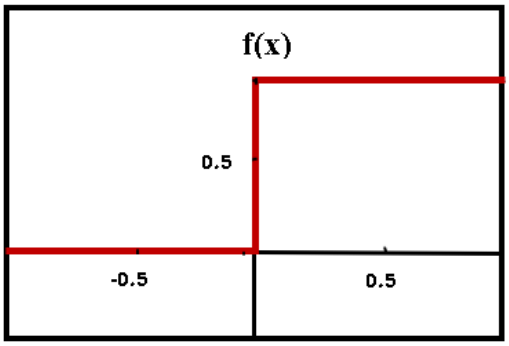

Figure 3.26. Threshold function in the binary case

B. The bipolar state: This function determines the output of the nodes, meaning the output becomes equal to one if it is greater than or equal to zero, and the output becomes negative one if it is less than zero, as in Figure (3.26).

$$f(x) = \begin{cases} 1 & if\ x \geq 0 \\ -1 & if\ x < 0 \end{cases} \tag{3.21}$$

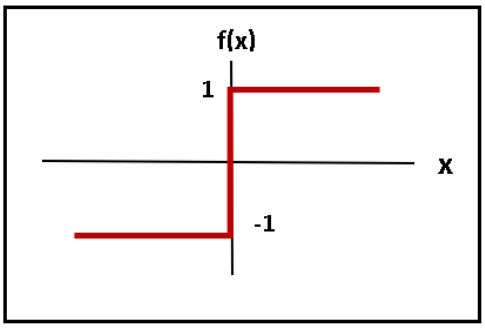

Figure 3.27. Threshold function in the bipolar state

Linear activation function

It is a linear function used in single-layer artificial neural networks to convert the network input into a form suitable for the output signal. It is a variable and continuous function, and Figure (3.28) shows the linear function.

$$f(x) = x \quad -\infty < x < \infty \tag{3.22}$$

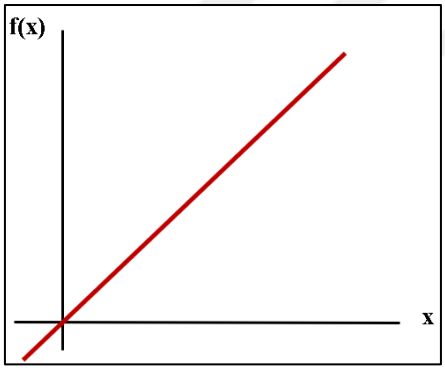

Figure 3.28. Linear activation function

Sigmoid function

They are the most widely used types of nonlinear functions, and they are of two types:

A. Binary Sigmoid Activation Function: The most widely used function is an activation function for networks in which the required output value is a binary value between zero and one. Figure (3.29) shows this function.

$$f(x) = \frac{1}{1 + \exp(-x)} \tag{3.23}$$

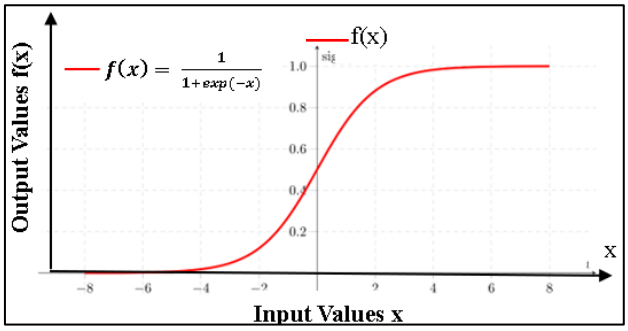

Figure 3.29. Binary sigmoid function

B. Bipolar sigmoid activation function: It is the most common use of the function when the desired range is between negative one and one, and Figure (3.30) shows this function:

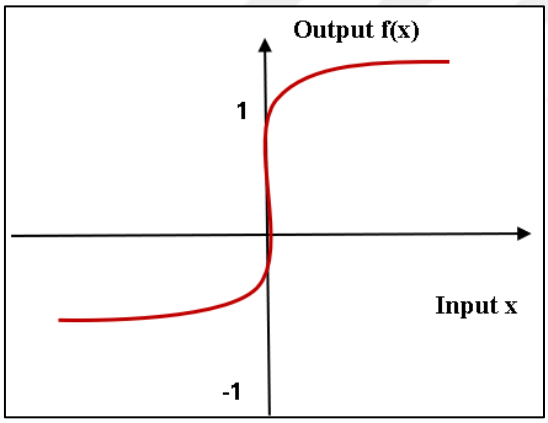

Figure 3.30. Bipolar sigmoid activation function

3.6.4. Artificial neural networks architecture

Artificial neural networks are a parallel group of simple processing units (nodes). Moreover, the interconnections between these units are of particular importance when creating the network, as the arrangement of the nodes in the layers and the form of the connections between the layers is called the architectural structure of the neural network. The number of layers in the network is calculated without counting the input layer because no mathematical

operation is performed, or it can be known from the number of connections between neurons (Hagan et al., 1996; Fausett, 1994).

In general, the typical artificial neural network architecture consists of three levels or layers (Fausett, 1994; Wu and McLarty, 2000):

Input layer: The first layer in the neural network. Includes input neurons that receive input from the outside world, and their number changes according to the type of issue.

Hidden layer: The middle layer lies between the first layer (input) and the last layer (output) and receives input from the previous layers.

Output layer: represents the last layer of the network, and the nodes in this layer provide the network's output or solve the problem.

Each of the above three layers consists of: (Fausett, 1994)

a. Synapses: These form the neural connection points between the layers of the neural network.
b. Layer: represents a group of nodes that receive input and have output.
c. Weights: Weights indicate the strength of the neural connection between the levels (layers) of the artificial neural network. Each node has a weight that links it to the previous layer and a weight that links it to the subsequent layer. It includes the weights of the information associations based on which the network will learn. Choosing the initial values for the weights and the network at the beginning of training is critical in the speed of network convergence and in reaching the minimum global error. The initial values are small random numerical values that are generated from statistical distributions (Bishop, 1995; Wu and McLarty, 2000).
d. Bias: It can be a component of the input vector and always takes the value ($X_0 = 1$) within the input vector $X = (1, X_1, X_2, \dots, X_n)$. Adding the bias unit to the input units changes the shape of the activation function and its typical form:

$$f(Net) = \begin{cases} 1 & if\ Net \geq 0 \\ -1 & if\ Net\ > 0 \end{cases} \qquad (3.24)$$

$$Net = b_0 + \sum_{i=1}^{n} X_I W_i \quad , \; b_0 = X_0 W_0$$

Because the work of the bias is similar to the work of the weights of the link between the units, but it has a constant activation that is always equal to one (and usually, these units are not shown in network diagrams), any neuron without the bias is like a linear equation without a constant that passes through the origin. (Gupta et al., 2003)

There are multiple types of networks whose names and characteristics differ depending on the number of layers (the connection of cells from one layer to another). Therefore, artificial neural networks are classified in terms of the number of layers:

Single layer networks

In this type of network, there is one layer of weight connections between the input and output layers, and it does not have a hidden layer. The flow of signals entering from the input layer to the output layer is in a Forward Direction (Figure 3.31). It is also called the Single Layer Perceptron or Perceptron Network. (Haykin, 1999)

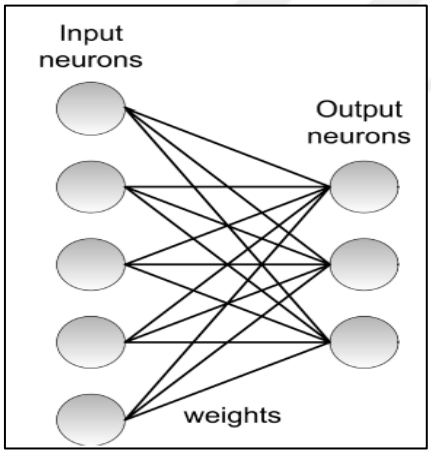

Figure 3.31. Single layer network

The Single layer perceptron architecture includes a layer of input neurons that are fully connected to a layer of output neurons.

Multilayer networks

These networks consist of three layers: the input and output, with the hidden layers in the middle (one or more layers). Each layer is connected to the next layer through links (weights) and sends its output to each node (cell) in it, and the number of nodes (cells) may vary. In

the hidden layer, depending on the degree of complexity of the problem and the number of inputs, and because of the presence of hidden layers, these networks have become distinguished by their ability to solve complex problems, especially non-linear problems that cannot be solved using a single layer method and that training takes a long time, and that training multi-layer networks despite It takes longer. However, training is more successful than single layer, knowing that the characteristics of the problem decide which type of network is used. As in Figure (3.32): (Bishop, 1995; Haykin, 1999)

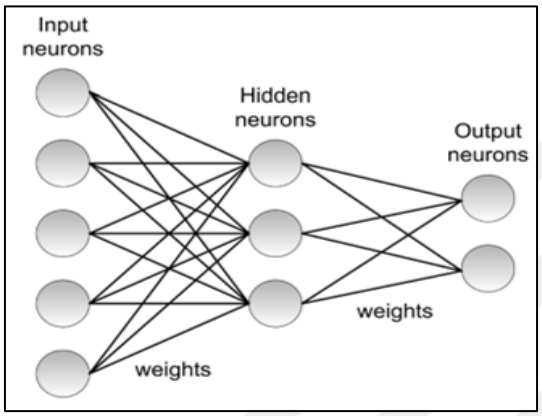

Figure 3.32. Multilayer layer network

A Multilayer perceptron can include more than one layer of trainable weights. In this example, the network has 3 layers: an input layer, a hidden layer, and an output layer. Each neuron connection is represented by a particular weight.

3.6.5. Training artificial neural networks

The training process is the basis for the process of learning neural networks. According to the training process, there are two types of artificial neural networks: (Fixed Network Nets), whose weights do not change during training, and (Adaptive N. Nets), which can change the associated weights with all cells. Artificial neural networks are not programmed; they learn through specialized algorithms that change the strength of the internal connection (weights). This is called weight adaptation, which occurs during the training phase. Initial weights are chosen for training randomly or based on some previous experiences. The training process begins by giving examples (the external input set) called the training set for the network or the training sample to form the output (output). These weights are gradually adjusted with

each example until the error reaches the smallest possible between the input and the output (Fausett, 1994).

Dividing the data during the training process into automated groups: (Bishop, 1995; Xu and Chin, 2008).

a. Training Set: This set is used to train the neural network to determine a model for the data.
b. Validation Set: This set performs the final test of the neural network's performance.
c. Testing Set: Through which the network skill can be determined and the possibility of using it in general in applied fields.

There are two types of training artificial neural networks:

Supervised training

According to this training method, artificial neural networks must have two pairs of vectors: the input vector and the desired output vector. The neural network is trained on several vectors for these pairs. The training begins with the input vector, which is applied to the network and produces the output. It is compared with the corresponding expected output vector, and the difference represents the error. This error is used to adjust the weights according to the training algorithm. All vectors are applied, and training continues until the error for the input training set reaches the lowest possible level. It is also called supervised training or “Training with Teacher” (Haykin, 1999).

Unsupervised training

Unsupervised training, or as it is known as self-organizing networks, consists of an input vector, a weight vector, and a training algorithm to change the network weights to produce the real output vector. It does not know the nature of the correct answer, and the network does not know the desired output. It does not need unsupervised learning algorithms (unsupervised Training) to the desired signal, but rather, the network is fed with the input data only. This Algorithm automatically adjusts the network’s weights. It extracts the characteristics in the input signals and groups them by strengthening the weights of the cells that respond to a specific characteristic carried by the input data. This type of network uses

a competitive learning rule between neurons. One of the famous algorithms in this type of training is the Algorithm of Kohonen (Bishop, 1995; Haykin, 1999).

3.6.6. Types of artificial neural networks

Artificial neural networks are classified according to the connections of their cells. They are classified into two main types:

Feed forward neural network

In it, the direction of propagation of signals entering the network is always forward, and thus, the signal leaving any cell depends on the signals entering only.

Feedback neural network

In these networks, it is possible to re-feed (all or some) the signals coming out of the network and change their direction to become an input signal as well (re-feed), and these networks are called “Recurrent Networks”.

3.6.7. Backpropagation neural network

It is one of the most important and commonly used artificial neural networks in learning. It uses multilayer neural networks (with one or more hidden layers) that are feed-forward according to the type of supervised training. The (Criterion) standard uses the total least square error of the output value calculated by the network until the optimal weights are reached, i.e., the lowest possible sum of squared error between the outputs of the neural network and the model data. It is also called the Generalized Delta Rule algorithm (Haykin, 1999; Fausett, 1994).

Activation functions which are used in backpropagation network

The activation function in feedback networks must possess several important properties. It must be the function is continuous and capable of non-decreasing streamline differentiation, and most importantly of all, its derivative should be easy to calculate. For most of the activation functions used, the value of the derivative can be expressed in the activation

language. Usually, the function is expected to approach the saturation region, meaning it asymptotically reaches a specific maximum or minimum value (Fausett, 1994).

One of the most essential activation functions is the Binary Sigmoid Function, whose values fall within the range [0.1] and is known by the following relationship:

$$f_1(x) = \frac{1}{1 + \exp(-x)} \tag{3.25}$$

Its derivative is:

$$f_1'(x) = f_1(x)[1 - f_1(x)] \tag{3.26}$$

The other common form of the activation function is the bipolar squirmed function (tansig function), which takes its values within the range [-1, 1] and is given by the following relationship:

$$f_2(x) = \frac{2}{1 + \exp(-x)} - 1 \tag{3.27}$$

Its derivative is:

$$f_2'(x) = \frac{1}{2} f_2(x)[1 + f_2(x)][1 - f_2(x)] \tag{3.28}$$

Error backpropagation algorithm

Training the artificial neural network using feedback is a gradual regression method to find the least square value of the total error. It includes three stages after choosing the initial values of the weights, which are as follows:

1- Error Forward Propagation stage
2- Error Back Propagation stage
3- Adjustment weights stage.

- Feed-forward stage: Each input unit $(1, X_1, X_2, \dots, X_n)$ receives the input signal and then propagates. This signal is sent to each unit of the hidden layer $(H_j, j = 1,2, \dots, P)$, then each unit of the hidden layer calculates a value using one of the activation functions for this signal, and then these units send their signal to each unit of the output layer. Then, each output layer unit $(Y_k, K = 1,2, \dots, m)$ calculates an activation function to form the network's response to the given input sample (Fausett, 1994).
- Training stage: Each output unit compares its calculated activations γ_K with the target output value t to determine the value of the error occurring for this sample for that unit. Depending on the value of the error happening, factor (δ_k) is calculated, as factor δ_k is used to distribute the error to the units of the output layer γ_K so that it is returned to all units in the previous layer (the hidden units associated with the output units γ_K). This factor is also used (later) to update the weights between the output layer and the hidden layer. Similarly, the factor δ_j is calculated for each unit of the hidden layer H_j and is not necessary to propagate the inverse error to the input layer units. However, (factor δ) updates the weights between the hidden and input layers. After the factors (δ) have been determined, the weights are combined for all classes simultaneously. This tuning is done concerning the weights W_{jk}(coming from the hidden layer units H_j to the output units γ_K) based on factor δ_k and the activation h_j with respect to the hidden layer units H_j. As for adjusting the weights of V_{ij} (coming from the input layer unit X_i to the hidden layer unit H_j) depends on factor δ_j and the activation of the input unit (X_i) (Fausett, 1994).

The error feedback network's operating algorithm can be summarized in the following steps:

1 - Generating initial values for weights from "one of the statistical distributions".

2- As long as the stopping condition has not yet been met (meaning the error is a certain value), perform the steps from

3 ⟶ 10

3- For each training pair, perform steps 4 ⟶ 9.

- Feed Forward stage (feed forward direction):

4- Each input unit ($X_i, i = 1, 2, \ldots\ldots, n$) receives its input signal X_i and then sends it to all hidden layer units.

5- Each hidden unit ($H_j, j = 1,2,3, \ldots\ldots, p$) collects its weighted input values and signals as follows:

$$H_j = V_{0j} + \sum_{i=1}^{n} X_i V_{ij} \tag{3.29}$$

The activation function is then applied to estimate the output of the hidden layer unit.

$$h_j = f(H_j) \tag{3.30}$$

The activation values are sent to all units (nodes) in the output layer.

6- Each output unit ($Y_k, k = 1, 2, \ldots\ldots, m$) collects its weighted input signals as follows:

$$Y_k = W_{0k} + \sum_{j}^{p} h_j W_{jk} \tag{3.31}$$

The activation function is then applied to estimate the output of the output layer unit.

$$y_k = f(Y_k) \tag{3.32}$$

Error Back Propagation stage

Calculated error for the output nodes by calculating the difference between the activation value, i.e., the output value of the nodes y_k, and the real value t_k, i.e.

$$E_k = t_k - y_k \tag{3.33}$$

Then, the outputs of the neural network are compared with the real values to estimate the error according to the formula:

$$\delta_k = (t_k - y_k) f'(net) \tag{3.34}$$

Whereas:

$$f'(net) = f'(Y_k) \tag{3.35}$$

Then, calculate the change in the size of the error ΔW_{jk} according to the equation:

$$\Delta W_{jk} = \alpha . \delta_k h_j \tag{3.36}$$

Then calculate the bias correction term (used to update the W_{0k} weight later)

$$\Delta W_{0k} = \alpha . \delta_k \tag{3.37}$$

The δ_k is then sent to the units in the layer below (hidden layer).

8- Each unit in the hidden layer ($H_j, j = 1,2,3, \dots . , p$) collects the weighted input signals, the δ, as in the formula

$$\Delta_j = \sum_{k=1}^{m} \delta_k W_{jk} \tag{3.38}$$

This value is then multiplied with the derivative of the activation function to calculate δ_j

$$\delta_j = \Delta_j f'(net(j)) \tag{3.39}$$

$$f'(net(j)) = f'(\Delta_j) \tag{3.40}$$

Then, the change in the error size ΔV_{ij} is calculated.

$$\Delta V_{ij} = \alpha . \delta_k X_i \tag{3.41}$$

Then, its bias correction term (used to update the weight V_{oj} later) is calculated.

$$V_{oj} = \alpha . \delta_j \tag{3.42}$$

- The stage of updating weights and biases:

9- The weights and biases for each unit in the output layer are determined according to the following formula:

$$W_{jK}(new) = W_{jK}(old) + \Delta W_{jk}\ , j=0,1, \ldots\ldots, p \qquad (3.43)$$

Then, the weights and biases are generated for each unit in the hidden layer according to the formula below, and then its activation function is applied to estimate the output of the hidden layer unit.

$$V_{ij}(new) = V_{ij}(old) + \Delta V_{ij} \quad , \quad i = 0,1 \ldots\ldots, n \qquad (3.44)$$

10- The network updates the weights, i.e., the training process, until the optimal weights are obtained. Then, the desired outputs are obtained, i.e., reaching the best fit for the model under investigation (meaning testing the stopping condition). Figure (3.33) shows the flow of the weights updating algorithm.

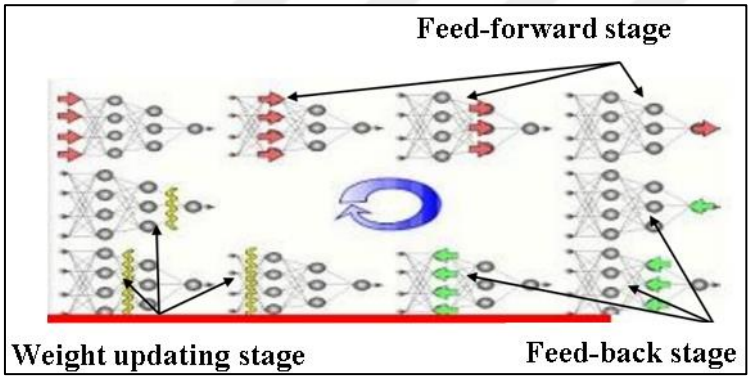

Figure 3.33. Streamlined weight updating algorithm

Traingdm and momentum (μ)

When adjusting the weights in a neural network, incorporating the momentum rate serves as a constraint, ensuring a controlled alteration of weights. This foster balanced and stable adjustments, enhancing the training process and thereby reducing the overall training duration. Additionally, it safeguards against the network converging to a local minimum. The advantages of momentum become particularly evident when the network adapts to vastly dissimilar or potentially erroneous data, as the momentum factor facilitates smoother adjustments (Fausett, 1994; Wu and McLarty, 2000).

$$w_{jk}(t+1) = w_{jk}(t) + \alpha.\delta_k.z_j + \mu\left[w_{jk}(t) - w_{jk}(t+1)\right] \quad (3.45)$$

Levenberg-marquardt (l-m) algorithm and trainlm

It is one of the feedback algorithms for training neural networks, as it updates the values of weights and biases according to the Levenberg Marquardt method and the (L-M) algorithm, which was developed independently by (Kenneth Evenberg) and (Donald Marquardt), to solve problems existing in both progressive regressions. Combining these two algorithms, the Gauss-Newton method in neural networks is used, but it is slower than the Gauss-Newton method. It provides a numerical solution to the problem of nonlinear functions to reduce the error to a minimum through a Hessian matrix.

$$W_{k+1} = W_k - (J_k^T J_k + \tau I)^{-1} J_k^T e \quad (3.46)$$

Where the parameter τ changes during the training process, and when it equals zero, $\tau = 0$, the LM algorithm matches the Gauss-Newton algorithm, and when the value of τ is large, it works as a gradual regression algorithm to train the network. The LM algorithm needs help with problems when the pattern sizes are large because the Jacobian matrix (J) is suitable for small and medium (Wilamowski and Irwin, 2011).

3.6.8. Artificial neural network performance factor

Multiple factors affect the efficiency of a neural network, including:

Number of vectors or exemplars in neural network

When conducting the training or learning process for a neural network, its performance is directly affected by the number of vectors (explanatory variables) that represent the inputs to the network. When the number of vectors or models entering the network is greater than the number of inputs or explanatory variables for each model, in this case, the neural network extracts a model. From that data represent all the relationships and properties of a complex nature that the network was able to learn from that data. In other words, if the inputs to the neural network, i.e., the models provided to it, are of a degree of complexity, then a more

significant number of explanatory variables must be introduced for each model when training the network on those inputs (Sarowar et al, 2009).

Training rate ($\boldsymbol{\alpha}$)

The training rate is one of the factors affecting the control of the size of weight changes during the process of training an artificial neural network. Choosing the optimization training rate is very important in trying to reach the lowest general error (Global minimum error). The optimization value practically cannot be allocated before training the network, so the training rate changes. During training, instead of customizing it, using a tremendous training rate leads to accelerating the training process but affects the accuracy of the network, and it becomes unstable. On the other hand, using a tiny percentage makes the network learn very slowly. To choose appropriate training rate value, we choose values between "positive one" and "zero" (Gupta et al., 2003; Wu and McLarty, 2000).

Number of hidden layers and hidden nodes

The efficiency of training a neural network in the classification process depends on several factors, including the number of inputs and outputs, the number of training samples, the amount of error in the objective function, the algorithm and structure used, and the number of nodes in the hidden layer. Moreover, could you determine the appropriate number of hidden nodes. The process of training the neural network begins with one hidden layer that contains several hidden units and determines the correct number of neurons used in hidden layers according to the points:

- The number of hidden units must be between the number of inputs and output layers.
- The number of hidden units must be 2/3 of the number of input layers as well as the number of output layers.
- The number of hidden units must be less than twice the number of the input layer.

Sometimes, it starts with one node, and the number of hidden nodes gradually increases until the lowest possible error is reached. However, in some cases, with the increase in the number of nodes, it is impossible to get the lowest error, so we add another hidden layer. The most common method for determining the number of nodes and hidden layers is (trial and error).

Although theoretical studies have shown that a single hidden layer is sufficient for an artificial neural network to approximate any complex nonlinear function, the large number of nodes in the hidden layer will lead to a longer time spent during training and excessive adaptation. Having more than one hidden layer can be helpful in some applications. Two hidden layers can make the network to train easier in some exceptional cases (Panchal et al., 2011; Stathakis, 2009).

Several rules, including the Baum-Haussler Rules, are used to determine the number of hidden layers in an artificial neural network (Chiang and Braun, 2004).

$$N_{Hidden} \leq \frac{N_{Train} E_{Tolerance}}{N_{Pts} + N_{output}} \tag{3.47}$$

$N_{Hidden} =$ Represents the number of hidden neurons (nodes).

$N_{Train} =$ Represents the number of Training Examples

$E_{Tolerance} =$ Represents the amount of permissible error (probability)

$N_{Pts} =$ Represents the number of data points for each training example.

$N_{Output} =$ Represents the number of output neurons.

3.6.9. Statistics and artificial neural networks

Do artificial neural networks have an essential relationship to statistics? From this question, researchers in both fields began to study the relationship between them, and interest in it increased increasingly in the mid-eighties for the purpose of understanding the differences and similarities between the two fields and encouraging scientists and researchers to work in it. Multiple research papers were presented on this relationship, and the prominent or fundamental question for discussion in this research was whether networks Can artificial neural science be an alternative to statistics in data analysis. Or is there a relationship of correspondence or contradiction? The conclusion from this research, through comparison, was that statistics and neural networks are consistent in data analysis. Still, there is a great overlap between the two fields, and they look at the same goal with a common lens. Many

of the ideas that researchers have in neural networks are based on the basics of statistics, and many statistical tools and methods can be programmed by employing networks and building an algorithm for them. They may be better suited to training neural networks than the algorithms proposed for education, and neural network models can be described correctly within a statistical framework. The reason for this is the excellent similarity and correspondence between the methods used in both fields. For example, single-layer networks were limited to solving problems for many years, and only a way was known to train multi-layer networks once statistical training appeared, which provided a way out of this dilemma. (Fletcher, 2002; Flexer, 1995; Sarle, 1994).

Since feed-forward artificial neural networks are statistical in nature, their training relies on statistical theories and models. Multiple neural networks have been used in statistical applications, including the Perceptron network, in estimating the simple regression model, nonparametric regression, and discriminant analysis capable of dealing with a particular type of regression or prediction problems with an indication of the variable or vector as a response. (Hagan et al., 1996; Fausett, 1994)

Several models of neural networks can be observed that are similar or identical to known statistical models. Through table (3.2). (Fletcher, 2002)

Table 3.2. Similarity between neural network models and statistical models

Equivalent Statistical Model	Neural Network Model
Multivariate Multiple Linear Regression	Simple Linear perceptron
Logistic Regression	Simple Nonlinear perceptron
Linear Discriminate Function	Adeline
Simple Nonlinear Regression, Multivariate Multiple Nonlinear Regression	Multilayer Perceptron
Principal Components	Unsupervised Habbian Learning
Least-square Cluster Analysis	Adaptive Vector Quantization
Variation of Nearest-neighbor Discriminate Analysis	Training Vector Quantization
Kernel Regression Methods	Radial Basis Function

Although many models of neural networks are similar or completely identical to the well-known statistical models, the terminology in the neural network's literature is entirely different from its counterparts in statistics. For example, there are terms in the neural network's literature, as shown in Table (3.3).

Table 3.3. Statistical terminology and its equivalent in artificial neural network

Neural Network	Statistic
Features	Variables
Input	Independent Variable
Output	Predicted Values
Target or Training Values	Dependent Variable
Errors	Residuals
Training, Learning, Adaptation Self-organization	Estimation
Error Function, Cost Function	Estimation Criterion
Patterns or Training Pairs	Observations
(Synaptic) Weights	Parameters Estimates
Supervised Training	Regression and Discriminate analysis
Unsupervised Training	Cluster analysis or data reduction
Bias	Intercept
Noise	Error term
Prediction	Forecasting

3.7. Wavelet Shrinkage

Wavelet shrinkage is a method utilized in image processing to reduce noise or compress images. It is based on the concept of thresholding wavelet coefficients to suppress noise or enhance signal features.

Wavelet-based methods are now widely used in various fields, including mathematics, signal and image processing, geophysics, bioinformatics, and many more. In statistics, wavelets are mainly used for non-parametric regression, density estimation, scaling assessment,

functional data analysis, and stochastic processes. These methods make use of the possibility of representing functions that belong to specific functional spaces as expansions on a wavelet basis. Wavelet expansions have unique characteristics that make them particularly useful. They are adaptive in both time and scale or frequency and have coefficients that are typically sparse. Fast computational algorithms can obtain the coefficients, and the magnitudes of the coefficients can be linked to the smoothness properties of the functions they represent. Wavelet representations enable flexible analysis of time and frequency data, providing computational benefits and facilitating statistical data modelling across different resolution scales (Gencay, 2002).

3.7.1. Wavelet properties

Wavelets possess certain crucial properties that define their behaviour. These properties are outlined below (Hamad, 2010):

1. The wavelet function $\psi(.)$ has zero average on interval $(-\infty, \infty)$; that is

$$\int_{-\infty}^{\infty} \psi(x)dx = 0 \tag{3.48}$$

To meet the requirement, the wavelet should have an oscillatory waveform. Furthermore, $\psi(.)$ should be localized.

2. The square of $\psi(.)$ in integrates to unity:

$$\int_{-\infty}^{\infty} \psi(x)dx = 1 \tag{3.49}$$

This condition ensures that the wavelet function is either compactly supported or of finite length.

3. Wavelet function has n-vanishing moments

$$\int_{-\infty}^{\infty} x^k \psi(x) dx = 0 \qquad k = 0,1, \ldots., N \tag{3.50}$$

The wavelet finite sum accurately represents the signal.

4. The integral of the scaling function $\emptyset(.)$ over the entire interval $(-\infty, \infty)$ equals one:

$$\int_{-\infty}^{\infty} \emptyset(x) dx = 1 \tag{3.51}$$

The area under the scaling function is normalized to one.

3.7.2. Some of wavelet function

Haar, Daubechies, Biorthogonal, Morlet, Mexican Hat, Meyer, Coiflets, Symlets

Here, we will discuss and explain some wavelet functions:

Haar Wavelet

The Haar wavelet is simple, low-cost, and easy to apply. However, it has limitations in providing effective compression and noise removal for audio signal processing applications. An alternative to the Haar wavelet is the Daubechies wavelet, which, despite offering better performance.

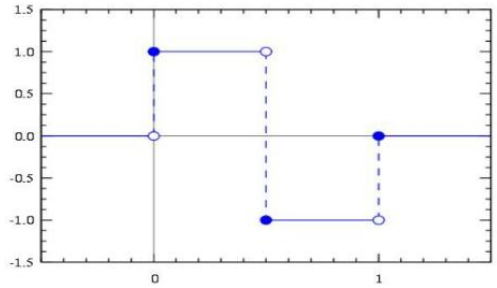

Figure 3.34. Haar wavelet

Daubechies wavelet

This function is an essential aspect of wavelet theory and is the focus of our work in this thesis. The Haar wavelet has limited effectiveness in discrete wavelet transform (DWT) applications. While it performs well for signals primarily composed of square waves, it falls short in providing clear information when processing more general harmonic waves. To address this, Daubechies introduced compactly supported orthonormal wavelets. Their unique shape has made discrete wavelet analysis both attractive and practical. In 1988, Ingrid Daubechies created a set of compact support wavelets that are commonly known as the Daubechies family wavelets are written as (dbN). These wavelets are identified by their "surname", db, and order number, N, to obtain the coefficients that meet the necessary conditions, one can use the length-4 coefficient sequences (Taha and Dlshad, 2021).

$$\sum_{k=0}^{N-1} h_k = h_0 + h_1 + h_2 + h_3 = 2 \tag{3.52}$$

$$\sum_{k=0}^{N-1} {h_k}^2 = {h_0}^2 + {h_1}^2 + {h_2}^2 + {h_3}^2 = 2 \tag{3.53}$$

$$\sum_{k=0}^{N-1} h_k h_{k+2m} = h_0 h_2 + h_1 h_3 = 0 \tag{3.54}$$

A unique solution exists for the given equations:

$$h_0 = \frac{1+\sqrt{3}}{4} \quad , \quad h_1 = \frac{3+\sqrt{3}}{4}$$

$$h_2 = \frac{3-\sqrt{3}}{4} \quad , \quad h_3 = \frac{1-\sqrt{3}}{4}$$

Figure 3.35 illustrates the wavelet functions of this family. It shows how the state of regularity changes as the wavelet length or rank increases.

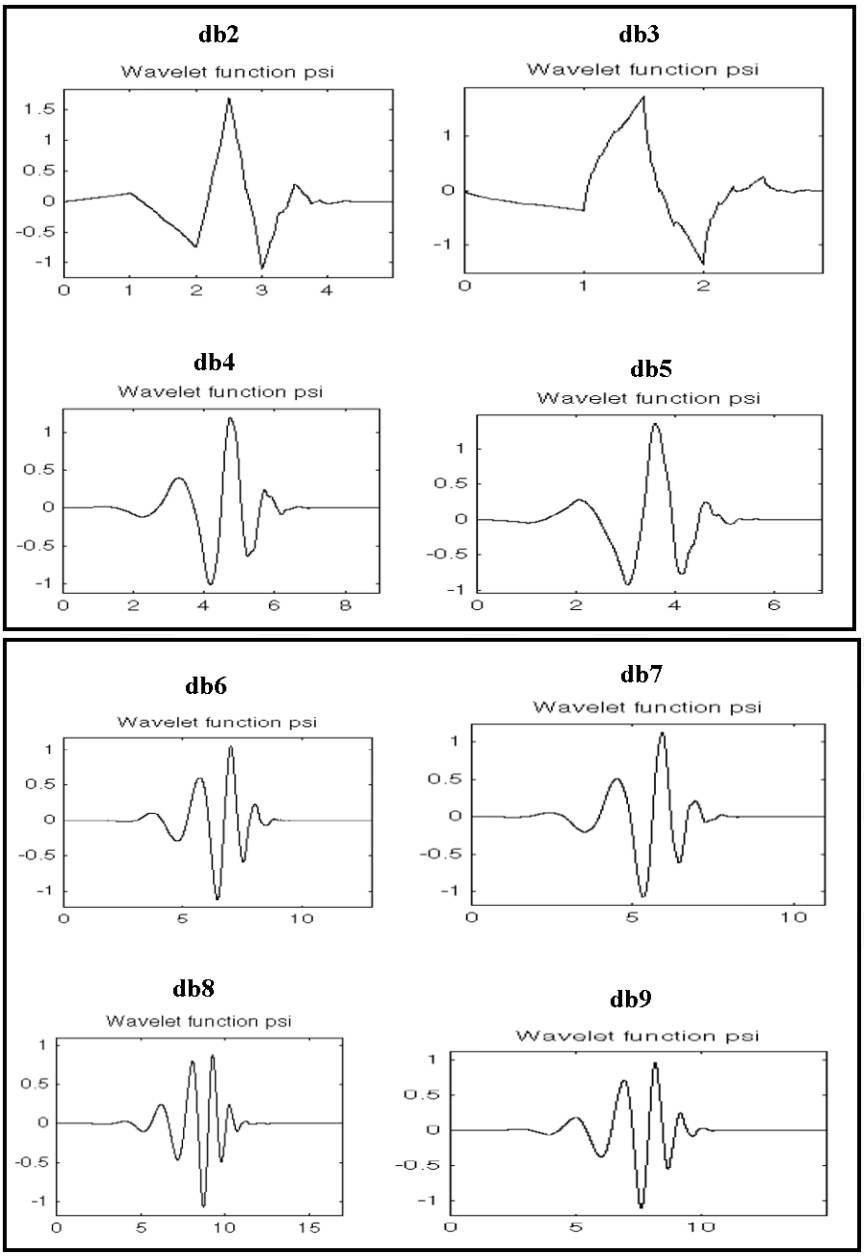

Figure 3.35. Daubechies wavelets (dbN)

Biorthogonal wavelet

Linear reconstruction of signals and observations is greatly enhanced by using a biorthogonal wavelet, which is more versatile than an orthogonal wavelet. By combining biorthogonal wavelets with a filter that has a finite impulse response, precise and symmetrical reconstruction is possible. We can create a reverse biorthogonal wavelet by using two biorthogonal wavelets. Figure 3.36 shows the graph of the biorthogonal wavelet.

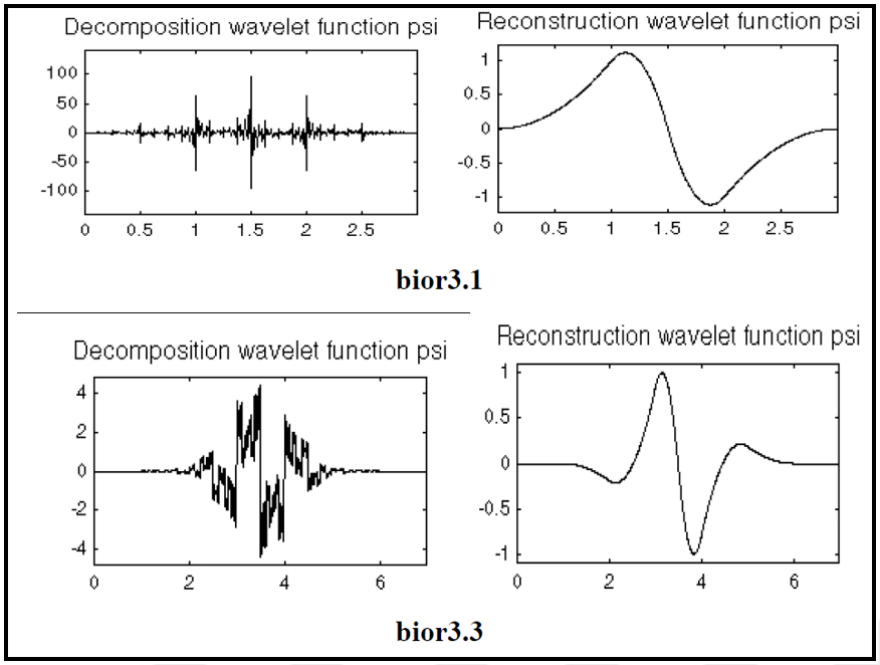

Figure 3.36. Biorthogonal wavelet Pairs (biorNr.Nd)

3.7.3. Discrete wavelet transform (DWT)

The Discrete Wavelet Transform (DWT) is a frequently used technique in image processing because it can break down an image into various frequency components. This enables efficient representation and analysis. When DWT is used in combination with Artificial Neural Networks (ANNs), it can be utilized for various purposes, such as feature extraction, denoising, compression, and image classification.

The Discrete Wavelet Transform (DWT) is a form of signal decomposition that produces a non-redundant image representation. This representation provides better spatial and spectral localization of image formation compared to other multi-scale representations, such as Gaussian and Laplacian pyramids. Recently, the DWT has attracted more interest in image de-noising.

The DWT can be interpreted as a signal decomposition in a set of independent, spatially oriented frequency channels. In this process, the signal S is passed through two complementary filters and emerges as two signals, approximation and details. This is called decomposition or analysis. The components can be reassembled back into the original signal without any loss of information. This process is called reconstruction or synthesis.

The mathematical manipulation that involves analysis and synthesis is known as discrete wavelet transform and inverse discrete wavelet transform. An image can be decomposed into a series of various spatial-resolution images using DWT. In the case of a 2D image, an N-level decomposition can be performed, resulting in 3N+1 various frequency bands, namely, LL, LH, HL, and HH, as shown in Figure (3.37).

The sub-bands may be called a1 or the first average image, h1 horizontal fluctuation, v1 vertical fluctuation, and d1 the first diagonal fluctuation. The sub-image a1 is formed by computing the trends along rows of the image followed by calculating trends along its columns. In the same way, fluctuations can be generated by calculating trends along rows and then trends along columns.

The low frequency subband image LL undergoes wavelet transform at the next level. The Gaussian noise is nearly averaged out in low-frequency wavelet coefficients. Therefore, only the wavelet coefficients in the high-frequency levels need to be thresholder.

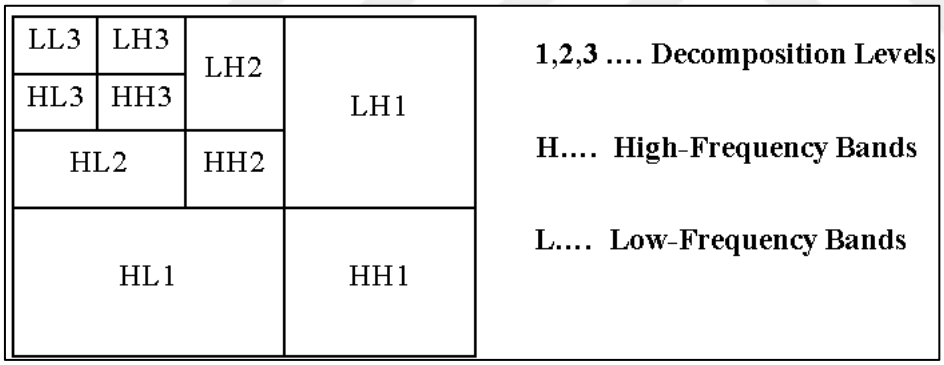

Figure 3.37. 2D-DWT with 3-Level decomposition

The Discrete Wavelet Transform (DWT) is a mathematical technique that summarizes the information from a large set of observations into a smaller set of coefficients. These coefficients are located in the time and frequency domains. DWT is beneficial in cases where the data is contaminated. It decomposes a signal using scaled and shifted versions of a compactly supported basis function. This technique finds application in various fields of life. (Mohideen et al., 2008; Abramovich, 2000; Walker, 1999).

Given a vector of a signal (y) consisting of 2^j observation. The (DWT) of Y is

$$W = wY \tag{3.55}$$

Where W is a $(n\ x\ 1)$ vector consisting of both discrete scaling and wavelet coefficients, the vector of wavelet coefficients can be organized into $j + 1$ vectors.

$$W = [W_1, W_2, \ldots\ldots, W_{j0}, V_{j0}]' \tag{3.56}$$

The given equation is represented as follows: (W_j) is a vector having a length of $(N_j = N/2^j)$, and it represents the wavelet coefficients (Details) that are associated with changes on a scale of length $(\lambda_j = 2^{j-1})$ symbol, known as CD. On the other hand, (V_{j0}) is a vector of scaling coefficients (approximation or smoothing) that is associated with an average on a scale of length $(\lambda_{j0} = 2^{j0})$ symbol, known as CA. (W) is an orthonormal matrix of size $(N * N)$ and is associated with the chosen orthonormal wavelet basis (Antoniadis, 2007; Gencay, 2002)

The DWT coefficients for the observation Y can be defined using the following hierarchy:

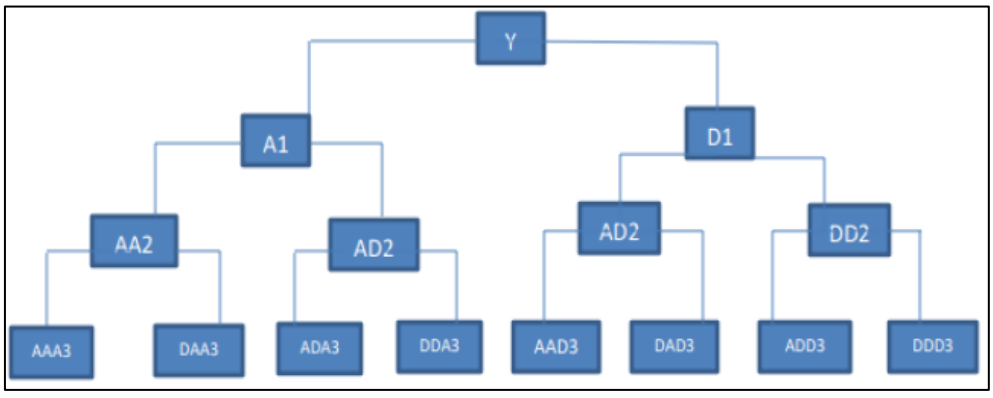

Figure 3.38. The hierarchical process for DWT coefficients

The process of Discrete Wavelet Transform (DWT) involves splitting the approximation coefficients obtained after each iteration into bands using the same filter as before. This adds the details to the most recent decomposition. At each level, the denoised observation can be used to reconstruct the original observation by applying the inverse transform.

$$Y = w'W = \sum_{j=1}^{j0} W_j' W_j + V_{j0}' V_{j0} = \sum_{j=1}^{j0} D_j + S_{j0} \quad (3.57)$$

3.7.4. Thresholding

Thresholding methods

The thresholding method proposed by Donoho for signal estimation uses the wavelet transform's denoising capabilities. It is an efficient and easy-to-use method for removing noise as it eliminates coefficients that are small compared to a specific threshold. The choice of thresholding parameter plays a crucial role in the effectiveness of the denoising process and has a significant impact on the dependent variable. (Nabeel and Alan, 2022; Gencay, 2002).

a) Universal Threshold

The first step in the wavelet threshold denoising process is to estimate the threshold. The universal threshold calculation approach is as follows:

$$\delta_{uni} = \sqrt{2\hat{\sigma}_{MAD}^2 * logN} \quad (3.58)$$

Where N denotes the length of the data series, and $\hat{\sigma}_{\mathrm{MAD}}$ represents the standard deviation estimator of the detail coefficients, which is calculated as (Hamad, 2010):

$$\hat{\sigma}_{\mathrm{MAD}} = \frac{median[|W_{0,1}|, |W_{1,1}|, \dots, \left|W_{1,\frac{N}{2}-1}\right|]}{0.6745} \quad (3.59)$$

b) SURE Threshold

Donoho and Johnstone created the Sure Shrink threshold by combining the Universal threshold and the SURE threshold. The main objective of Sure Shrink is to minimize MSE.

In the case of the soft threshold estimator, a threshold value δ_j is specified for each level j of the wavelet coefficients.

$$SURE(\delta, W) = N - 2 \neq \{j: |W_j| \leq \delta\} - \sum_{j=0}^{d} \min(|W_j|, \delta) \tag{3.60}$$

Given a wavelet coefficient (W_j: $j = 1,2, \ldots, d$) in level $j - th$, select a δ_{SUR} that minimizes SURE (δ, W).

$$\delta_{SUR} = \arg\min SURE(\delta, W) \tag{3.61}$$

Thresholding rules

Soft and hard thresholding are two primary rules utilized in thresholding:

a) Soft Thresholding

Donoho and Johnstone proposed soft thresholding in 1995. This method zeros all signal values smaller than δ and subtracts δ from values larger than δ, which is defined as follows:

$$Wn^{ST} = sign\{Wn\}(|Wn| - \delta) + \tag{3.62}$$

Where

$$sign\{W_n\} = \begin{bmatrix} +1 & if\ W_n > 0 \\ 0 & if\ W_n = 0 \\ -1 & if\ W_n < 0 \end{bmatrix} \tag{3.63}$$

And

$$(|W_n| - \delta)_+ = \begin{bmatrix} (|W_n| - \delta) & if\ (|W_n| - \delta) \geq 0 \\ 0 & if\ (|W_n| - \delta) < 0 \end{bmatrix} \tag{3.64}$$

b) Hard Thresholding

Donoho and Johnstone proposed Hard Thresholding, which is a simple technique for Thresholding that follows the "keep or kill" principle. Hard Thresholding sets all signal values smaller than δ to zero. The wavelet coefficient is then set to the vector $({W_n}^{HT})$ with element (Gencay et al., 2002).

$$ {W_n}^{HT} = \begin{bmatrix} 0 & if\ |W_n| \leq \delta \\ W_n & if\ |W_n| > \delta \end{bmatrix} \tag{3.65} $$

3.7.5. Wavelet with shrinkage as a filter

Wavelet shrinkage is a method used for estimating an underlying signal or observation when it is noisy. To apply this method, the noisy observation is transformed into a wavelet domain, resulting in wavelet coefficients. These coefficients are then shrunk, and the inverse transform of the shrunk coefficients is taken as an estimator of the original observation. Wavelet shrinkage is a well-established field of research with many techniques available in the literature. Some of the main works in this area are Donoho (1993a, 1993b) and Donoho and Johnstone (1994a, 1994b, and 1995), (Alex et al, 2021). This refers to a primary method of reducing wavelet coefficients that Donoho and Johnston proposed. The technique involves setting a threshold value and only keeping the coefficients whose absolute value exceeds this threshold. These significant coefficients are used to create a summary of the transformation. The goal of this approach is to recover an observation, $\tau(t)$, from a noisy observation, $x(t)$.

$$ x(t) = \tau(t) + \zeta(t) \quad t = 0,1, \dots \dots, N-1 \tag{3.66} $$

Where $\zeta(t)$ represent a noise.

Therefore, to summarize the wavelet shrinkage method, the following basic steps should be performed:

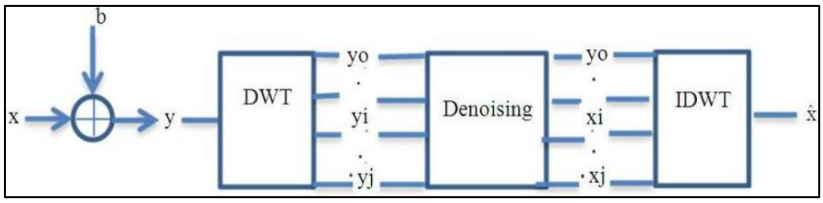

Figure 3.39. Steps wavelet shrinkage

3.8. Performance Evaluation and Variables Importance

In this section, we will explore various approaches to evaluating the efficiency of (ANNs) and (ANNs with Wavelet). We will also discuss the essential factors for determining the significance of independent variables. It is crucial to assess the performance of these models to understand their efficacy in solving complex tasks. Additionally, selecting the most critical independent variables is essential in improving the model's predictive abilities and overall robustness. We aim to illuminate the various techniques available for scrutinizing and refining neural network architectures, enabling practitioners to make well-informed decisions in their modelling pursuits.

3.8.1. Confusion matrix

The classification matrix serves as a statistical gauge of a model's adequacy and its coherence with the dataset. It specializes in the classification of binary events, facilitated by the Confusion Matrix, which presents a comparison between actual and predicted affiliations for each group (Soderstrom and Leitner, 1997). This analytical method operates on the premise that accurate classification of cases by the model, according to a predetermined criterion, provides evidence of its fidelity to the observed data. In essence, this approach offers a structured means of assessing how well the model aligns with real-world observations. The general format of the classification Table (3.4) is: (Ferrer and Wang, 1999)

The Confusion Matrix is an important tool used in predictive analysis in the field of machine learning. It is used to evaluate the performance of classification-based machine learning models. The matrix presents a summary table that shows the number of accurate (correct) and inaccurate predictions made by a classifier, especially for binary classification tasks.

Table 3.4. Confusion matrix

Classification		Observations	
		Negative	Positive
Prediction	Negative	True negative (TN)	False negative (FN)
	Positive	False positive (FP)	True positive (TP)

The confusion matrix is represented by a square matrix in which the columns indicate the actual values, the rows indicate the predicted value, and vice versa.

The confusion matrix helps us see how bad our misclassifications are. The numbers on the main diagonal of this matrix are correct classifications.

True Positive (TP): When the model correctly predicts a positive outcome, and the actual value is positive, it is called a True Positive (TP).

False Positive (FP): In statistical analysis, Type 1 error occurs when the model predicts a positive outcome while the actual result is negative.

False Negative (FN): Here, the model makes a Type 2 error by predicting a negative outcome when the actual value is positive.

True Negative (TN): In these instances, the model accurately predicts a negative outcome when the actual value is indeed negative (Barulina, M. et al, 2023; Visa, et al., 2011).

Accuracy

Accuracy measures the number of correct predictions generated by your model on the entire test dataset. It is an essential indicator for evaluating your model's performance, providing a clear assessment of its effectiveness. However, in cases where datasets are unbalanced, accuracy may not be an effective metric.

$$\text{Accuracy} = \frac{TN + TP}{TN + FP + FN + TP} \tag{3.67}$$

Sensitivity

Sensitivity, also known as True Positive Rate or Recall, is a significant metric that is commonly used in the context of a Confusion Matrix. It measures the percentage of actual positive cases that the model has correctly identified. You can mathematically calculate sensitivity by dividing true positives by the totality of true positives and false negatives. This formula captures the model's ability to accurately detect positive instances, which reflects its effectiveness in recognizing the presence of a particular condition or class.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{3.68}$$

Sensitivity is an important measure in situations were identifying positive cases correctly is critical, such as medical diagnoses, where failing to identify an optimistic case (false negative) could have severe consequences. This metric provides an understanding of how efficiently the model can recognize the true positives within the dataset (Barulina, M. et al, 2023; Visa, et al., 2011).

Specificity

Specificity is a measure that tells us how well a model can identify true negative instances, i.e., the number of cases where the model correctly predicts that a condition is not present. The closer the number of true negative instances (TN) is to the actual number of negative instances (N), the better the model's specificity. A well-trained network should have a specificity value close to one. This measure is useful in identifying how many instances without a condition are predicted by the model not to have the condition.

$$N = TN + FP$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{3.69}$$

Precision (positive prediction value)

The precision metric helps us understand the accuracy of our model in predicting positive cases. It measures the proportion of correctly predicted cases that actually turned out to be

positive. This metric is essential in evaluating the trustworthiness of our model. It is beneficial in situations where minimizing False Positives is more important than minimizing False Negatives (Barulina, M. et al, 2023).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.70}$$

False positive rate (negative prediction value)

False positive rate measures the proportion of negative cases that are incorrectly identified as positive (Ferrer and Wang, 1999; Soderstorm and Leitner, 1997).

$$\text{False Positive rate} = \frac{TN}{TN + FN} \tag{3.71}$$

3.8.2. Receiver operating characteristic curve (ROC curve)

The accuracy of a test is determined not only by its quality but also by how we define positive and negative results. This definition is known as the cut-off point. The cut-off point affects the test's ability to accurately identify true positives, true negatives, false positives, and false negatives.

ROC (Receiver Operating Characteristic) analysis is a method for evaluating and selecting a rating tool based on its effectiveness. It plots sensitivity (the test's ability to detect true positives accurately) against specificity (its ability to detect true negatives accurately) across different cut-off points. The ROC curve shows the balance between sensitivity and specificity, and optimal test performance is achieved when both sensitivity and specificity reach 100% (Bradley, 1997).

The ROC curve is an essential tool in medical research, offering valuable insights into the performance of a diagnostic test:

1. Comparing a test's sensitivity and specificity can reveal trade-offs between the two. An increase in sensitivity is typically accompanied by a decrease in specificity, and vice versa.

2. The curve that approaches the top-left side of the ROC space and reaches the upper limit represents the most accurate test.
3. The slope of the tangent line to the cutoff point provides a great odds ratio for that test value.
4. The area under the curve indicates the test's overall accuracy. A larger area under the curve means better discrimination ability and, thus, higher accuracy.

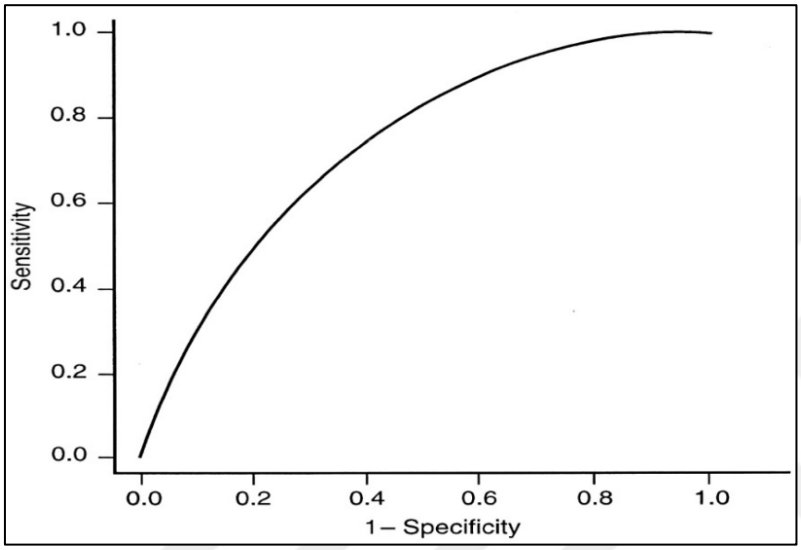

Figure 3.40. Receiver Operating Characteristics (ROC) curve

The area under curve (AUC)

The Area Under the Curve (AUC) is a measurement used to evaluate the overall performance of a classifier across all possible threshold values. When we know the probability distributions for both detection and false alarms, we can construct a ROC curve by plotting the cumulative distribution of probabilities ranging from $(-\infty\ to\ +\infty)$. The area under the ROC curve is typically used as a measure of the quality of probability classification. We calculate this area using the following formula: (Hosmer and Lemshow,2000)

$$A_{ROC} = \int_0^1 \frac{TP}{P}\ d\ \frac{FP}{N} = \frac{1}{PN}\int_0^N TP * dFP \tag{3.72}$$

3.8.3. Cohen's Kappa coefficient

Cohen's Kappa (known as Kappa) provides a useful metric for evaluating the consistency between two binary variables that aim to measure the same phenomenon. Kappa computes the percentage of data values present on the main diagonal of the table and adjusts these values to reflect the agreement that could occur randomly.

To calculate Kappa, we must first establish the observed level of agreement:

$$P_0 = \frac{TP + TN}{N} \tag{3.73}$$

This value should be compared to the expected value if the two raters were completely independent:

$$P_e = \left(\frac{(TP + FP)}{N} * \frac{(TP + FN)}{N}\right) + \left(\frac{(TN + FN)}{N} * \frac{(TN + FP)}{N}\right) \tag{3.74}$$

Then, the value of Kappa is defined as:

$$K = \frac{P_0 - P_e}{1 - P_e} \tag{3.75}$$

When the level of agreement observed between two raters reaches 1, Cohen's Kappa statistic achieves its maximum value. This happens because the numerator matches the denominator. As the observed probability of agreement decreases, both the numerator and denominator decrease accordingly. Although it's theoretically possible for Kappa to be negative, such cases are rare. Below is a breakdown of Kappa's values, along with corresponding estimates for each explanation:

Table 3.5. Interpretation of Kappa (Allouche et al., 2006; Altman, 1991; Carletta, 1996).

Cohen's Kappa	Interpretation
0	No agreement
0.10-0.20	Slight agreement
0.21-0.40	Fair agreement
0.41 0.60	Moderate agreement
0.61-0.80	Substantial agreement
0.81-0.99	Near perfect agreement
1	Perfect agreement

3.9. Variables Importance

Constructing high-performance models requires identifying the most crucial independent variables that explain a significant portion of the variance in the response variable.

Typically, a variable's importance is measured by how much the model's accuracy decreases when that variable is removed. This measure is relative and can vary depending on the modelling algorithm used. To assess the importance of each predictor in a model, one approach is to examine the absolute value of the t-statistic for each parameter.

The VarImp function in the caret package is commonly used to evaluate the importance of variables, especially in general and generalized linear models. This function calculates the (t-statistic) for each model parameter, indicating whether it is significantly different from zero. The VarImp function can be used with other models, such as random forests, to provide an importance score that offers insights into the relative importance of variables within those models as well (Strobl and Zeileis, 2008).

In this thesis, the previous technique was used to find and identify the important variables in the artificial neural network.

The Garson function, introduced by (Garson, 1991), is an algorithm that evaluates the significance of explanatory variables for a single response variable through model weights analysis. This function breaks down the model weights to identify each variable's importance.

The Garson algorithm works by identifying all weighted connections between the layers in the network. It identifies the weights that link the input nodes to the response variable, which passes through the hidden layer. This process is repeated for each explanatory variable, producing a list of weights that are specific to each input variable. These connections are then tallied for each input node and scaled relative to all other inputs. Thus, a single value is generated for each explanatory variable, indicating its association with the response variable in the model.

The Garson function provides results that show the relative importance of variables, presented as absolute magnitudes varying from zero to one. It's crucial to note that this function cannot determine the response's direction. Furthermore, it's limited to assessing neural networks with only a single hidden layer and one output node (Barulina, M. et al, 2023; Beck, 2018; Goh, 1995).

3.10. Methodology

In this section, we present a review of the proposed algorithm for medical mamogram image processing for breast cancer (malignant and benign). For the first time a range of statistical and geometric measurements have been used to analyze and classify regions of interest (ROI). by Using artificial neural networks and wavelet transformation. The classification accuracy has been evaluated by using the confusion matrix with (ROC) curves, the area under a curve (AUC), and Cohen's kappa coefficient. To determine the most important independent variables affecting breast cancer and whether wavelet transformation inclusion gives us a better classification and model than the medical image. The following programs have been used for digital medical analysis:

- Image J 1.45r from (National Institutes of Health, USA).
- R programming language (4.2.1)
- MATLAB (2022)

3.10.1. Data collection

As it has been mentioned earlier, digital image processing is a two-dimensional data source in digital image processing. In this thesis, (150) digital images were collected and stored in

TIFF (8bit) format. (75 of benign and 75 images of malignant tumors) of mammography images of breast cancer.

The data source was as follows:

1. Breast Cancer Diagnostic Center at the Media Diagnostic Center (MDC) in the city of Erbil / for patients over the age of (40) for the period (10/11/2022 to 13/03/2023)
2. The Digital Database for Screening Mammography at the University of South Florida (DDSM) is an extensive database containing many mammograms. Many researchers rely on this rule (DDSM, 2022).
3. Note that all the images were subjected to a laboratory (pathological) examination, and the specialist doctor diagnosed the quality of the mass.

The type of Variables is:

Dependent variable: It consists of variables (benign and malignant) coded as follows:

Y=0 (benign) image

Y=1 (malignant) image

Independent variable: It consists of variables (Energy, Entropy, Area, Mean, StdDev, Mode, Min, Max, X, Y, XM, YM, Perim, BX, Width, Height, Major, Minor, Angle, Circ, Feret, IntDen, Median, Skew, Kurt, Area A, Slice, FeretX, FeretY, FeretAngle, MinFeret, AR, Round, Solidity).

The researcher has processed digital images by extracting measurement (variables) in digital form (data) using the program (Image J 1.8.0), and they are explained and demonstrated in the following steps:

1. Open the program, and the image (Mammography) is required.
2. Determining the (ROI) (benign or malignant tumour) previously determined by clinicians.
3. Determination of statistical and geometric measurements to describe the selected mass shape.

4. Getting the results and saving data in Microsoft Excel (2010).
5. All data is transferred to the (R) program.
6. We then can perform the required statistical analyses.

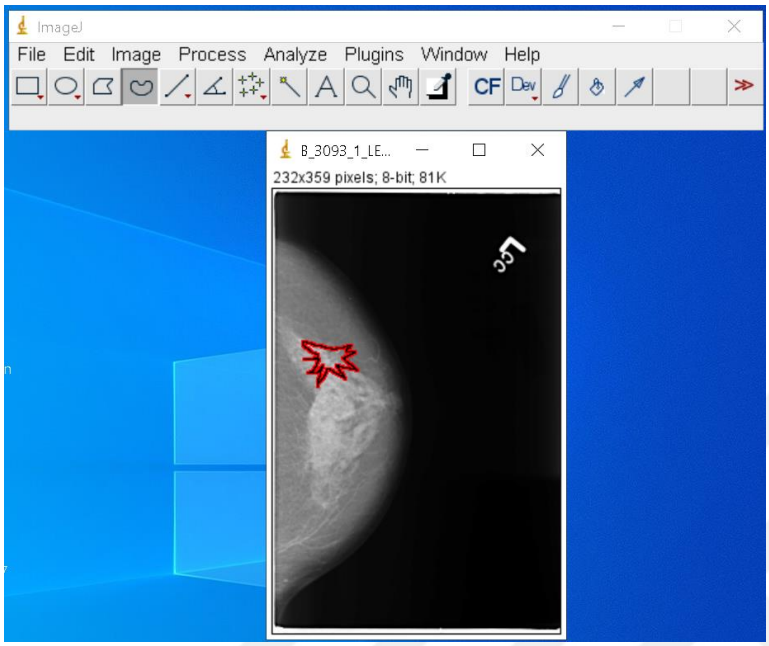

Figure 3.41. Select ROI of (Mammography) images

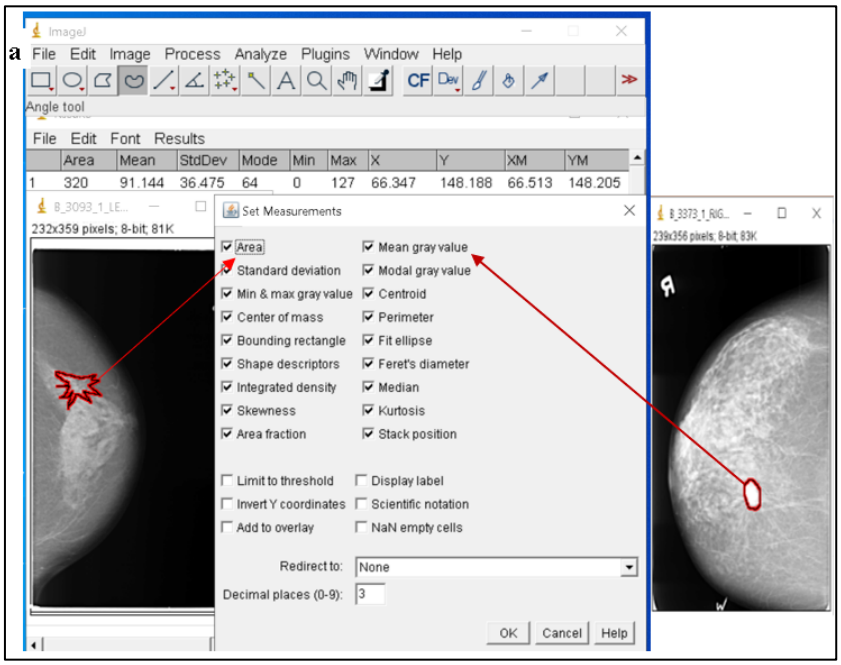

Figure 3.42. Determine statistical measure of ROI

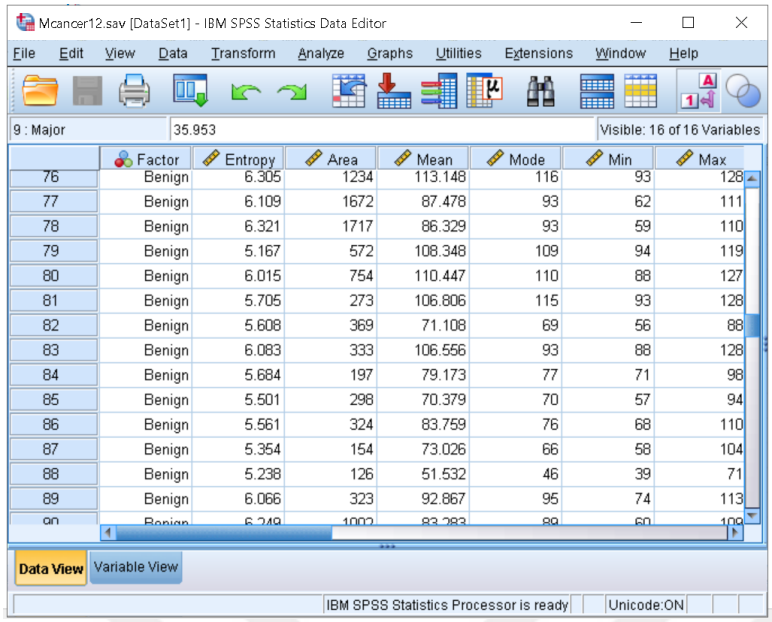

Figure 3.43. Algorithm can be organized for analyzing medical digital images (Malignant and Benign)

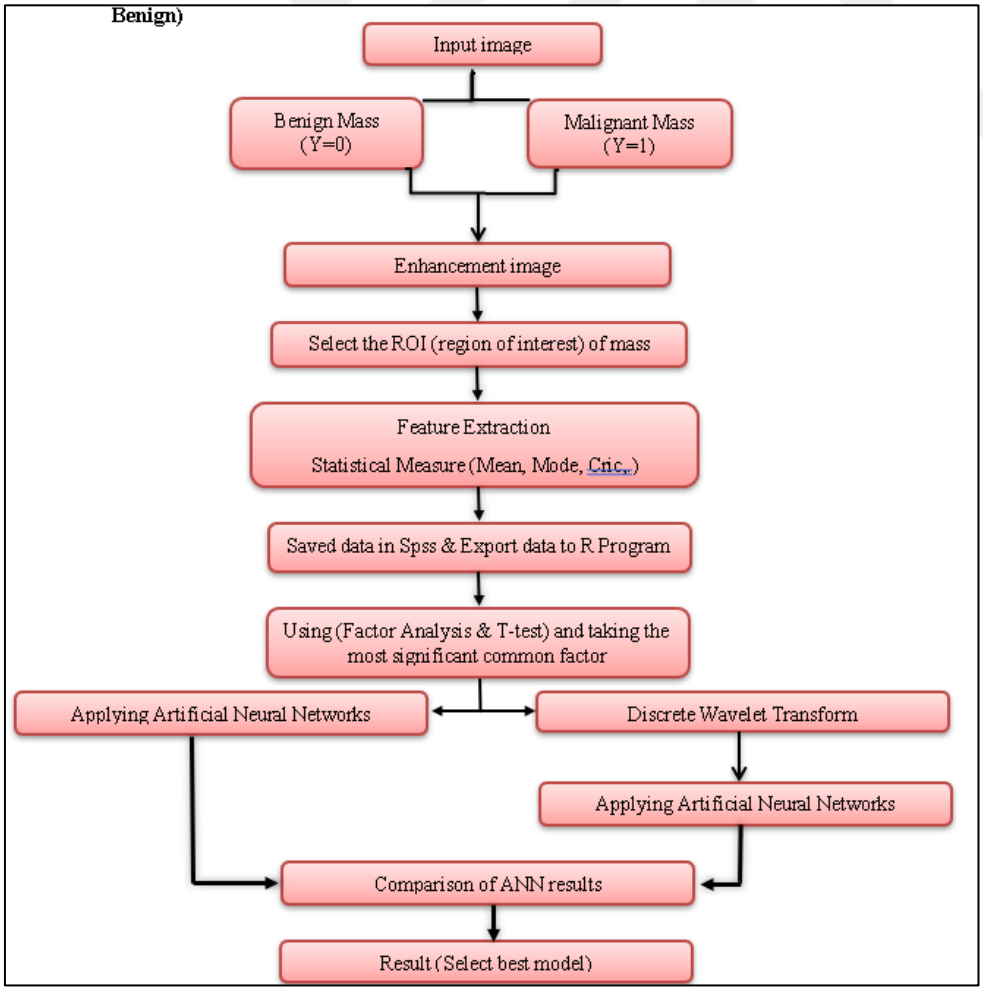

Figure 3.44. Input image

3.10.2. Proposed methods

We used artificial neural network (ANN) techniques to address contamination issues by employing ANN wavelet transformations. This approach involves using wavelets treated with a threshold to reduce the effects of contamination, such as outliers and noise. The transformed outputs are then used to perform an inverse of the Discrete Wavelet Transformation (DWT) and reach the filtered data. This filtered data is then used to estimate parameters for an artificial neural network model.

To effectively remove outliers or noise from response variable observations, a common practice is to use thresholding methods such as hard or soft thresholding. This involves shrinking the detail coefficients obtained from the original observations, which are split into two components using wavelets. The first component represents the sum of the coefficient details, while the second represents the smoothing coefficients, as determined by the artificial neural network.

The threshold level is estimated by one of the well-known methods, including the fixed-from-threshold method. Soft thresholding is then applied to the DWT coefficients, with the remaining coefficients returned to their vector elements. The modified wavelet coefficients facilitate this process.

Depending on the choice of wavelet matrix (db7), the processed dependent variable observations are obtained. These observations, along with independent variables, are then used to estimate parameters for the artificial neural network model.

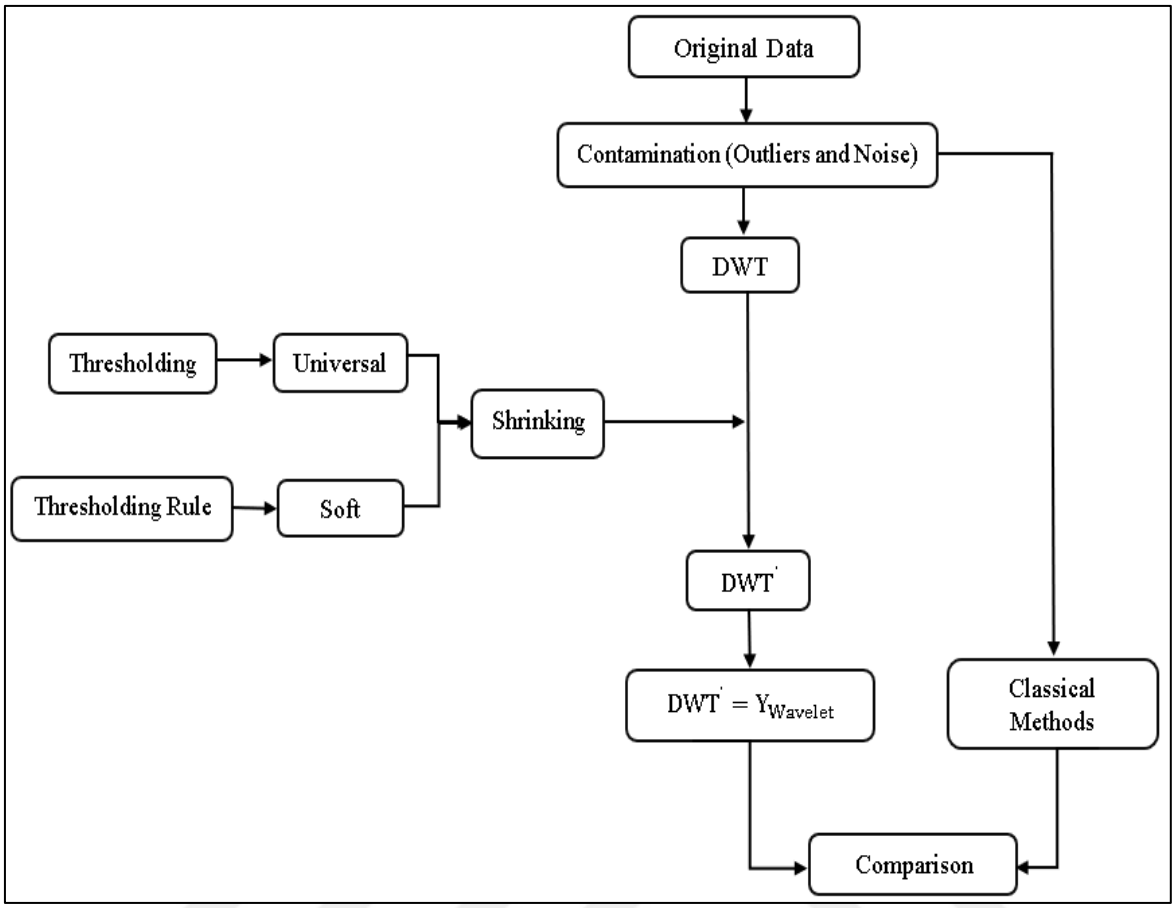

Figure 3.45. Proposed algorithm to compare the proposed and classical methods

4. RESULT

4.1. Applied T-Test

Identify the essential variable and which variable is significantly or insignificantly different between the two masses (malignant and benign). We use the (T-Test).

$H_0: \mu_{Malignant} = \mu_{Benign}$

$H_1: \mu_{\text{Malignant}} \neq \mu_{\text{Benign}}$

We reject the null hypothesis if it is ($0.05 \leq$ Sig), that is, the differences are significant between the two types, and on the contrary, we accept the null hypothesis, and the results of the t-test can be observed for the two independent samples (benign and malignant), as shown in Table (4.1).

Table 4.1. Compares the means of measurements using (t-test)

Factor		N	Mean	t-test	p-value	Factor		N	Mean	t-test	p-value
Energy	Malignant	75	0.016	-0.463	0.644	Major	Malignant	75	38.223	4.558	0.000
	Benign	75	0.018				Benign	75	24.981		
Contrast	Malignant	75	427.149	-0.479	0.633	Minor	Malignant	75	24.914	4.352	0.000
	Benign	75	455.549				Benign	75	17.418		
Entropy	Malignant	75	5.680	2.687	0.008	Angle	Malignant	75	91.537	0.424	0.672
	Benign	75	5.332				Benign	75	88.235		
Area	Malignant	75	924.350	4.039	0.000	Circ.	Malignant	75	0.368	0.424	0.002
	Benign	75	424.920				Benign	75	0.482		
Mean	Malignant	75	110.865	7.919	0.000	Feret	Malignant	75	46.972	5.181	0.000
	Benign	75	92.611				Benign	75	28.361		
StdDev	Malignant	75	5.667	-1.829	0.069	IntDen	Malignant	75	104866.990	4.727	0.000
	Benign	75	6.931				Benign	75	40009.250		
Mode	Malignant	75	111.710	6.962	0.000	Median	Malignant	75	111.130	7.402	0.000
	Benign	75	90.790				Benign	75	92.520		
Min	Malignant	75	93.770	5.655	0.000	Skew	Malignant	75	-0.204	-2.782	0.006
	Benign	75	76.190				Benign	75	0.077		
Max	Malignant	75	123.210	7.064	0.000	Kurt	Malignant	75	-0.116	-1.128	0.261
	Benign	75	109.850				Benign	75	0.178		
X	Malignant	75	109.161	1.299	0.196	Slice	Malignant	75	26.670	-3.433	0.001
	Benign	75	97.523				Benign	75	38.000		
Y	Malignant	75	183.678	0.138	0.891	FeretX	Malignant	75	94.160	0.463	0.644
	Benign	75	182.579				Benign	75	90.050		
XM	Malignant	75	109.171	1.303	0.195	FeretY	Malignant	75	170.970	-0.420	0.675
	Benign	75	97.500				Benign	75	174.560		
YM	Malignant	75	183.698	0.141	0.888	FeretAngle	Malignant	75	87.978	-1.050	0.296
	Benign	75	182.574				Benign	75	96.124		
Perim.	Malignant	75	211.795	4.563	0.000	MinFeret	Malignant	75	30.205	4.978	0.000
	Benign	75	106.063				Benign	75	19.770		
BX	Malignant	75	89.670	0.394	0.694	AR	Malignant	75	1.537	1.374	0.171
	Benign	75	86.210				Benign	75	1.463		
Width	Malignant	75	37.970	5.482	0.000	Round	Malignant	75	0.683	-1.141	0.256
	Benign	75	22.890				Benign	75	0.709		
Height	Malignant	75	39.550	4.419	0.000	Solidity	Malignant	75	0.749	-2.726	0.007
	Benign	75	25.170				Benign	75	0.797		

Table (4.1) show that:

1) There is a significant difference between the measurements of shape and size of the two types and in favour of malignant masses, as its value is greater than the measurements of benign tumors such as (Entropy-Area-Perim-Width-Height- Major-Minor....). This result confirms that the measurements of the malignant masses have a heterogeneous and forked shape (it is irregular in shape) of benign masses.

This result does not contradict the medical side. Instead, it is equivalent to the medical side, as medical research indicates: (Al-Naami et al, 2009; D'Orsi, 1993; Sickles EA, 1986).

2) There is a significant difference between the measurements of circularity, such as (Circ -Feret Angle, Round…), and in favour of benign masses as their value is greater than malignant masses, which means that benign masses tend to be more circular than malignant masses.

4.2. Applied Factor Analysis

Application of principal components analysis to extract the most important variables (measurements)

Principal components analysis was used to identify the most critical variables and measures taken. Extracting them from mammography images of tumors (benign and malignant) and showing the sequence of variables in terms of their importance and impact on the studied phenomenon and the extent of their contribution to explaining the variance and obtaining easily interpretable components by relying on:

1. On the Matrix Correlation, due to the different units of measurement for the variables included in the analysis.
2. (Kaiser) criterion for extracting principal components equal to the number of distinct roots greater than one ($\lambda > 1$).

The results were as follows:

Table 4.2. Test the sample's adequacy and the correlation matrix's significance

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.740
Bartlett's Test of Sphericity	Approx. Chi-Square	13403.638
	df	595
	p-value	2.22e-16

We note in Table (4.2) that two tests were performed before starting to use the (PCA) analysis, which is:

KMO adequacy test

The (KMO) scale is used to determine the adequacy of the sample so that it compares the amounts of correlation coefficients for observation with the quantities of partial correlation coefficients, and the value of the scale falls within the range ($0 \leq KMO \leq 1$). If the scale value exceeds (0.6), the sample is sufficient to conduct the analysis. Here We note that the value of the scale is equal to (0.740), meaning that we can use PCA analysis.

Bartlett's test

The Bartlett test is applied to test the total significance of all correlations within the matrix. It is used to test the hypothesis that the correlation matrix for the community represents the identity matrix, which means that all variables are not correlated. Suppose the test statistic is large and the probability value (Sig.) is minimal. In that case, this indicates that the correlation matrix of the community is not equal to the unit matrix, and in this case, it is recommended to use (PCA). This statistic depends on the (χ^2chi-square transformation) of the determinant of the matrix. We note here that the value of (χ^2) is very large and equal to (13403.638) and with a significant probability value of (<0.001). That is, we reject the hypothesis (that all variables are unrelated).

Table 4.3. interpretation of the total variance of the extracted components

Importance of components:				
	Component 1	Component 2	Component 3	Component 4
Standard deviation	3.3193326	2.1736091	2.0005688	1.79761005
Proportion of variance	0.3147991	0.1349879	0.1143507	0.09232577
Cumulative Proportion	0.3147991	0.4497870	0.5641377	0.65646351
	Component 5	Component 6	Component 7	Component 8
Standard deviation	1.52306226	1.3695501	1.28294694	1.05152543
Proportion of variance	0.06627768	0.0535905	0.04702722	0.03159159
Cumulative Proportion	0.72274119	0.7763317	0.82335891	0.85495050
	Component 9	Component 10	Component 11	Component 12
Standard deviation	1.04340542	0.97068572	0.83196279	0.7469809
Proportion of variance	0.03110557	0.02692088	0.01977606	0.0159423
Cumulative Proportion	0.88605607	0.91297695	0.93275301	0.9486953

We note in Table (4.3) the main extracted components and their interpretation ratio for the total variance so that it shows the number of extracted components, eigenvalues, and the percentage of the variance interpretation to it, for example, the eigenvalues of the first component, amounted to (3.3193), which explains (31.4799%) of the total variance, for the original variables and so on for the rest of the extracted components, whose number is (9) (each eigenvalue is greater than one). Moreover, at the same time, the sum of the nine main components together equals (88.6056%) of the total variance, which is a high rate for explaining the variance. The eigenvalues of the last components are close to zero, indicating an unnoticeable linear correlation in the data.

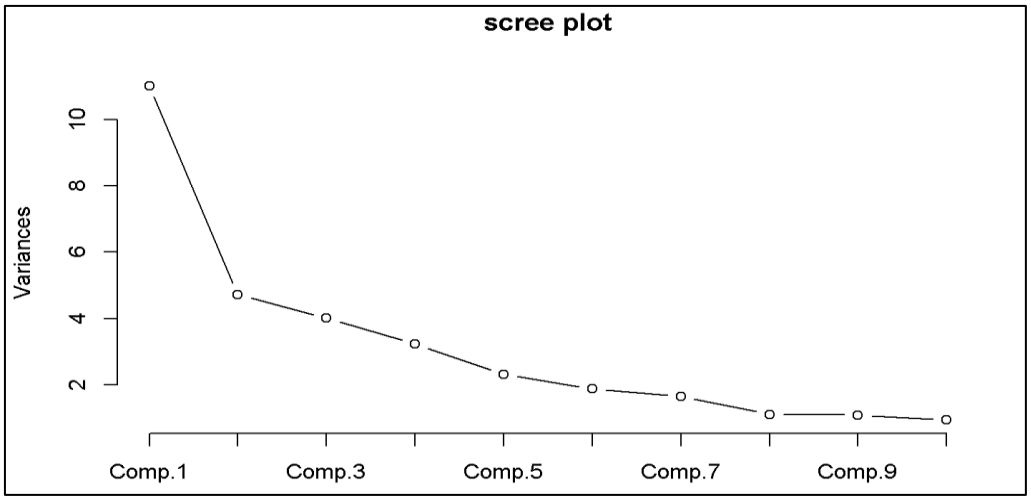

Figure 4.1. Plot screen is a graphical method for extracting the main components.

We notice in Figure (4.1) the plot screen, where the x-axis represents (the number of components) and the y-axis represents (the eigenvalues of each component). The figure suggests that there are (9) main components whose eigenvalues are greater than one ($\lambda>1$) and explain a large percentage of the total variance (88.6056%). The other components were neglected so that we notice in the plot that their distribution is similar as their eigenvalues are less than one ($\lambda<1$), and it falls within the range (0-1). In other words, it determines the significant components. Those located in a very steep area, and until the curve begins to moderate, the components located in this area are neglected.

As for Table (4.4), which represents the Matrix Loading, each load value represents the relationship between the variables and the extracted components, and the load value of each variable was compared with the value (0.5 ±) (to extract highly loaded variables) as a standard value for comparison regardless of the signal, and thus the variables that have a load greater than the value referred to are highly significant and have an impact on the studied phenomenon. The variables within the components were arranged according to their size (in ascending order) and shaded, as shown in the Table above.

Table 4.4. Component matrix

Variables	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7	Factor8	Factor9
Area	0.975								
IntDen	0.963	0.138							
Height	0.942				0.164	0.151			
Major	0.942	0.114			0.180	0.203			
MinFeret	0.937	0.149			0.206		-0.127		
Feret	0.932	0.120			0.160	0.185	-0.142		-0.107
Minor	0.924	0.161			0.211	-0.172			
Width	0.873	0.174			0.194		-0.207		-0.193
Perim	0.860				0.143		-0.326		
Entropy	0.550	0.109			0.743		-0.223		-0.145
Mean	0.244	0.951						0.114	
Median	0.244	0.939						0.124	
Min		0.859						0.399	
Max	0.317	0.851			0.123			-0.220	
Mode	0.276	0.850						0.370	
BX			0.990						
FeretX			0.986						
X	0.140		0.976						
XM	0.140		0.976						
Y				0.982					
YM				0.982					
FeretY				-0.954					-0.197
Energy	-0.287		-0.108		-0.788				
Contrast	-0.172				-0.516			-0.155	
Round	-0.129					-0.960	0.130		
AR	0.145					0.956	-0.133		
Solidity	-0.153					-0.171	0.940		
Circ	-0.412				-0.336	-0.170	0.693		
StdDev	0.161	-0.344		-0.105				-0.886	
Area_A		0.342	-0.100		0.292			0.635	
FeretAngle	-0.119		-0.158						0.710
Angle									0.706
Skew	-0.400	-0.260			0.209		-0.232	0.102	0.134
Slice	-0.396				-0.407		0.164	0.118	
Kurt	0.121	-0.218	-0.274		-0.134		0.101		

The extracted components can be arranged and named as follows:

1. First component: includes the variables and measurements of the shape of the block, which are saturated with a load value ranging between (0.550-0.975), which are high load values, and include:

(Area, IntDen, Height, Major, MinFeret, Feret, Minor, Width, Perim, Entropy)

$$PC_1 = 0.975\text{Area} + 0.963\text{IntDen} + 0.942\text{Height} + 0.942\text{Major} + 0.937\text{MinFeret} + 0.932\text{Feret} + 0.924\text{Minor} + 0.873\text{Width} + 0.860\text{PeriM} + 0.550\text{entropy}$$

It can be called the component of the shape of the masses because it contains the most measures of the shape.

2. Second component includes five variables, the most loading (Mean, Median, Min, Max, Mode). And it can be called the Measures of Central Tendency.
3. Third component: includes four variables that are the most loaded (BX, FeretX, X, XM).
4. Fourth component: includes three variables that are the most loaded (Y, YM, FeretY).
5. Fifth component: includes two variables that are the most loaded (Energy, Contrast).
6. sixth component: includes two variables that are the most loaded (Round, AR). It can be called Circle.
7. Seventh component: includes two variables that are the most loaded (Solidity, Circ). It can be called Circle.
8. Eighth component: includes two variables that are the most loaded (StdDev, Area-A).
9. Ninth component: includes two variables that are the most loaded (FeretAngle, Angle).

Common factors

After factor analysis and t-test, we take common factors between the factor analysis and t-test. as explained in Table (4.5).

Table 4.5. Common factors

Factor Analysis		Two Sample T-Test	
Data	Factor	Data	p-value
Area	0.975	Area	<0.001
IntDen	0.963	IntDen	<0.001
Height	0.942	Height	<0.001
Major	0.942	Major	<0.001
MinFeret	0.937	MinFeret	<0.001
Feret	0.932	Feret	<0.001
Minor	0.924	Minor	<0.001
Width	0.873	Width	<0.001
Perim	0.860	Perim	<0.001
Entropy	0.550	Entropy	0.008
Mean	0.951	Mean	<0.001
Median	0.939	Median	<0.001
Min	0.859	Min	<0.001
Max	0.851	Max	<0.001
Mode	0.850	Mode	<0.001

After extracting common factors, we start using artificial neural networks.

4.3. Data Fragmentation

Our data is randomly divided into a training dataset, and a testing dataset from (70%) of the data (105 patients) selected for the training dataset and (30%) of the data (45 patients) chosen for the testing dataset are selected randomly. Then exit and multiple attempts, the best result is extracted. So that there are no random errors in the method, we use R programming.

The codes are in the appendix.

4.4. Applying Artificial Neural Network in Mammogram Image

The first step before we start analysing the data is to standardize the data. In this case, we have adopted (min-max) standardization. This case (usually called the feature scaling) is to get all the scaled data in the range (0,1) (Graupe, 2013). as follows:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

As mentioned earlier, our data were divided into two groups (test and training), then the neural network was implemented in the training group. The training was repeated five times at 10,000 steps. We used all the hidden layers and the Back-Propagation algorithm to achieve the best result.

Table 4.6. Artificial neural networks summary

hidden: 6	thresh: 0.01	rep: 1/5	steps:	4398	error: 1.04146	time: 0.48 secs
hidden: 6	thresh: 0.01	rep: 2/5	steps:	5100	error: 1.51371	time: 0.57 secs
hidden: 6	thresh: 0.01	rep: 3/5	steps:	6447	error: 1.50883	time: 0.69 secs
hidden: 6	thresh: 0.01	rep: 4/5	steps:	3659	error: 1.54147	time: 0.39 secs
hidden: 6	thresh: 0.01	rep: 5/5	steps:	2867	error: 2.51826	time: 0.31 secs

Table 4.6 shows that we have one hidden layer. Within this layer are six neurons that are repeated five times. In the first iteration and at step (4398), it has the lowest error in the network. which is equal to (1.04) in (0.48) seconds. So, we rely on the first iteration. Moreover, we will plot, analyze, and explains the result.

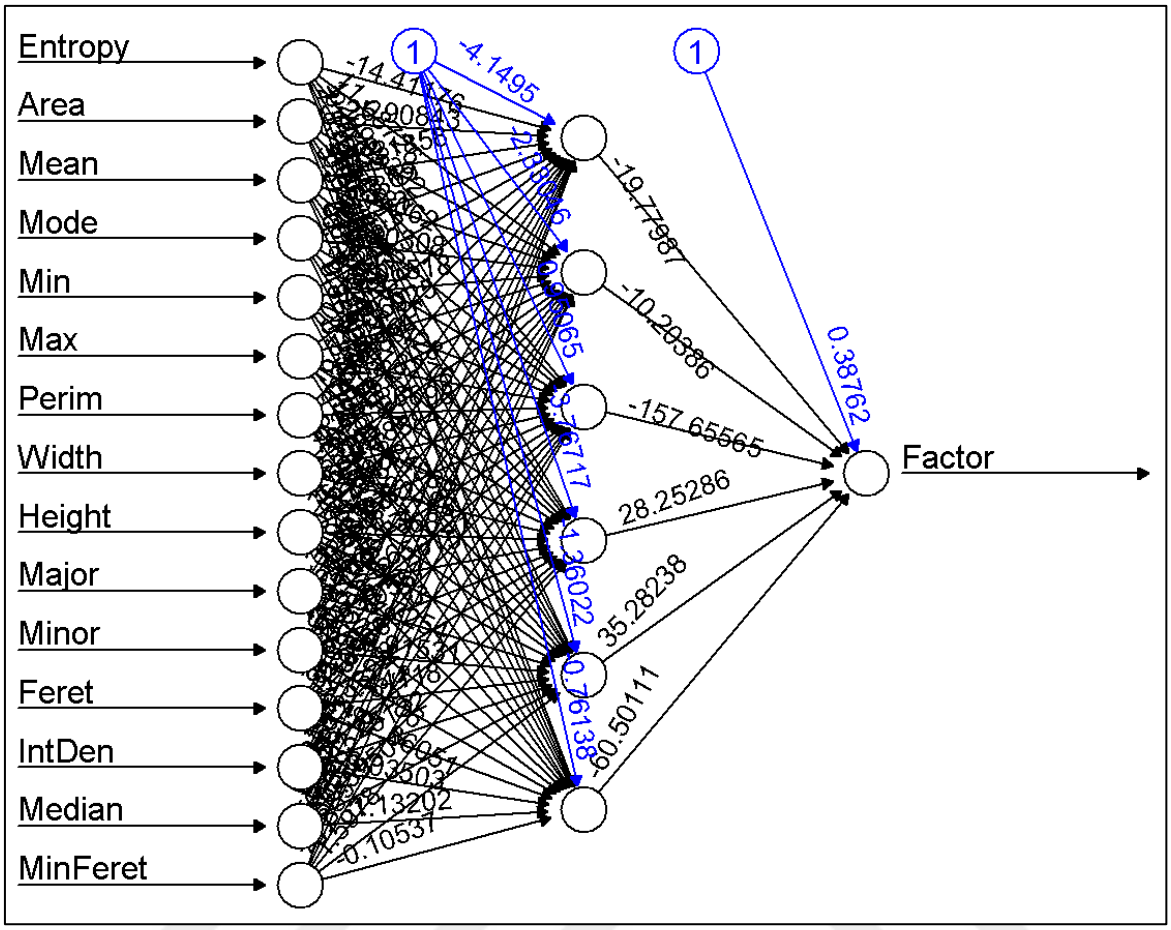

Figure 4.2. Neural network plot

Figure (4.2) shows we have fifteen input, six hidden, and one output neural network. The black lines starting at the input nodes show the relationship between each layer and the weight of each connection. Blue lines (distinguished by number one, these lines start at bias nodes). At each step, they show added bias terms.

We finally have the network trained and ready for use. Now we can check the performance of our classifier. Construct a confusion matrix.

Table 4.7. Confusion matrix and statistics for training dataset of ANNs

Classification		Observation	
		Benign	Malignant
Prediction	Benign	53	1
	Malignant	1	50
Model Accuracy		98.1 %	
Model Sensitivity		98.04 %	
Model Specificity		98.15 %	
Positive Prediction Value		98.04 %	
Negative Prediction Value		98.15 %	
Error Rate of Classification		1.9 %	
Kappa Coefficient		96.19 %	

Table (4.7) shows that the neural network model correctly classified 103 out of 105 cases, which is an excellent rate. The Model Accuracy is equal to (98.1), and the Error Rate of Classification is equal to (1.9), confirming that we have a good classification and model. The Rate of Model Sensitivity is equivalent to (98.04), and the model specificity is equal to (98.15). Another accuracy indicator is the Kappa Coefficient, which shows us how the classification results compare to the values determined by chance. Here the kappa coefficient is equal to (96.19), which obtained a near perfect agreement with the model. Here we find that the artificial neural network models for the training data set have achieved outstanding results.

Table 4.8**.** Confusion matrix and statistics for testing dataset of ANNs

Classification		Observation	
		Benign	Malignant
Prediction	Benign	16	2
	Malignant	6	21
Model Accuracy		82.22 %	
Model Sensitivity		91.30 %	
Model Specificity		72.73 %	
Positive Prediction Value		77.78 %	
Negative Prediction Value		88.89 %	
Error Rate of Classification		17.78 %	
Kappa Coefficient		64.29 %	

Table (4.8) shows that the neural network model correctly classified 37 out of 45 cases, which is an excellent rate. However, compared to the Table (5.7), all the coefficients decreased. In addition, it gives us a good amount of accuracy and sensitivity. The model's accuracy is equal to (82.22), and the model's sensitivity is equivalent to (91.30). This means that our model has a correct prediction ability based on the independent variables proportionately (91.30) for those with breast cancer. The Model Specificity is equal to (72.73), meaning it can make predictions. We also found that positive and negative prediction value proportions are equal to (77.78 and 88.89). Moreover, the Error Rate of Classification is equivalent to (17.78). Also, the kappa coefficient is equal to (64.29), which obtained a substantial agreement with the model. Here we find that the artificial neural network models for the testing data set have achieved good results.

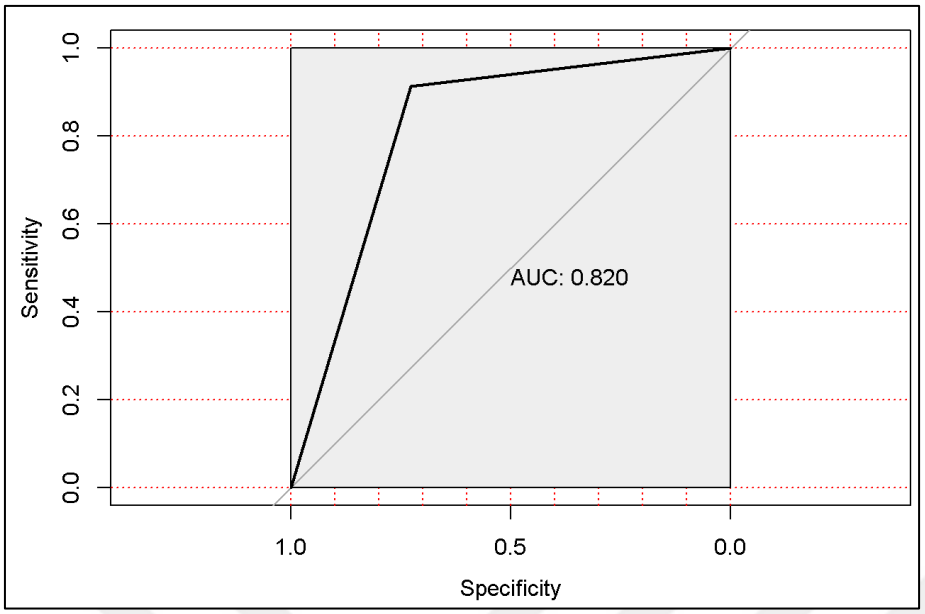

Figure 4.3. (ROC) curve of Testing Dataset for ANNs

Figure (4.3) explains whether the classification is good for the predicted values. It also describes how many values are distinguished and correctly classified by the model. Here the ratio of the area under the curve $(AUC)_{ROC}$ equals (0.820). Where the ROC is equal to (82%). So, the model classified the predicted values mostly well.

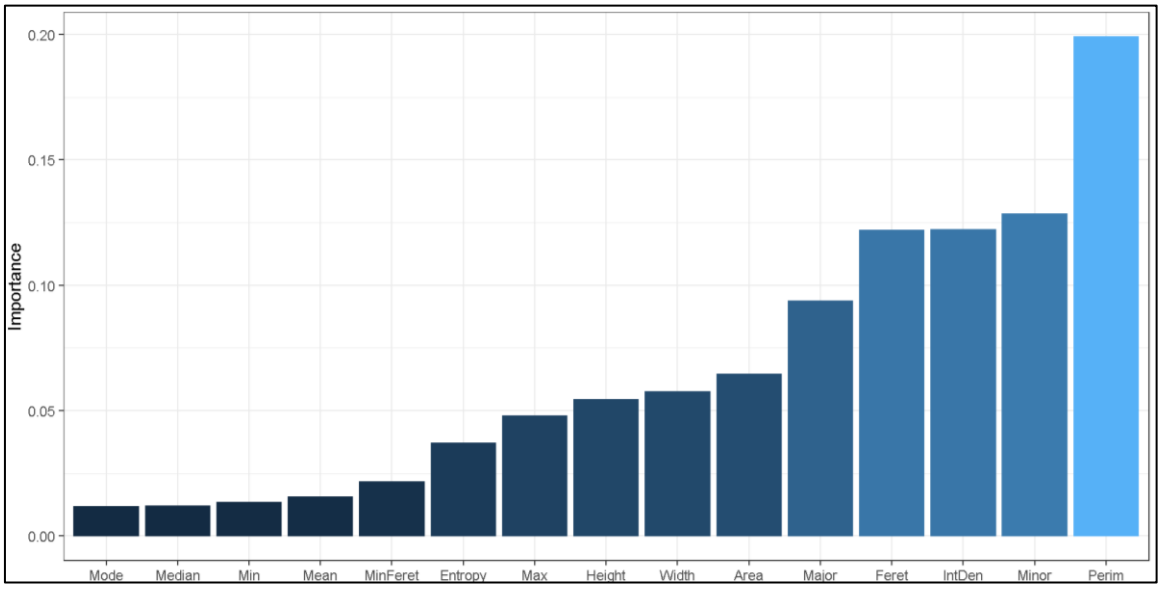

Figure 4.4. Variables importance for ANN

Figure (4.4) shows the degree of importance of each factor or variable affecting breast cancer's morbidity using artificial neural networks. The (Perimeter) factor was the most

influential factor on breast cancer proportionately (0.20) on the dependent variable (Factor). Then each of the variables (Minor), (InDen), and (Feret) proportionately (0.14), (0.13), and (0.13). These independent variables had minor effects: (Mode) and (Median) proportionately (0.01).

4.5. Using Artificial Neural Networks After Adopting Discrete Wavelet Transform

Using the MATLAB program, we obtained wavelet transforms for all images from breast cancer (Benign and malignant). We transferred the data from MATLAB to Spss and then applied the artificial neural networks by R programming. To see if our data classification was better after or before the wavelet transforms. Moreover, see which one will help doctors detect breast cancer earlier.

Table 4.9. Artificial neural networks summary after wavelet

hidden: 8	thresh: 0.01	rep: 1/5	steps: 1754	error: 1.04308	time: 0.33 secs
hidden: 8	thresh: 0.01	rep: 2/5	steps: 2533	error: 0.54558	time: 0.49 secs
hidden: 8	thresh: 0.01	rep: 3/5	steps: 1620	error: 1.01971	time: 0.46 secs
hidden: 8	thresh: 0.01	rep: 4/5	steps: 1869	error: 1.11388	time: 0.31 secs
hidden: 8	thresh: 0.01	rep: 5/5	steps: 3992	error: 0.05741	time: 0.62 secs

Table (4.9) shows that we have one hidden layer. Within this layer are eight neurons that are repeated five times. In the five iterations and at step (3992), it has the lowest error in the network. which is equal to (0.05741) in (0.62) seconds. So, we rely on the five iterations. Moreover, we will plot, analyze, and explains the result.

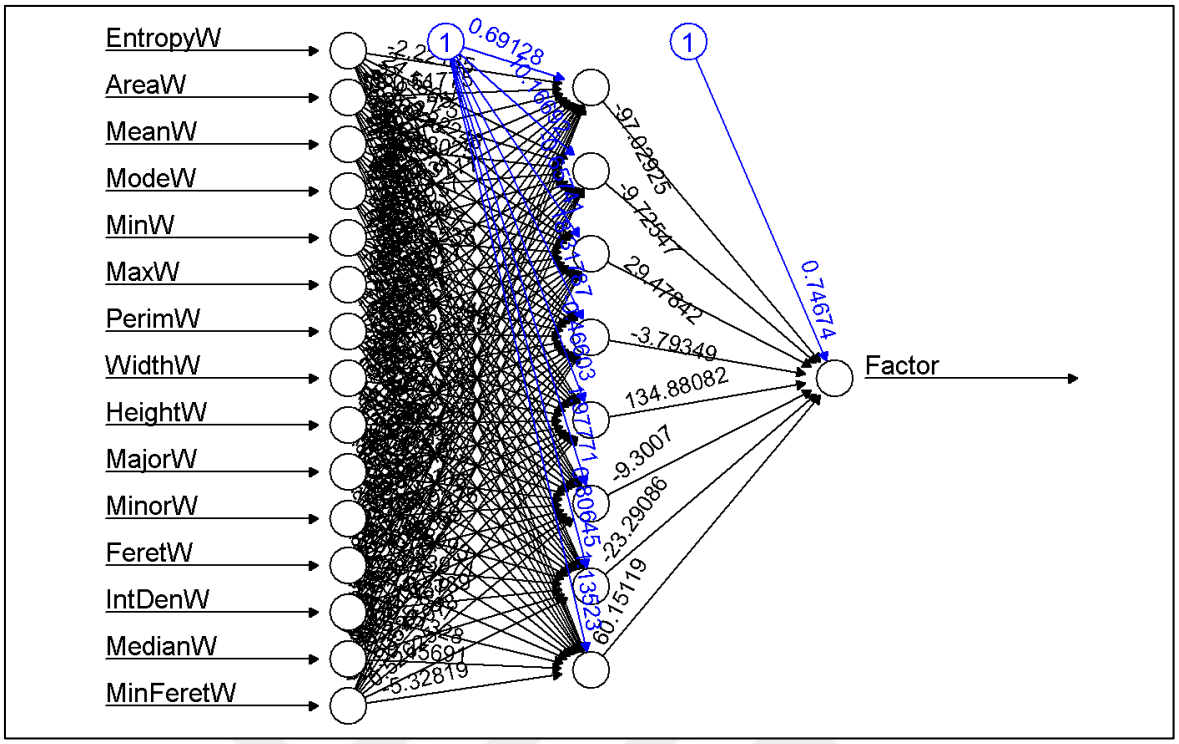

Figure 4.5. Neural network plot

Figure (4.5) shows we have fifteen input, eight hidden, and one output neural network. The black lines starting at the input nodes show the relationship between each layer and the weight of each connection. Blue lines (distinguished by number one, these lines start at bias nodes). At each step, they show added bias terms.

We finally have the network trained and ready for use. Now we can check the performance of our classifier. Construct a confusion matrix.

Table 4.10. Confusion matrix and statistics for training dataset of ANNs after wavelet

Classification		Observation	
		Benign	Malignant
Prediction	Benign	53	0
	Malignant	0	53
Model Accuracy		100 %	
Model Sensitivity		100 %	
Model Specificity		100 %	
Positive Prediction Value		100 %	
Negative Prediction Value		100 %	
Error Rate of Classification		0 %	
Kappa Coefficient		100 %	

Table (4.10) shows that the neural network model correctly classified 106 out of 106 cases, which is an excellent rate. The Model Accuracy is equal to (100%), and the Error Rate of Classification is equal to (0%), confirming that we have an excellent classification and model. The Rate of Model Sensitivity is equivalent to (100%), and the model specificity is equal to (100%). Another accuracy indicator is the Kappa Coefficient, which shows us how the classification results compare to the values determined by chance. Here the kappa coefficient is equal to (100%), which obtained a perfect agreement with the model. Here we find that the artificial neural network models for the training data set have achieved outstanding results. This is much better than the results before the wavelet transformation.

Table 4.11. Confusion matrix and statistics for testing dataset of ANNs after wavelet

Classification		Observation	
		Benign	Malignant
Prediction	Benign	21	2
	Malignant	1	20
Model Accuracy		93.18 %	
Model Sensitivity		90.91 %	
Model Specificity		95.45 %	
Positive Prediction Value		95.24 %	
Negative Prediction Value		91.30 %	
Error Rate of Classification		6.82 %	
Kappa Coefficient		86.36 %	

Table (4.11) shows that the neural network model correctly classified 41 out of 44 cases, which is an excellent rate. However, compared to the table (6.10), all the coefficients decreased. In addition, it gives us a very good amount of accuracy and sensitivity. The model's accuracy is equal to (93.18%), which is an exciting result. and the model's sensitivity is equivalent to (91.30%). This means that our model has a correct prediction ability based on the independent variables proportionately (91.30%) for those with breast cancer. The Model Specificity is equal to (95.45), meaning it can make predictions. We also found that positive and negative prediction value proportions are equal to (95.24% and 91.30%). Moreover, the Error Rate of Classification is equivalent to (6.82%). Also, the kappa coefficient is equal to (86.36%), which obtained a near-perfect agreement with the model. Here we find that the artificial neural network models for the testing data set have achieved excellent results. This is much better than the results before the wavelet transformation.

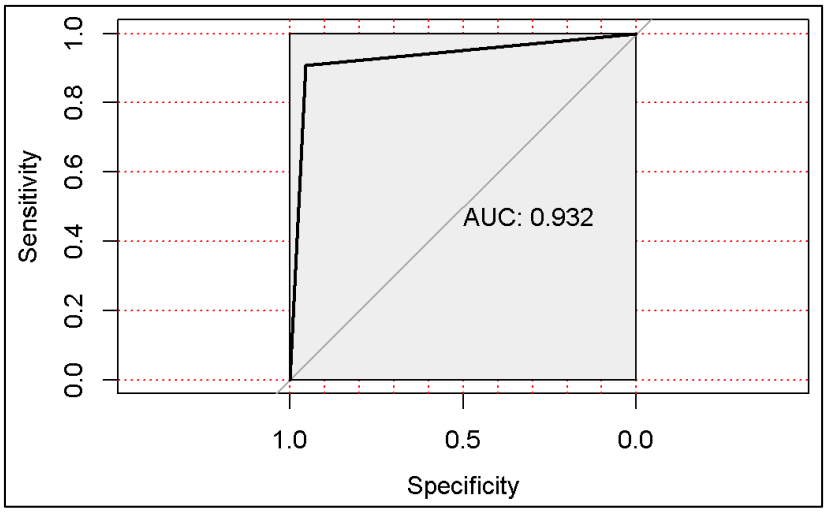

Figure 4.6. (ROC) curve of testing dataset for ANNs

Figure (4.6) explains whether the classification is good for the predicted values. It also describes how many values are distinguished and correctly classified by the model. Here the ratio of the area under the curve $(AUC)_{ROC}$ quals (0.932). Where the ROC is equal to (93.2%). So, the model classified most of the predicted values very well.

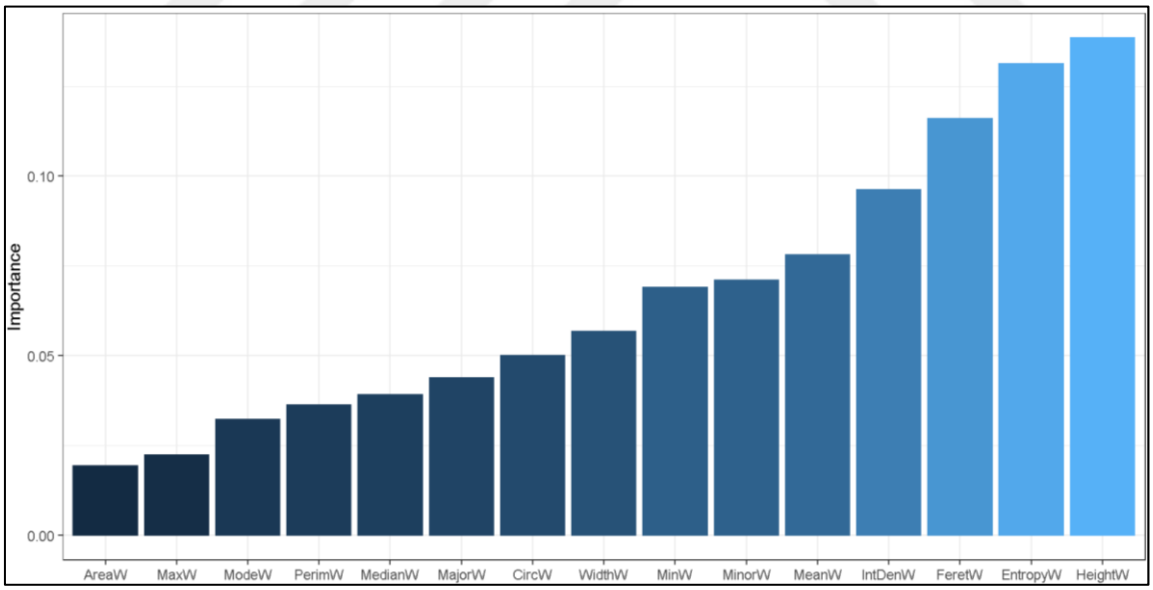

Figure 4.7. Variables importance for ANN after wavelet

Figure (4.7) shows the degree of importance of each factor or variable affecting breast cancer's morbidity using artificial neural networks. The (Heigh) factor was the most influential factor on breast cancer proportionately (0.20) on the dependent variable (Factor). Then each of the variables (Entropy), (Feret), and (InDen) proportionately (0.18), (0.16), and

(0.14). These independent variables had minor effects: (Max) and (Area) proportionately (0.01 and 0.02).

4.6. Comparison Between Methods

After finding classifications using (artificial neural networks) and (artificial neural networks after receiving wavelet transforms) based on (model accuracy, model sensitivity, model specificity, classification error, positive and negative prediction, area under the curve (ROC), and kappa coefficient), the results showed that the artificial neural networks method was better, more accurate, and more efficient after adopting the wavelet transform. The following table shows the comparisons for testing the data sets:

Table 4.12. Performance evaluation criteria between methods

Methods	Artificial Neural Networks	Artificial Neural Networks after Wavelet Transform
Model Accuracy	82.22 %	93.18 %
Model Sensitivity	91.30 %	90.91 %
Model Specificity	72.73 %	95.45 %
Positive Prediction Value	77.78 %	95.24 %
Negative Prediction Value	88.89 %	91.30 %
Error Rate of Classification	17.78 %	6.82 %
$(\mathrm{AUC})_{\mathrm{ROC}}$	82.0 %	93.2 %
Kappa Coefficient	64.29 %	86.36 %

Table (4.12) shows that the model accuracy in ANN is equal to (82.22) but in ANN after discrete wavelet transform is equal to (93.18). So, the model accuracy in Ann after wavelet is much better than in ANN. The model sensitivity in ANN is equal to (91.30), but in ANN after the wavelet transform is equal to (90.91). The sensitivity of ANN is better. The model property in Ann is equal to (72.73), but in ANN after wavelet is equal to (95.45), so ANN after wavelet is much better than Ann. We found positive and negative predictions in ANN after wavelet are equal to (95.24 and 91.30), giving us a more accurate prediction classification than Ann. Moreover, the ROC curve in ANN after wavelet gave us better

predictive values than in Ann. The kappa coefficient of ANN after wavelet is equal to (86.36) more perfect than Ann's.

This shows that ANN after wavelet is much better than ANN by all measures. And we rely on the ANN after wavelet classification.

5. CONCLUSIONS AND RECOMMENDATION

5.1. Conclusions

Based on the results and their interpretation, several following conclusions can be drawn:

1. Utilizing both (t-test and Factor analysis) facilitates the selection of significant metrics extracted from digital images(benign and malignant).
2. The research has compared two classification techniques for differentiating between (benign and malignant) mass shapes by using mammogram data. According to the evaluation criteria, the (ANN with DWT) model emerged as the most effective and superior method in this investigation.
3. ANNs were used to analyze breast cancer images, with and without DWT. Results have shown higher accuracy with DWT, equal to (93.18%) compared to without DWT, which is equal to (82.22%), emphasizing DWT's importance in enhancing predicting and classification performance.
4. The research shows that malignant masses display more significant heterogeneity and irregularity in shape compared to benign masses, which tend to be more rounded. This finding is supported by various medical studies (Al-Naami et al., 2009; D'Orsi, 1993; Rizgar, M. A., 2013; Santos et al., 2021; Sickles, 1986; Yap et al., 2013) and this shows that the medical aspects are on the same paper with the statistical and geometric measures.
5. This study shows that (height, entropy, and Feret) are essential factors that affect the breast shapes. When these factors increase, the breast shape becomes more irregular, which resembles the characteristics of malignancy. On the other hand, decreases in these factors result in a shape that is closer to benign characteristics, which is the same to the interpretations of medical professionals (Al-Naami et al., 2009; D'Orsi, 1993; Rizgar, M. A., 2013; Santos et al., 2021; Sickles, 1986; Yap et al., 2013).
6. The different methods, which are used to classify mass shapes (benign and malignant), have produced varied outcomes, showing that each technology has unique mathematical characteristics. Some techniques perform well in specific applications while not doing as well in others. This was evident while creating the training dataset model for artificial neural networks, complete accuracy was achieved but significantly decreased when

using the testing dataset. In contrast, the (ANNs with DWT) method showed balanced results between the two groups (training and testing).

5.2. Recommendations

The thesis recommends some recommendations and future work, including:

1. We recommend that specialist doctors rely on statistical analyses and measurements of tumours and follow up on the changes occurring during treatment periods to increase their information when making their decisions. A computer program can be designed to extract these statistical measurements and focus on the variables of circularity and the amount of homogeneity or heterogeneity of the masses.
2. Using statistical tests on mammograms or magnetic resonance images (MRI) of (malignant and benign) masses for various diseases to identify and study the differences between the two types.
3. Emphasizing the importance of leveraging comprehensive information databases for women diagnosed with breast cancer and ensuring the availability of skilled personnel proficient in utilizing specialized computer systems tailored for this crucial purpose.
4. Exploring alternative methodologies for processing medical digital images, such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Convolutional Neural Networks (CNN), while conducting comparative analyses to ascertain their efficacy and applicability.
5. Employing time series analysis techniques on mammogram images to monitor treatment progressions among women battling breast cancer, discerning tumour-related changes over successive intervals, and conducting in-depth investigations into their dynamics.
6. Studying the shapes of tumours on mammogram images in terms of linking them to statistical distributions, especially the normal distribution

REFERENCES

Abdolmaleki, P., Buadu, L. D., Naderimansh, H. (2001). Feature extraction and classification of breast cancer on dynamic magnetic resonance imaging using artificial neural network. *Cancer Letters,* 171, 183-191.

Abdolmaleki, P., Dizagi, M., Vahead, M. R. and Gity, M. (2004). Logistic discriminant analysis of breast cancer using ultrasound measurements. *Iranian Journal of Radiation Research,* 2(1), 1-8.

Abdolmaleki, P., Yar Mohammadi, M. and Gity, M. (2004). Comparison of logistic regression and neural network models in predicting the outcome of biopsy in breast cancer from MRI findings. *Iranian Journal of Radiation Research,* 1(4), 217-228.

Abdul Hameed, W., Anuradha, D., Kaspar, S. (2021). Logistic regression and artificial neural network: A comparative study in diagnosing breast cancer. *Research Journal of Pharmacy and Technology,* 14(12), 6330-6334.

Abdul Karim, N. (2010). An identification of persons utilizing the prosperities of the eyes. *AL-Mansour Journal,* 14, 24-25.

Abeloff, M. D., Wolff, A. C., Weber, B. L., Zaks, T. Z., Sacchini, V., and McCormick, B. (2008). Cancer of the breast. In Abeloff M. D., Armitage J. O., Niederhuber J. E., Kastan M. B., MacKenna W. G. (Eds.), *Abeloff's Clinical Oncology* (4th ed.). Churchill Livingstone, Philadelphia, 1875-1943.

Abramovich, F., Bailey, T., Sapatinas, T. (2000). Wavelet analysis and its statistical applications. *The Statistician,* 49(1), 1-29.

Acharya, T., Ray, A. K. (2005). *Image processing: Principles and applications.* New Jersey, USA: John Wiley & Sons, 102-108.

Al-Amri, S., Kalyankar, N. V., Khamitkar, S. D. (2010). Image segmentation by using threshold techniques. *Journal of Computing,* 2(5), 83-86.

Al-Asmari, A. bin Khuzaym (2010). *Introduction to digital image processing.* King Saud University, Scientific Publishing, and Printing Presses, Riyadh, Saudi Arabia, 45-52.

Alex, R. S., Nancy, L., Brani, V. (2021). Bayesian wavelet shrinkage with beta priors. *Computational Statistics,* 36, 1341–1363.

Al-Hinawy, Abdel-Razzaq. (2004). Mechanical measurements to distinguish malignant calcifications from healthy calcifications in digitized breast images. *Damascus University Journal of Engineering Sciences,* 20, 209-225.

Allouche, O., Tsoar, A., Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, Kappa and the True Skill Statistic (TSS): Assessing the accuracy of distribution models. *Journal of Applied Ecology,* 43, 1223-1232.

Al-Naami, B., Adnan, B., Hani, A., Jamal, A. and Abdul-Majeed, A. (2009). Statistical approach for brain cancer classification using a region growing threshold. *Springer Science + Business Media, LLC,* 35, 463-471.

Al-Nuaimi, J. S. (2006). D*iagnosis of brain tumours using hybrid intelligence techniques.* Doctoral Dissertation, College of Computer Science and Mathematics, University of Mosul, Iraq, 85-93.

Altman, D. G. (1991). *Practical statistics for medical research.* London: Chapman and Hall, 75-93.

Ansari, I., Borse, R. (2013). Image processing and analysis. *International Journal of Engineering Research and Applications,* 3(4), 1655-1658.

Antoniadis, A. (2007). Wavelet methods in statistics: Some recent development and their applications. *Statistics Surveys,* 1, 24-28.

Ayer, T., Chhatwal, J., Alagoz, O., Kahn, C., Woods, R. and Burnside, E. (2010). Comparison of logistic regression and artificial neural network models in breast cancer risk estimation. *Radio Graphics Journal,* 30(1), 13-22.

Barulina, M., Okunkov, S., Ulitin, I. and Sanbaev, A. (2023). Sensitivity of modern deep learning neural networks to unbalanced datasets in multiclass classification problems. *Applied Sciences,* 13, 2-22.

Beck, M. W. (2018). Neural net tools: Visualization and analysis tools for neural networks. *Journal of Statistical Software,* 85(11), 1-20.

Beura, S., Majhi, B., Dash, R. (2015). Mammogram classification using two-dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer. *Elsevier Journal,* 154, 1-14.

Bishop, C. M. (1995). *Neural networks for pattern recognition.* Oxford, Birmingham, UK: Clarendon Press, 93-105.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition,* 30(7), 1145-1159.

Calle, E. E., Rodriguez, C., Walker-Thurmond, K. and Thun, M. J. (2003). Overweight, obesity, and mortality from cancer in a prospectively studied cohort of US adults. *New England Journal of Medicine,* 348(17), 1625-1638.

Carletta, J. (1996). Assessing agreement on classification tasks: The Kappa statistic. *Computational Linguistics,* 22, 249-254.

Carol, T., Karen, K. (2005). *The encyclopedia of breast cancer* (Facts on File Library of Health and Living). New York, USA: Facts on File, 102-108.

Chiang, F., Braun, R. (2004). Intelligent failure domain prediction in complex telecommunication networks with hybrid rough sets and adaptive neural nets. *3rd International Information and Telecommunication Symposium Technologies Symposium,* 1-8.

Das, S., Chakraborty, G., Sarkar, T., Kar, S., Roy, S. and Panda, S. (2020). Detecting breast cancer using neural networks. *International Journal for Research in Applied Science & Engineering Technology (IJRASET),* 8, 330-337.

Devita, V. T., Lawrence, T. S., Rosenberg, S. A. (2023). *Cancer principles & practice of oncology* (12th ed.). Wolters Kluwer Lippincott: Williams & Wilkins, 1606-1654.

Donoho, D. L. (1993). Nonlinear wavelet methods of recovery for signals, densities, and spectra from indirect and noisy data. In *Proceedings of Symposia in Applied Mathematics,* 47, 173-205.

Donoho, D. L. (1993). Unconditional bases are optimal bases for data compression and statistical estimation. *Applied and Computational Harmonic Analysis,* 1, 100–115.

Donoho, D. L., Johnstone, I. M. (1994). Ideal denoising in an orthonormal basis chosen from a library of bases. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics,* 319, 1317–1322.

Donoho, D. L., Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika,* 81, 425–455.

Donoho, D. L., Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association,* 90, 1200–1224.

D'Orsi C.J. and Kopans D.B. (1993). Mammographic Feature Analysis. *Seminars in Roentgenology*. 28(3), 204-230.

Dove, E. L. (2004). *Physics of medical imaging - An introduction,* Biomedical Engineering, University of Iowa, USA. Springer-Verlag Berlin Heidelberg, 35-38.

Eric, O. (2011). Particle shape factors and their use in image analysis – Part 1: Theory. *Journal of GXP Compliance,* 15(3), 85-96.

Fausett, L. V. (1994). *Fundamentals of neural networks: Architectures, algorithms and application.* Prentice Hall, Englewood, 75-93.

Ferreira, T., Rasband, W. (2012). ImageJ user guide. *IJ Technologies,* 1, 46.

Ferrer, A. J., Wang, L. (1999). Comparing the classification accuracy among nonparametric, parametric discriminant analysis and logistic regression methods. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada. (ERIC Document Reproduction Service No. ED 432 591).

Fletcher, F. (2002). *Statistical modeling by neural networks*. Doctoral Dissertation, University of South Africa, 78-82.

Flexer, A. (1995). Connectionists and statisticians, friends or foes? In *From Natural to Artificial Neural Computation*. Springer, Berlin, Heidelberg, 454-461.

Furundzic, D., Djordjevic, M., Bekic, A. (1998). Neural networks approach to early breast cancer detection. *Journal of Systems Architecture,* 44(8), 617-633.

Garson, G. D. (1991). Interpreting neural network connection weights. *Artificial Intelligence Expert,* 6(4), 46–51.

Gençay, R. (2001). *An introduction to wavelets and other filtering methods in finance and economics*. New York: Academic Press, 96-160.

Goh, A. T. C. (1995). Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering,* 9(3), 143–151.

Gomes, J., Velho, L. (1997). *Image processing for computer graphics.* New York: Springer, INC, 78-93.

Gonzales, R. C., Woods, R. E. (2018). *Digital image processing* (4th ed.). Pearson, 330 Hudson Street, New York, USA, 102-108.

Graupe, D. (2013). *Principles of artificial neural networks* (3rd ed.). Singapore: World Scientific Publishing Co., 145-162.

Gupta, M. M., Jin, L., Homma, N. (2003). *Static and dynamic neural networks: From fundamentals to advanced theory.* New York: John Wiley & Sons, Inc., 45-63.

Hagan, M. T., Demuth, H. B. and Beale, M. (1996). *Neural network design.* New York: Pws Publishing Co., 85-93.

Hamad, A. S. (2010). *Using some thresholding rules in wavelet shrinkage to denoise signals for simple regression with application in Rezgary hospital – Erbil*. Doctoral Dissertation, College of Administration and Economic, University of Sulaimania, Iraq, 88-93.

Hamad, Y., Simonov, K., Naeem, M. (2018). Breast cancer detection and classification using artificial neural networks. *1st Annual International Conference on Information and Sciences,* 120, 126-131.

Haykin, S. (1999). *Neural networks: A comprehensive foundation* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall, 33-36.

Hosmer, D., Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: John Wiley & Sons, 65-93.

İnternet: American Cancer Society. (2022). Diagnostic mammogram, national breast cancer. *Breast Cancer Early Detection and Diagnosis*. Web: https://www.cancer.org Son Erişim Tarihi: 08/04/2024.

İnternet: Cancer.org (2006). Learn about breast cancer. İnternet: Web: http://www.cancer.org Son Erişim Tarihi: 19 Ekim 2024.

İnternet: Mayo Clinic. (2022). Breast cancer, diagnosis and treatment. İnternet: Web: https://www.mayoclinic.org Son Erişim Tarihi: 26/04/2024.

İnternet: WHO. (2021). World cancer report. *World Health Organization.* İnternet: Web: https://www.iarc.who.int/featured-news/world-cancer-day-2021/ Son Erişim Tarihi: 20/03/2024.

Jain, A. (1989). *Fundamentals of digital image processing.* New Delhi: Prentice Hall of India, 55-63.

Jannat, H., Abdulshahed, A., Alturas, A. (2019). *Swish heart diagnosis using wavelet transform and artificial neural networks.* Master Thesis, Misurata University, Libya, 102-106.

Jemal, A., Siegel, R., Xu, J. and Ward, E. (2010). Cancer statistics 2010. *CA: A Cancer Journal for Clinicians,* 60(5), 277-300.

Khandezamin, Z., Naderan, M., Rashti, M. (2020). Detection and classification of breast cancer using logistic regression feature selection and GMDH classifier. *Journal of Biomedical Informatics,* 111, 103591.

Kopans, D. (1989). Mammography. In *Breast Imaging*. Philadelphia, PA, USA: JB Lippincott Company, 34-59.

Kumar, T., Verma, K. (2010). A theory based on conversion of RGB image to gray image. *International Journal of Computer Applications,* 7(2), 7-10.

Leon, A. D. (1989). A prospective study of the independent effects of parity and age at first birth on breast cancer incidence in England and Wales. *International Journal of Cancer,* 43(6), 986-991.

Levkowitz, H. (1997). *Color theory and modeling for computer graphics, visualization, and multimedia applications.* Berlin: Kluwer Academic Publishers, 402.

Lillesand, T. M., Kiefer, R. W. (2000). *Remote sensing and image interpretation* (4th ed.). New York: Wiley & Sons, 724.

Ammar, M. (1992). *Digital image processing,* R. Gonzalez and P. Wintz (Translation from English to Arabic, 728 pages). The Arabic Center for Arabization, Syria: Translation, Authorization and Publishing, Damascus, 46-48.

Mohammed, A., Ahmed, A., Mohammed, W., Viju, G. and Taha, M. (2020). Mammogram images classification using linear discriminant analysis technique. *International Research Journal of Engineering and Technology (IRJET),* 7(6), 6656.

Mohideen, S. K., Perumal, S. A., Sathik, M. M. (2008). Image de-noising using discrete wavelet transform. *IJCSNS International Journal of Computer Science and Network Security,* 8(1), 213-216.

Muhammad, Z., Sami, A. (2014). Study of bone diseases using wavelet transform and artificial neural networks. *Iraqi Journal of Information Technology,* 6(3), 16-31.

Murugaraja, S., Balamurali, K. (2014). An efficient face tracker using active shape model. *Journal of Computer Engineering,* 16(2), 91-96.

Nabeel, G. S., Alan, G. R. (2022). Using a proposed method for wavelet shrinkage to estimate the tuning parameter of penalized linear regression. *Palarch's Journal of Archaeology of Egypt/Egyptology,* 19(3), 424-439.

Omer, A., Rizgar, A. (2019). Prediction and Factors Affecting of Chronic Kidney Disease Diagnosis using Artificial Neural Networks Model and Logistic Regression Model. *Iraqi Journal of Statistical Science*, (28), 1-19.

Panchal, G., Ganatra, A. (2011). Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers. *International Journal of Computer Theory and Engineering,* 3(2), 332-337.

Pearson, D. (1991). *Image processing.* A volume of the Essex Series in Telecommunication and Information Systems. New York: McGraw-Hill International Editions, 46-52.

Poznyak, A., Sancbez, E. N., Yu, W. (2001). *Differential neural networks for robust nonlinear control.* Singapore: World Scientific Publishing Co., 99-103.

Rabunal, J. R., Dorado, J. (2006). *Artificial neural networks in real-life applications.* New York: IDEA Group Publishing, Hershey, 85-93.

Ragab, D., Sharkas, M., Al-Sharkawy, M. (2013). A comparison between support vector machine and artificial neural network for breast cancer detection. *Recent Advances in Circuits, Communications and Signal Processing,* 171-176.

Rashed, E., Awad, M. (2006). Neural networks approach for mammography diagnosis using wavelets features. *First Canadian Student Conference on Biomedical Computing,* Egypt.

Rizgar, M. A. (2013). *Employment of principal component and discriminant function in digital image processing to distinguish between benign and malignant breast cancer in Erbil.* Doctoral Dissertation, College of Administration and Economics, Salahaddin University, Erbil, Iraq, 58-72.

Rosenthal, M. S. (2000). *The breastfeeding sourcebook* (3rd ed.). Lincolnwood: Lowell House, 46-48.

Roses, D. F. (1999). *Breast cancer* (1st ed.). Philadelphia: Churchill Livingstone, 493-502.

Russ, J. C. (2006). *The image processing handbook* (5th ed.). London: CRC Press, 481-526.

Santos, J., Prado, M., Morais, H., Sousa, S., Silva-Pinto, E., Cançado, L. and Neves, B. (2021). Topological vectors as a fingerprinting system for 2D-material flake distributions. *npj 2D Materials and Applications,* 5(1), 51.

Sarle, W. S. (1994). *Neural networks and statistical models.* In Proceedings of the Nineteenth Annual SAS Users Group International Conferenc, SAS Institute Inc., Cary, NC, USA, 1538-1550.

Sarowar, G. N., Nizamuddin, M. A., Hamid, S. M., Nafiz, I. B. and Mahmud, A. (2009). Enhancing Bengali character recognition process applying heuristics on neural network. *IJCSNS International Journal of Computer Science and Network Security,* 9(6), 154-158.

Scott, E. Umbaugh (1998). *Computer vision and image processing: A practical approach using CVIP tools.* New York: MacGraw-Hill, 85-102.

Sehrawat, D., Abhishek, S., Jaiswal, D. and Sen, A. (2017). Detection and classification of tumor in mammograms using discrete wavelet transform and support vector machine. *International Research Journal of Engineering and Technology (IRJET),* 4(5), 1328.

Sepandi, M., Taghdir, M., Rezaianzadeh, A. and Rahimikazerooni, S. (2018). Assessing breast cancer risk with an artificial neural network. *Asian Pacific Journal of Cancer Prevention,* 19(4), 1017–1019.

Shankar, T. N. (2008). *Neural networks.* India: Laxmi Publications, 122-136.

Sickles EA. (1986). Breast Calcifications: Mammographic Evaluation. Radiology. 160, 289-293.

Soderstrom, I. R., Leitner, D. W. (1997). *The effects of base rate, selection ratio, sample size and reliability of predictors on predictive efficiency indices associated with logistic regression models.* Presented at the Annual Meeting of Mid-Western Educational Researchers Association, Chicago, IL, 22-36.

Stathakis, D. (2009). How many hidden layers and nodes? *International Journal of Remote Sensing,* 30(8), 2133–2148.

Strobl, C., Zeileis, A. (2008). *Danger: high power! - Exploring the statistical properties of a test for random forest variable importance.* COMPSTAT 2008 - Proceedings in Computational Statistics*,* 17, 59-66.

Surendiran, B and Vadivel, A. (2010). Feature selection using stepwise ANOVA discriminant analysis for mammogram mass classification. *ACEEE International Journal on Signal & Image Processing,* 2(1), 17-19.

Taha, H. A., Dlshad, M. S. (2021). Comparison between wavelet Bayesian and Bayesian estimators to remedy contamination in linear regression model. *Palarch's Journal of Archaeology of Egypt/Egyptology,* 18(10), 3388-3409.

Tahmooresi, M., Afshar, A., Nowshath, K., Bamiah, M. and Rad, B. (2018). Early detection of breast cancer using machine learning techniques. *Journal of Telecommunication, Electronic and Computer Engineering,* 10(2-3), 21-27.

Thune, I., Brenn, T., Lund, E. and Gaard, M. (1997). Physical activity and the risk of breast cancer. *New England Journal of Medicine,* 336(18), 1269-1275.

Tian, D. P. (2013). A review on image feature extraction and representation technique. *International Journal of Multimedia and Ubiquitous Engineering,* 8(4), 385-395.

Tryggvadottir, L., Tulinius, H., Eyfjord, J. E. and Sigurvinsson, T. (2001). Breastfeeding and reduced risk of breast cancer in an Icelandic cohort study. *American Journal of Epidemiology,* 154(1), 37-42.

Umbaugh, E. (2010). *Digital image processing and analyses: Human and computer vision applications with CVIPtools* (2nd ed.). New York: CRC Press, 122-136.

Vijayarani, S., Dhayanand, S. (2015). Kidney disease prediction using SVM and ANN algorithms. *International Journal of Computing and Business Research (IJCBR),* 6(2).

Visa, S., Ramsay, B., Ralescu, A. and Knaap, E. (2011). *Confusion matrix-based feature selection.* In 22th Midwest Artificial Intelligence and Cognitive Science Conference, 710, 120–127.

Wadkar, K. (2019). Breast cancer detection using ANN network and performance analysis with SVM. *International Journal of Computer Engineering & Technology (IJCET),* 10(3), 75-86.

Walker, J. S. (1999). *A primer on wavelets and their scientific applications.* Studies in Advanced Mathematics (1st ed.). Florida, USA: CRC Press, 22-36.

Wander, R., Machado, A., Fatima, L. (2020). Toward an approach using grammars for automatic classification of masses in mammograms. *Computational Intelligence*, 37(3).

Wilamowski, B. M., Irwin, J. D. (2011). *The industrial electronics handbook: Intelligent systems* (2nd ed.). New York, USA: Taylor & Francis Group, 88-93.

Wu, C. H., McLarty, J. W. (2000). *Neural networks and genome informatics.* Methods in Computational Biology and Biochemistry (Vol. 1). Amsterdam: Elsevier Science, 45-63.

Xu, S., Chin, L. (2008). *A novel approach for determining the optimal number of hidden layer neurons for FNN's and its application in data mining*. In 5th International Conference on Information Technology and Application (ICITA), 683-686.

Yap, F. Y., Bui, J. T., Knuttinen, M. G., Walzer, N. M., Cotler, S. J., Owens, C. A. and Gaba, R. C. (2013). Quantitative morphometric analysis of hepatocellular carcinoma: development of a programmed algorithm and preliminary application. *Diagnostic and Interventional Radiology*, 19(2), 97.

APPENDICES

APPENDIX 1. R Programing for Artificial Neural Network

```
library(haven)
cancer33 <- read_sav("C:/Users/karzan/Desktop/Mcancer12.sav")
View(cancer33)
data=cancer33
attach(data)
sapply(data, class)
head(data)
library(caret)
str(data)
##Normalize (Mim-Max 0-1)##
data$Entropy<-(data$Entropy-min(data$Entropy))/(max(data$Entropy)-
min(data$Entropy))
data$Area<-(data$Area-min(data$Area))/(max(data$Area)-min(data$Area))
data$Mean<-(data$Mean-min(data$Mean))/(max(data$Mean)-min(data$Mean))
data$Mode<-(data$Mode-min(data$Mode))/(max(data$Mode)-min(data$Mode))
data$Min<-(data$Min-min(data$Min))/(max(data$Min)-min(data$Min))
data$Max<-(data$Max-min(data$Max))/(max(data$Max)-min(data$Max))
data$Perim<-(data$Perim-min(data$Perim))/(max(data$Perim)-min(data$Perim))
data$Width<-(data$Width-min(data$Width))/(max(data$Width)-min(data$Width))
data$Height<-(data$Height-min(data$Height))/(max(data$Height)-min(data$Height))
data$Major<-(data$Major-min(data$Major))/(max(data$Major)-min(data$Major))
data$Minor<-(data$Minor-min(data$Minor))/(max(data$Minor)-min(data$Minor))
data$Feret<-(data$Feret-min(data$Feret))/(max(data$Feret)-min(data$Feret))
data$IntDen<-(data$IntDen-min(data$IntDen))/(max(data$IntDen)-min(data$IntDen))
data$Median<-(data$Median-min(data$Median))/(max(data$Median)-min(data$Median))
data$MinFeret<-(data$MinFeret-min(data$MinFeret))/(max(data$MinFeret)-
min(data$MinFeret))
### partitioning of the data ###
library(caret)
str(data)
#dataset transformation after Min-Max
```

APPENDIX 1. (continued) R Programing for Artificial Neural Network

```
set.seed(3033)
intrain <- createDataPartition(y = data$Factor, p= 0.7, list = FALSE)
training <- data[intrain,]
testing <- data[-intrain,]
dim(training); dim(testing);
library(neuralnet)
set.seed(825)
n = names(data[2:16])
f = as.formula(paste("Factor ~", paste(n, collapse = " + ")))
net = neuralnet(f,data=training,hidden=6,linear.output=FALSE,lifesign =
            'full',rep=5,algorithm = 'rprop+',stepmax = 10000)
plot(net,rep = 1)
#prediction
output<-compute(net,training[,-1],rep = 1)
p1<-output$net.result
pred1<-ifelse(p1>0.5,1,0)
tab1<-table(pred1,training$Factor)
tab1
1-sum(diag(tab1))/sum(tab1)
g<-factor(pred1)
h<-factor(training$Factor)
confusionMatrix(g,h,positive="1")
output1<-compute(net,testing[,-1],rep = 1)
p2<-output1$net.result
pred2<-ifelse(p2>0.5,1,0)
tab2<-table(pred2,testing$Factor)
tab2
#error rate classification
1-sum(diag(tab2))/sum(tab2)
q<-factor(pred2)
w<-factor(testing$Factor)
```

APPENDIX 1. (continued) R Programing for Artificial Neural Network

```
confusionMatrix(q,w,positive="1")
#Area under curve (ROC)
library(pROC)
RocObj2<-(roc(testing$Factor,pred2))
```

APPENDIX 1. (continued) R Programing for Artificial Neural Network

```
RocObj2
plot.roc(RocObj2)
plot(RocObj2, print.auc=TRUE, auc.polygon=FALSE, grid=c(0.1, 0.2),
    grid.col=c("red", "red"), max.auc.polygon=TRUE,
    auc.polygon.col="blue",print.thres=FALSE)
#importance varibles via garson algorithim
library(RSNNS)
library(NeuralNetTools)
garson(net)
```

APPENDIX 2. Matlab Programing for Discrete Wavelet Transformation

```
clear WST
yre= [ ]
n=150
[ca,cd] = dwt(yre,'db7',n);
%estimate of delta(level of thresholding)for haar wavelet filter
MAD=median(abs(cd));
sigmaMAD=MAD/0.6745;
D=sigmaMAD*((2*log(n))^0.5);
%soft thresholding for haar wavelet filter
W=[ca;cd];
for i=1:n
   if W(i)>0
     signW(i)=1;
   else if W(i)==0
           signW(i)=0;
 else
           signW(i)=-1;
       end
   end
end
 signW';
 for i=1:n
    plus(i)=abs(W(i))-D;
    if plus(i)<0
       plus(i)=0;
    else
       plus(i)=plus(i);
    end
 end
 plus';
for i=1:n
```

APPENDIX 2. (continued) Matlab Programing for Discrete Wavelet Transformation

```
WST(i)=signW(i)*plus(i);
 end
 WST=WST';
cd=WST(n/2+1:n);
  yh1 = idwt(ca,cd,'db7',n)
```

Gazili olmak ayrıcalıktır…