IDOKO JOHN BUSH

MACHINE LEARNING TECHNIQUES FOR BREAST TISSUE CLASSIFICATION

NEU 2017

MACHINE LEARNING TECHNIQUES FOR BREAST TISSUE CLASSIFICATION

A THESIS SUBMITTED TO THE GRADUATE SCHOOL OF APPLIED SCIENCES OF NEAR EAST UNIVERSITY

By
IDOKO JOHN BUSH

In Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Engineering

NICOSIA, 2017

MACHINE LEARNING TECHNIQUES FOR BREAST TISSUE CLASSIFICATION

A THESIS SUBMITTED TO THE

GRADUATE SCHOOL OF APPLIED SCIENCES

OF

NEAR EAST UNIVERSITY

By
IDOKO JOHN BUSH

In Partial Fulfillment of the Requirements for the Degree of Master of Science
in
Computer Engineering

NICOSIA, 2017

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name, Surname: John Bush Idoko

Signature:

Date:

ACKNOWLEDGMENT

I would like to sincerely thank my supervisor Prof. Dr. Rahib Abiyev for his understanding, patience, and guidance throughout my graduate studies at Near East University. His supervision was paramount in providing a well-rounded experience in projecting my long-term career goals. He encouraged me to be confident in everything I do. I graciously thank you for all you have done for me Prof. Dr. Rahib Abiyev.

I would also like to thank every other lecturer in Computer Engineering Department and the Faculty of Engineering at large for given me the opportunity to be their assistant.

Furthermore, I would like to thank my family for their consistent prayers and love even when I am away. Conclusively, I extend a big thank you to my very good friends; Abdulkader Helwan and Samuel Nii Tackie for their prompt responses to my calls.

ABSTRACT

This thesis presents an automated classification of breast tissue using three machine learning techniques: Radial Basis Function Network (RBFN), Naïve Bayes (NB) and Back Propagation Neural Network (BPNN). These three neural network models were considered basically to identify the best model for breast tissue classification after an intense comparison of experimental results. An electrical impedance spectroscopy method was used for data acquisition while RBFN, NB and BPNN were the models used for the execution of the classification task. The approach considered in this thesis is made out of the following steps; feature extraction, feature selection and classification steps. The features were obtained using the electrical impedance spectroscopy (EIS) at the feature extraction stage. These extracted features are area under spectrum, the maximum of spectrum, the normalized area, etc. Information theoretic criterion is the strategy used in the proposed algorithm for feature selection and classification phase executed using the RBFN, NB and BPNN. The performance measure of the framework is the total performance accuracies obtained from the experimental results of the three models. The obtained experimental result depicts that the BPNN outperforms the NB and the RBFN in terms of accuracy in classifying breast tissues, minimum square error reached, and learning time as demonstrated in the experimental results.

Keywords: Breast tissue; electrical impedance spectroscopy; back propagation neural network; naïve Bayes; radial basis function network

ÖZET

Bu tez, üç makine öğrenme tekniğini kullanarak radyal temel işlev ağını (RBFN), Naïve Bayes (NB) Algoritması ve Geri Yayılım Yükselişi Yapan Sinir Ağı'nı (BPNN) kullanarak göğüs dokusunun otomatik olarak sınıflandırılmasını sunmaktadır. Üç sinir ağı modeli temel olarak, deney sonuçlarının yoğun bir şekilde karşılaştırılmasından sonra göğüs dokusu sınıflandırması için en iyi modelin tanımlanması için desteklenmiştir. Veri toplama için bir elektrik impedans spektroskopisi yöntemi kullanılmışken, sınıflandırma görevinin uygulanması için tasarlanan modeller RBFN, NB ve BPNN idi. Bu tezde öne sürülen yaklaşım aşağıdaki adımlardan oluşur; Özellik çıkarımı, özellik seçimi ve sınıflandırma adımları. Özellikler, özellik ekstraksiyon aşamasında elektriksel impedans spektroskopisi (EIS) kullanılarak elde edilmiştir. Çıkarılan bu özellikler, sıfır frekansta (I0) empedans, faz açısının yüksek frekans eğimi, 500KHz'de faz açısı, spektrum altındaki alan, maksimum spektrum, normalize alan, spektral uçlar arasındaki empedans mesafesi, I0'daki impedans ve maksimum frekans noktasının gerçek kısmı ve spektral eğrisinin uzunluğu. Bilgi teorik kriter, RBFN, NB ve BPNN kullanılarak yürütülen özellik seçimi ve sınıflandırma aşaması için önerilen algoritmada kullanılan strateji. Çerçevenin performans ölçütü, üç modelin deneysel sonuçlarından elde edilen toplam performans doğruluklarıdır. Elde edilen deneysel sonuç, göğüs dokularının sınıflandırılmasında doğruluk, minimum karesel hata ve deney süresi sonuçlarında gösterilen öğrenme süresi açısından RBFN'nin NB ve BPNN'den daha iyi performans sergilediğini göstermektedir.

Anahtar Kelimeler: Meme dokusu; Elektriksel impedans spektroskopisi; Radyal temel işlev ağı; Naif Bayes; Geri yayılım sinir ağı

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Overview

Understanding the difference with the various breast tissues may help to have some basic knowledge about the normal structure of the mamma. Woman's breast functionality is to produce milk to feed newborn babies, inside it there are, in general, 15 to 20 sections, called lobes where milk is produced. Each lobe is made of many smaller sections called lobules. Milk from lobe is then carried to the nipple through little vessels called ducts. Skin surrounding the internal is about 1mm to 3mm. Fibrous tissue and fat fill the spaces between the lobules and ducts. Fat can be found in three different regions: subcutaneous, just under the skin, retro-mammary in the back of the breast, and intra-glandular between the glandular structures. Minor presence has nerves, vascular and lymphatic tissue as well as fewer lymph hubs inside mamma. The dataset used in this thesis is located at UCI repository under classification category (http://www.ics.uci.edu/~mlearn/MLRepository.html Retrieved 5 May 2017). The name of the dataset is breast tissue database. The dataset contains information about breast tissue measurements in samples of freshly excised breast tissue using electrical impedance. Considering an extraction of those observations in the entire database, several constraints were put into cognizance. The database consists of 106 instances and each instance belongs to one of the classes. Using electrical impedance spectroscopy, six classes of the freshly excised tissues were studied and they include; fibro-adenoma, Carcinoma, glandular, mastopathy, adipose and connective tissues. Characteristics (feature vectors/input attributes) utilized for the prediction task include: impedance at zero frequency (I0), phase angle at 500KHz, high frequency slope of phase angle, area under spectrum, maximum of spectrum, normalized area, distance between the impedivity at I0 and the real part of the maximum frequency point and the length of the spectral curve and finally, impedance distance between spectral ends. These feature vectors were obtained from the many raw breast tissue features with the aid of electrical impedance spectroscopy feature extraction method; the information theoretic criterion.

1.2 Electrical Impedance Spectroscopy

Electrical impedance spectroscopy will always remain breast tissue classification key screening tool as well as the detection of abnormalities/malignancy, since it allows/permits recognition of tumour before being palpable. (Vacek et al., 2002) exhibited that tumors of the breast extent identified in vermont by examining mamma expanded from 2% to 36% in the period between 1974−1984 and 1995−1999. In addition, already analyzed and suspicious lesions sent for examination, around 25% has been affirmed cancerous tumor while; roughly 75% has been analyzed to be benign tissue. The much rated wrong classification is connected to tediousness of achieving correct analysis as depicted in (Basset and Gold, 1987). For this reason, computerized image analysis plays an essential responsibility to improving issues with diagnosis. Set of tools in Computer-Aided Diagnosis (CAD) systems helps radiologists detect and diagnose new cases. Various late explores exhibited that; while the specificity of the frameworks remained moderately consistent, the affectability of these frameworks has altogether diminished as the thickness of the breast expanded (Ho and Lam, 2003). The dataset used in this paper was deduced from the operations of electrical impedance spectroscopy (EIS) and could be found at the UCI repository.

The electrical impedance procedures have for quite some time been utilized in classifying tissue as well as impedocardiography applications (Kubicek et al., 1970). Some of these strategies have additionally empowered mapping in impedance as seen in (Tachibana et al., 1970) and (Henderson et al., 1978) as well as recently dynamic imaging in (Brown et al., 1994). The equivalent of AC in resistivity for DC equivalent current is known as impedivity/particular impedance. The electric and dielectric properties dictates impedivity of a tissue and this depend, in addition to other things such as; membrane capacitance, cell concentration, intracellular medium and the interstitial space electric conductivity in (Schwan, 1959) as well as (Foster and Schwan, 1989). Some of the good features of impedance techniques includes; minimum invasiveness, easiness and low cost. Initially in the 80s, estimations of electric and dielectric has been performed using tissues of breast within a scope of test settings incorporating into ex-vivo/vivo estimations as well as utilizing different methods of measurement (Surowiec et al., 1988) to (Mad and Heinitz, 1995). Within 488Hz-1MHz level/range, an author in (Jossinet, 1998) found critical contrasts in phase angle and impedivity modulus from among the six tissue

classes of breast. EIS is conceivably applicable in breast cancer detection and breast tissue separation as proposed in the above discoveries. Using EIS, this study demonstrate a strategy for the classification of breast tissues. Feature set utilized in this research work is the same as those features defined in (Jossinet and Lavandier, 1998) and also extra features chose for their separation capacity. Twelve-point and seven-point spectra were used to choose the statistical hierarchical approach. A non-invasive strategy used to measure cells impedance within the scope of frequencies from a tissue surface is termed an electrical impedance spectroscopy. Changes that occur in the nature of tissues are as a result of changes in impedance. Along these lines, fitness level of the fundamental tissue can be demonstrated by the variation of impedances. The above ideology makes electric impedance spectroscopy an essential strategy for detecting/diagnosing irregularities, cancer and abnormalities particularly to diagnose women malignancy (Kerner et al., 2002), (Zheng et al., 2008).

1.3 Brief Review of the Implored Methods

Radial Basis Function Network (RBFN), Naïve Bayes (NB) and Back propagation Neural Network (BPNN) machine learning techniques as implored in this thesis are virtually applicable in situations where a relationship between predicted variables (dependents/outputs) and the predictor variables (independents/inputs) exists, also extending to relationship being difficult to understand and very complex as seen in some of the differences or correlations within sets. In neural network, the type of problem amenable to solution is defined by the way they are trained and the way they work. RBFN, NB and BPNN works by inputting some input variables to the classifier and generating some corresponding output variables. Therefore, they are used in scenarios where there are some known information and probably infer some corresponding unknown information.

1.4 Objective of this Study

In this thesis, the main objective is to train RBFN, NB and BPNN to classify to which set each of the breast tissue belongs when assigned different inputs known to be attributes. First thing needed to execute this task is to have a dataset. This research work considered a dataset taken from UCI repository under classification category. Breast tissue database is the name of the dataset as used in that repository. A number of constraints were considered in the extraction of

these observations from the general database. 106 instances were identified in the database where each instance belongs to one of the classes. Carcinoma, fibro-adenoma, glandular, mastopathy, connective and adipose tissues are the six classes of the freshly excised breast tissues examined by the electrical impedance spectroscopy. This dataset could neither be fed unto any of the proposed machine learning techniques in its original form for classification unless it is first normalized.

1.5 Anatomy and Physiology of the Breast

Considering researches (Bindu et al., 2006), (Lazebnik et al., 2007), and (Sha et al., 2002) it is recorded that the breast of woman is basically made up of three major types of tissue: the glandular tissue, the connective tissue (Cooper's ligament also known as the fibrous strands) and the breast fat (or adipose tissue). Critically, the proportion of these major types of breast tissue varies between persons. The quantity or amount of fat, fibro-glandular tissue and water could varies in different stages due to usual hormonal changes of location, menopause, pregnancy or menstruation as reported in (Bindu, et al., 2006), (Bland et al., 2004). From bioelectrical studies, breast anatomy is analyzed and demonstrated as:

- The adipose tissue layer is located beneath skin, consisting of vesicular cells covered with adipose, connected into lobules then distributed using Cooper's ligament.
- The lobules that produce milk (mammary glands) of the woman's breast is located in the innermost tissue. Also, about 15 to 20 sections of each woman's breast known as lobes having many smaller sections of mammary glands are commonly organized circularly. Thin tubes called lactiferous ducts terminate each section connected to the nipple and ultimately connected to a reservoir (ampulla). The Cooper's ligament surrounds these ducts and lobes.
- The basic function of the Cooper's ligament is to support the tissue attached to the chest wall as well as maintain the inner structure of the breast. Moreover, in (Jossinet, 1998) it is noted that the breast major muscle from pectoralis is separated by the retro-mammary adipose.

In both sagittal and frontal views, a healthy breast's anatomy is depicted in figure 1.1. Intense researches in (Hagness et al., 1999) and (Choi et al., 2004) demonstrated that, as represented in figure 1.1, Despite the fact that lymph hubs does not form the mamma essentially, it is as yet finding in the image to be mamma malignancy could be analyzed via discovery of metastasized

tumor organs where roughly half of breast lesion growth happen, especially in the axillary lymph hubs.

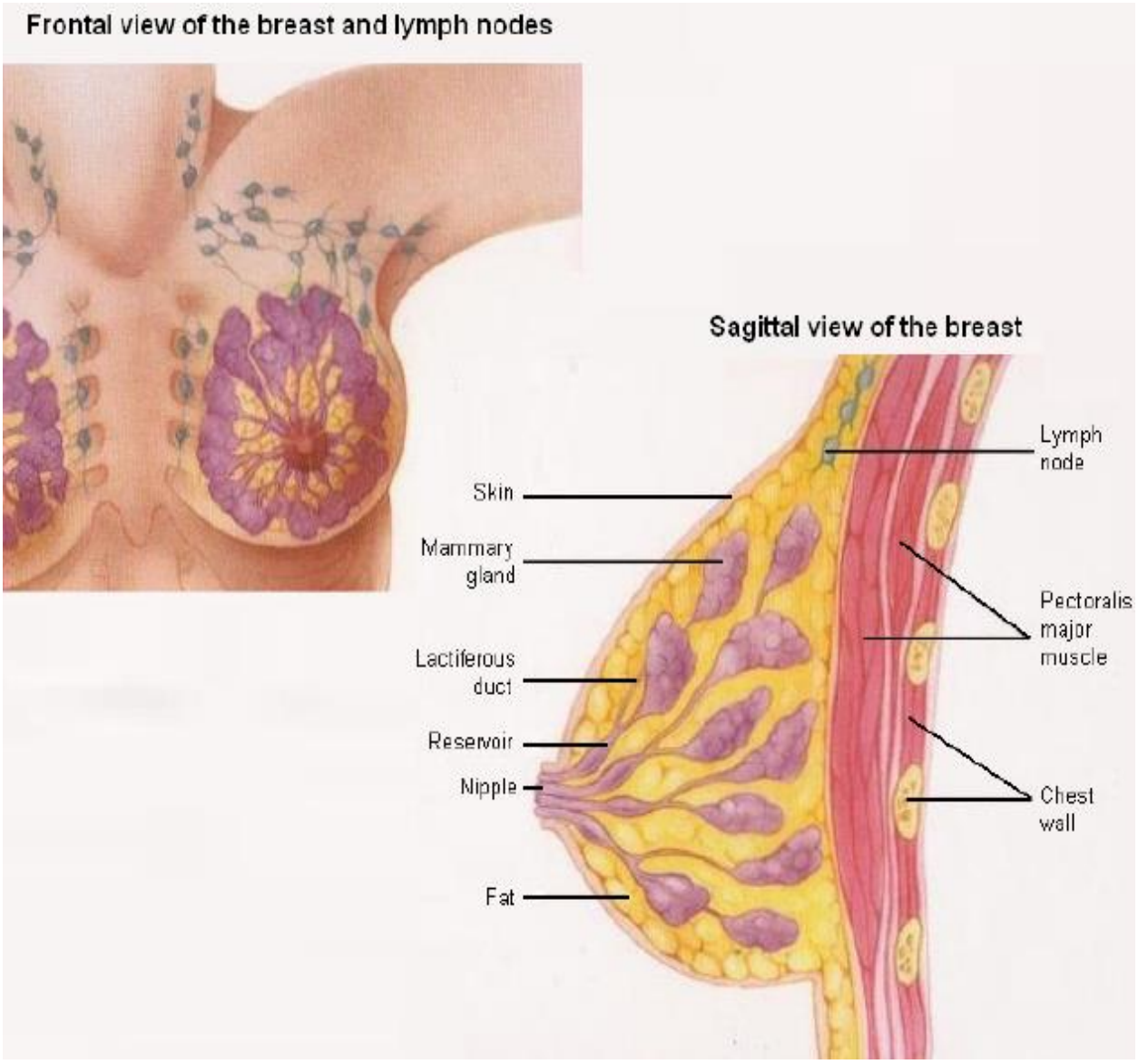

Figure 1.1: Sagittal and frontal views of the breast anatomy (Gorey et al., 2006)

1.6 Tumours Formation

As seen in in the UCI repository, tumours in the breast are defined as the growth of undifferentiated (unspecialized) cells from where lump is formed. Most often, the undifferentiated cell are destroyed by the capability of the immune system and these undifferentiated cells usually a approach know as cell self-destruction (apoptosis) leads to the formation of a tumour. Furthermore, masses of tumours are formed when several mutations take

place within organs in a given period and human cells is unable to respond appropriately as depicted in (http://www.cancerhelp.org.uk/help/default.asp?page=85 Received 10 May 2017). A carcinogen is the generic name for something that induces the mutation of cells, leading to the formation of tumour cells. For the most part, there are two nonexclusive conceivable roots for tumor cells. The first is the presence of oncogenes, which are the qualities in charge of the expansion of cells, and the second is the restraint of the qualities that more often than not control cell multiplication and enable apoptosis with a specific end goal to support a consistent development of cells (Enzinger and Weiss, 1995). The tumour growth simply the proliferation occurrence of tumour cells could demonstrate if a tumourof the breast is malignant or benign. The benign tumours encounters dangers/problems only if there is compression and a push in the nearby organs or when the tumors grows within skull or releases abandoned cells. In other words, cancerous tumors have an uncontrolled growth because of high rate of replication which often by the process of metastases, spread to different regions to destroying good tissues surroundings.

(Bindu et al., 2006) and (Cameron and Pool, 1981) depict several changes suffered by the tumour cell in terms of water state, cell surface, pH, cytoskeleton, viscosity, the inhibition of contact loss, membrane transport, growth regulation, several other factors and temperature.
Some of these changes will affect dielectric properties directly, so these will be studied in more detail. The tumour malignancy level could be obtained from pathological analysis of the premature level of the cells within the breast tumour. The various stages of growth where organs could be seen is known to be differentiation. The disorganization and the decrease of microfilaments and microtubules disorganizes the cytoskeleton of tumour (Cameron and Pool, 1981), causing the mitosis process known as cell replication to become very chaotic and cell original shape to be lost becomes more round leading to both loss of genetic information and an uncontrolled tissue growth. The regular osmosis process and the alteration of membrane permeability is often affected due to changes on the surface of the cells, making the tumour in the breast tissue retaining more fluid than normal organs. And such explains why cancerous cells having bound water shape accommodate more fluid.

Moreover, contact-inhibited are non-cancerous/malignant cells meaning that, the growth of large amount of cancerous organs on each other are piled up on one another creating multiple layers

coexisting in high concentrations. Because of the massive development in tumors organs, particularly in tumors of cancerous cells, networks of capillaries normally generated to properly enhance the newly generated organs as depicted in (Bridges et al., 2002). It is recorded in (Malich et al., 2007) that tumours with a dimension of at least 3mm induce neoangiogenesis. Capillaries networks could develop into arteries and even tiny veins which would join to major blood supply channels as the size of tumours gets larger (Bridges et al., 2002). Hence, the characterization of malignancy grade of a tumour is of great importance in the study of the level of vascularization. The high scattering in microwave imaging is as a result of the increase of water within cancerous tissue. In (Joines, 1984), (Pethig, 1984) and (Sha et al., 2002), it is reported that the increase of water and sodium, specifically in bound water inside an organ of tumour induce large amount of relative permittivity and conductivity within tissues of breast tumour. Another feature which can assist to detecting presence of malignancy in tumours is the existence of calcifications. In general, an occurrence of severe necrosis often leads to the formation of malignant tumours that is; sets of deadly organs that are not formally mixeed by the organism is as a result of disorderly apoptosis as demonstrated in (Sha et al., 2002).

Furthermore, malignant and benign tumours have other characteristics inherent that have demonstrated to be important with regards to different imaging modalities classification. Those inherent attributes/characteristics are mostly based on shape, depth, margins, localization, packing density, size and surface texture (Bridges et al., 2002), (Jossinet, 1998), (Rangayyan et al., 1997), (Davis et al., 2008) and (Malich et al., 2007). Features of a tumour that may be particularly beneficial in the context of classification of MI are basically texture and shape of the tumour surface.

Major characteristics of tumours woth malignancy are:

Asymmetric, irregular shapes and ill-defined;

- Lack of sharpness referred to as blurred boundaries.
- Complex and rough surfaces having micro-lobules or spicules.
- Breast architecture distortion.
- Non-uniform permittivity variations.
- Calcifications and masses caused by irregular increase of tissue density.

Finally, below are the main characteristics of benign tumours:

- Oval, spherical, or well-circumscribed contours presentation.
- Surfaces smoothness.
- Compact (Bridges et al., 2002), (Bindu and Mathew, 2007) to (Guliato et al., 2008).

1.7 Thesis Overview

Remaining part of the thesis is organized in the following ways:

- The second chapter (chapter 2) describes some medical applications of breast tissue where all the six classes of breast tissues are examined, cancerous tissues are differentiated from non-cancerous tissues and possible diagnosis and remedies are presented for the cancerous tissues.
- Chapter 3 presents some related research emanated from breast tissue classification induced by electrical impedance spectroscopy as well as dielectric properties of breast tissues.
- Chapter 4 examines the implored machine learning techniques for the breast tissue classification task where historical and detail analysis of RBFN, NB and BPNN classifiers are presented.
- Chapter 5 presents the system design and experimental result analysis of the three implored models. Experimental result comparison between the three models was made where radial basis function network outperformed naïve Bayes and backpropagation neural network classifiers.
- Finally, the overall conclusions as well as future work suggestions are vividly presented in the last chapter (chapter 6) where it is stated that repetition of the experiment would be made using other machine learning techniques such as co-adaptive neuro-fuzzy inference system (CANFIS), extreme learning machines (ELMs), deep learning and support vector machines (SVMs) to ascertain generalization report as well as a more optimal results.

CHAPTER 2
MEDICAL APPLICATIONS OF BREAST TISSUE

2.1 Breast Tissue Classes Overview

The basic function of the woman's breast is to produce milk for newly born babies' upkeep. In general, there are about 15 to 20 sections. Milk is produced in the major section known as lobes. Smaller sections of the lobe are known as lobules. The ducts (little vessels) are responsible for transferring the milk from the lobe to the nipple. About 1mm to 3mm of skin surrounds the internal components of the breast. Spaces between ducts and lobules are filled with fat and fibrous tissue. Subcutaneous, just under the skin, retro-mammary in the back of the breast, and intra-glandular between the glandular structures are the three major regions where fat could be found. Nerves, vascular and lymphatic tissue as well as a small number of lymph nodes forms the minor regions where fat is found inside the mamma. The six classes of breast tissue examined in this study include; carcinoma, fibro-adenoma, glandular, mastopathy, adipose and connective tissues.

2.2 Carcinoma Class

A type of cancerous tissue developed from epithelial cells is known as carcinoma (Lemoine et al., 2001). Carcinoma is particularly alluded to as a tumor that starts within the tissue lines at the external or inward breast areas and by and large emerges from cells beginning in the ectodermal or endodermal germ layer amid embryogenesis. Carcinogenic tissue happens when a cell's DNA is adjusted or harmed and the cell ends up noticeably malignant and develops wildly. Figure 2.1 depicts carcinoma tissues.

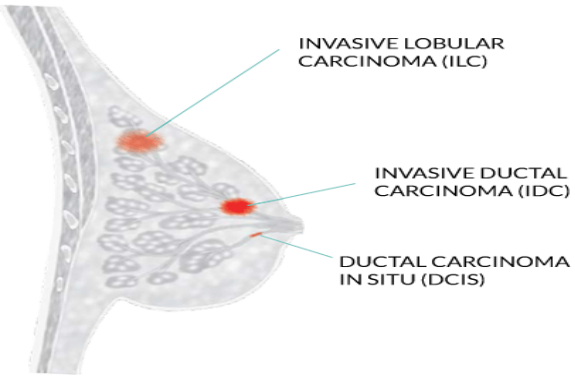

Figure 2.1: Carcinoma breast tissues

2.2.1 Classification of Carcinoma Tissue

No comprehensive and simple classification system as of 2004 was designed and accepted by the scientific community (Berman, 2004). Conventionally, using jointed criteria, malignancies are generally group into various categories as depicted in table 2.1.

Table 2.1: Classification based on cell type

Carcinoma tissues	Cell type
carcinoma	Epithelial
sarcoma	Non-hematopoietic mesenchymal
Leukemia and Lymphoma	Hematopoietic
Germinoma	Germ

Some other cancer diagnosis criteria include;

- Malignant cells degree of resemblance to their untransformed or normal counterparts.
- Local tissue appearance and architecture of stromal.
- The location of anatomic from which breast tumors arise.
- Molecular, epigenetic and genetic features.

2.2.2 Carcinoma Histological Types

- **Adenocarcinoma:** Adeno-organ alludes to a carcinoma highlighting minute glandular-tissue engineering, or potentially organ related molecular items and tissue cytology. An illustration is the mucin.
- **Squamous cell carcinoma:** This alludes to a carcinoma having attributes demonstrative of squamous separation (intercellular extensions, keratinization, and squamous pearls) and detectable components.

- **Adeno-squamous carcinoma:** Adeno-squamous historical type of carcinoma is referred to a tumour mixture containing both squamous cell carcinoma and adenocarcinoma where each of the cell types comprises at least tumor volume of 10%.
- **Anaplastic carcinoma:** This alludes to a high-review carcinomas heterogeneous gathering that components cells lacking particular cytological or histological confirmation of any of the all the more particularly separated neoplasms. These tumors are alluded to as undifferentiated or anaplastic carcinoma.
- **Large cell carcinoma:** Compose of unmistakably polygonal-formed or extensive dreary adjusted cells having bounteous cytoplasm.
- **Small cell carcinoma:** These cells are not exactly roughly 3 times the distance across of an idle lymphocyte and are as a rule round and minimal obvious cytoplasm. Little cell malignancies may themselves, periodically have huge segments of somewhat axle formed or potentially polygonal cells.

Generally, there are countless subclasses/sorts of undifferentiated and anaplastic carcinoma. Lesions are a portion of the all the more outstanding carcinomas comprising of pseudo-sarcomatous segments including spindle/axle cell carcinoma (containing prolonged cells taking after connective tissue diseases), sarcomatoid carcinoma (blends of spindle/axle and mammoth cell carcinoma) and the giant cell carcinoma (containing gigantic, unusual, multinucleated cells). Giant cells as well as spindle/axle cell segments are found in Pleomorphic carcinoma, moreover, not over 10% part of cells normal for all the more exceptionally separated sorts, for example, the squamous cell carcinoma or potentially adenocarcinoma. Moreover, it is very rare for tumors to contain individual's components resembling both true sarcoma and carcinoma to including; pulmonary blastoma and carcinosarcoma as seen in (Travis et al., 2004).

Carcinoma diagnosis: Biopsy is definitely an essential diagnosing tool for carcinomas. Different devices incorporate; center biopsy, fine-needle aspiration (FNA) as well as subtotal expulsion of single hub. Pathologist's microscopic examination is then important to legitimately distinguish and perceive molecular, tissue structural qualities and cell of epithelial cells.

2.2.3 Staging

Carcinoma staging is referred to the act of the combination of pathological review, physical/clinical examination of tissues and cells, imaging studies, surgical techniques and laboratory tests logically to obtain information about the extent of its invasion and metastasis and the size of the neoplasm.

Usually, Roman numerals are used for the staging of carcinomas. Many of the classifications uses Stages I and II carcinomas to confirm when the tumors have spread to local structures and/or has been found to be small. Typically, Stage III carcinomas is confirmed to have been spreading to organ structures, provincial lymph hubs as well as tissues, while Stage IV carcinomas affirms when tumors have as of now metastasized via blood to organs, tissues or inaccessible destinations.

Various carcinomas categories use Stage 0 carcinoma to describe carcinoma in occult carcinomas detectable and situ only through the testing of sputum for malignant/cancerous cells (carcinomas of the lung).

Staging criteria differs dramatically due to the organ system where the tumor grows. Such cases are shown in the bladder and colon malignancy organizing framework resultantly depending; in renal carcinoma, staging depends on both the profundity and size of the tumors intrusion into the renal sinus lastly, on the profundity of attack, breast carcinoma stagging is more reliant on the extent of the tumor. Lung carcinoma has a duller and confounded staging framework considering various anatomic factors and size as portrayed in (Pepek et al., 2011).

It has been recorded that the systems of the UICC/AJCC TNM are mostly utilized. But, for some normal tumors notwithstanding, traditional staging strategies; colon malignancy dukes grouping are as yet considered.

2.2.4 Grading

Carcinomas grading is alluded to the criteria exploration to semi-evaluate the level of tissue development and cell found in cells change in respect to the show of epithelial tissue of ordinary parent from which the carcinomas are inferred.

Grading in carcinoma is mostly performed after the surgeon and/or a treating physician obtains a suspected tumor tissue sample using surgical resection; sputum cytopathology, direct washing or brushing of tumor tissue, needle or surgical biopsy, etc. Then, a pathologist critically examines the stroma and the corresponding tumor, also utilizing flow cytometry, immunohistochemistry, or staining. Conclusively, the pathologist at that point groups the tumor into one of the 3 or 4 evaluations as depicted below:

- **Well Differentiated or Grade 1:** Here, there is a nearby similarity to the ordinary parent tissue and the tumor cells are effortlessly characterized and distinguished as a histological substance of a specific malignant.
- **Moderately Differentiated or Grade 2:** In this grade, there is resemblance considerably to the tissues and parent cells, but the much comprehensive attributes are not specifically well-built and easily lead to abnormalities.
- **Poorly Differentiated or Grade 3:** In grade 3, there is almost no likeness in original parent tissue and the malignant tumor; the more intricate design highlights are generally primitive or simple and variations from the norm are apparent.
- **Undifferentiated Carcinoma or Grade 4:** In this grade, the carcinomas have no noteworthy likeness to the tissues and the relating guardian cells, with no unmistakable arrangement of ducts, stratified units, spans, keratin pearls, organs or other known traits predictable having higher separated neoplasm.

Even when there is convincing and definite statistical resemblance between tumour prognosis and carcinoma grade for some sites of origin and tumor types, the degree of the association between them is still highly variable. In this scenario, it is generally proven that; a worse prognosis results to higher grade of lesion as seen in (Sun et al., 2006).

Epidemiology: Generally, cancer is seen as a disease of the aged but cancer could also be diagnosed in children. Moreover, contrast views to that of the aged, carcinomas are not rampageously found in children. Family history and age are the two biggest risk factors for ovarian carcinoma.

2.3 Fibro-adenoma Tissue

A scary experience could be investigation of a breast lump. Moreover, it is not all tumours and lumps that are cancerous. A type of non-cancerous (benign) tumor is known as the fibro-adenoma. Fibro-adenoma requires treatment even though it is not life-threatening.

Fibro-adenoma commonly found in the breast of women under the age of 30 and it is a non-cancerous tumour. In the United State, fibro-adenoma according to Mammotome is diagnosed in approximately 10 percent of women. It was further stated that African-American women are more likely to be diagnosed of these tumors.

Tumour mostly comprises of connective (stromal) tissue and the breast tissue. Although most women have only one tumour, 10 to 15 percent of women have multiple lumps. Fibro-adenomas not only occur in one breast, it can also occur in both breasts.

Small size of some fibro-adenomas makes them so tiny that they cannot be felt. Even when any is felt, the surrounding tissue makes it very distinct. The tumors have a detectable shape and the edges are clearly defined. Mostly, they are typically not tender and are moveable under the skin. These tumors may have a rubbery feel to them but often feel like marbles. Figure 2.2 shows right view of the mama with fibro-adenoma tissue.

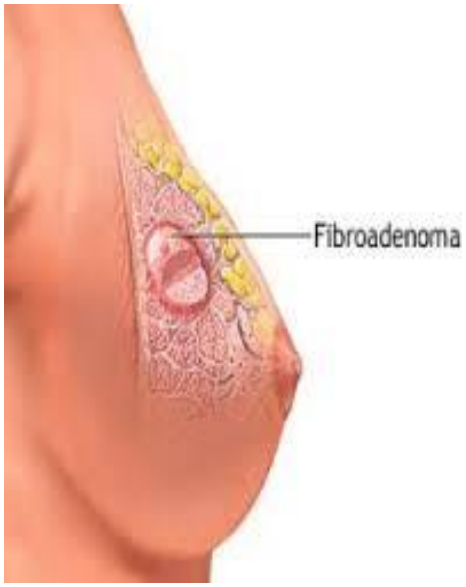

Figure 2.2: Fibro-adenoma tissue

Causes: It is still arguably that the exact cause of fibro-adenomas is yet known. An estrogen hormone plays a part in the development and growth of the tumours. Fibro-adenomas are higher risk of development in women associated to taking oral contraceptives before the age of 20. Particularly, the tumours develops/grows rapidly and faster during pregnancy. And for women under menopause, they shrink often. Possibly, fibro-adenomas could get resolved on radiologist's aid.

Types: Basically, fibro-adenomas are of two types: the complex fibro-adenomas and the simple fibro-adenomas. While simple tumours look the same all over when viewed under a microscope and does not increase breast cancer risk, the complex tumors contains calcifications; calcium deposits and macrocysts; fluid-filled sacs large enough to feel and to see without a microscope components. The complex fibro-adenomas have the ability of slightly increasing the breast cancer risk. An audit demonstrates the American Cancer Society (ACS) expressing that ladies having complex fibro-adenomas around have one and a half to two times more serious danger of having breast disease than ladies having no breast bumps.

Diagnosis: Diagnosis of fibro-adenoma includes leading a physical examination and the breasts will be palpated (physically inspected). A breast mammogram or ultrasound imaging test may likewise be directed. The breast ultrasound includes making a photo on the screen which is performed by moving a hand-held gadget brought a transducer over the skin of the breast of a lady lying on the table. An X-ray of the breast taken while the breast is packed between two level surfaces is known as mammogram.

Biopsy or a fine needle goal might be analyzed to expel tissue for testing. It is performed by embedding a needle into the breast and after that, expelling little bits of the tumor. To determine any type of fibro-adenoma and the cancerous degree, the tissue will then be sent to a lab for microscopic examination.

Remedy: If a patient is confirmed through diagnosis to having fibro-adenoma, removing is basically optional and should not be enforced. It as well depends on her personal concerns, family history and physical symptoms. The decision to removing it lies between the patient and the radiologist; whether to keep it or have it removed. Fibro-adenomas that are definitely not

cancerous and do not grow can be closely monitored with clinical breast examinations and imaging tests; ultrasounds and mammograms.

Typical illustrations given below describe decisions to removing a fibro-adenoma:

Examining if the breast natural shape is impacted;

- If it causes pain to the patient.
- If the patients are concerned about developing cancer.
- If the patient have a family history of cancer.
- If the patient received questionable biopsy results.

In a scenario where a fibro-adenoma is removed, the growth/development of one or more in its place is possible and induces slightly increase of risk of breast cancer; the patient advised schedule regular mammograms as well as regular checkups with the radiologist. It is also advisable for the patient to include breast self-exams in her regular schedule. A felt of any changes in the shape or size of an existing fibro-adenoma should prompt a visit to the radiologist (http://www.healthline.com/health/fibroadenoma-breast Received 10 May 2017).

2.4 Mastopathy Tissue

The word mastopathy encompasses all changes in benign breast, illustrated by indurated nodules, cysts or swelling. Often, these changes affect both the benign breasts. In general, a change in breast cancer may be as a result of severe form of mastopathy.

Glands and connective tissue are the components of the breast. Changes made to these tissues result to the occurrence of mastopathy. Glandular cysts occur when there is an increase of connective tissue that causes nodule changes. This is as a result of frequent combination of cysts and nodules. Mastopathy is known to be the most common breast disease because one in two women suffers from a mastopathy during its existence. When mastopathy is compared with other benign breast changes (tumours) such as fibro-adenomas, lipomas and adenomas, there is always a distinction. Figure 2.3 displays two mammas with the right mamma showing mastopathy effect.

Causes: Mastopathy is highly suspected when there is an imbalance between the progesterone and the female sex hormones estrogen with excessive amounts of estrogen. When a patient does not follow hormonal therapy due to menopause, mastopathy abruptly disappears since it is hypothesized to affecting women aged 30 to 50 years.

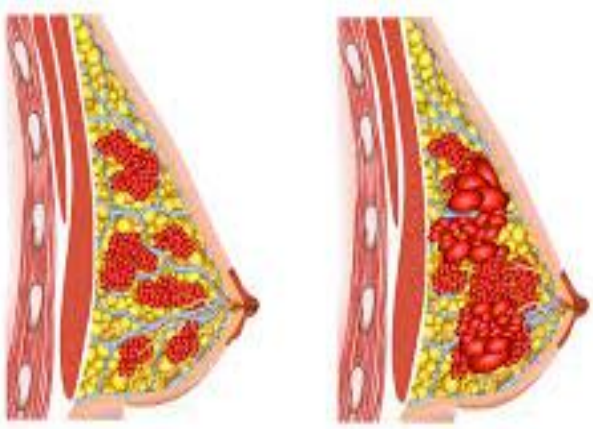

Figure 2.3: The right mamma showing mastopathy effect

Also, another hormonal cause of mastopathy is; its symptoms occur in the cycle particularly at end of cycle, just before the onset of the rules. Mastopathy could also be caused by hyper/hypothyroidism (thyroid disease).

2.4.1 Mastopathy Degree of Severity

Simple Mastopathy or Mastopathy Grade I: In this grade, tissue of the breast is indurated, thickened and probably have cysts or not. Some samples of histological tissue show that the cells appearance is normal but the tissue proliferated; a common form of simple mastopathy has cases with 70%.

Proliferative Matopathy or Mastopathy Grade II: This grade considers certain cells that grow faster than others; it is frequently referred to as cells of milk ducts. At this stage, appearance of the cells is not affected by mastopathy. The proliferative mastopathy is the second form in terms of frequency with about 25% of cases.

Severe Mastopathy or Mastopathy Grade III: With about 5% of cases there are rarest forms where the biopsy revealed pathological cells. At this point, it is not yet cancer, but subsequently,

the cells can later lead to cancer. 2 to 4 times higher, women having severe mastopathy have a breast cancer risk. And in order to control eventual degeneration, regular checks are imperative.

Mastopathy Symptoms: Intensity type greatly determines symptoms variability. But severity of mastopathy is neither influenced by symptom nor intensity.

One of the major characteristic of mastopathy is that the symptoms occur towards the end of the cycle meaning that; symptoms occur shortly before menstruation (premenstrual syndrome). Below are some of the symptoms:

- Feeling of tension up to pain in the breast and slenderness.
- Nodules swellings palpated or palpable nodule can be. The indurated nodules are so extensively arranged often and patients often have many small nodules concentrated in several places.
- Presence of cysts is often indicated by fluid secretion from the nipple.

Note that cancer is not often synonymous with all palpated nodule. Although any felt changes demand medical supervision palpable. Hence, it is advisable for every woman to essentially practice to self-examination regularly.

Mastopathy Diagnosis

- Consideration of symptoms with regards to history.
- Breast palpation.
- Breast scanning
- Regular mammography
- Use biopsy in a case of suspicion of a malignant change.

Treatment Options: Mastopathy is yet to be identified having a cure. So far, only identified symptoms are treated. Cysts or the nodules and nodules unsightly suspected cancer can be surgically resected. In case of pain, analgesics can be effective. A gel containing a gestagen could mitigate symptoms.

Need for surgical resection is not necessarily immediate but regular biopsy should be performed to timely detect any cancerous degeneration.

Possible Complications: It is recorded that cancer of the breast developed only by the severe mastopathy. Women diagnosed to having mastopathy other than severe mastopathy could live with it for those are some worth harmless.

Preventive Measures: Medical tests one to two per year and regular breast self-examination. (http://www.rayur.com/mastopathy-breast-changes-benign.html).

2.5 Glandular Tissue

Women's breasts consist of glandular tissue called the mammary glands that holds milk-producing cells. They also have connective tissue, which includes adipose or fatty tissue. These tissues make up the shape of your breasts. Glandular tissue are a mixture of both exocrine (have hormones secreted onto surfaces, ducts) and endocrine (hormones secreted into the blood, ductless) glands. Figure 2.4 depict glandular tissue.

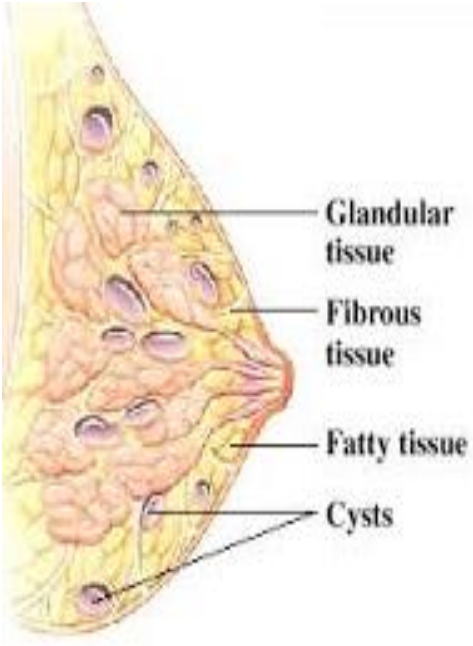

Figure 2.4: Mamma displaying glandular tissue

2.5.1 Breast dense tissue

As demonstrated in figure 2.5, breasts are made up of ducts, fatty, fibrous connective and lobules tissues.

- Generally, the lobules are often called glandular tissue because they produce milk.
- The milk produced by lobules is usually transferred to the nipple by tiny tubes called ducts.
- Fat and fibrous tissues give breasts their shape and size and hold the other tissues in place.

If the breast has a lot of glandular tissue or fibrous and not much fat, it is regarded to be dense. Proportion of breast density is not fixed since some women have denser breast tissue than others. Breast becomes less dense with age for most women. There is little change in some women.

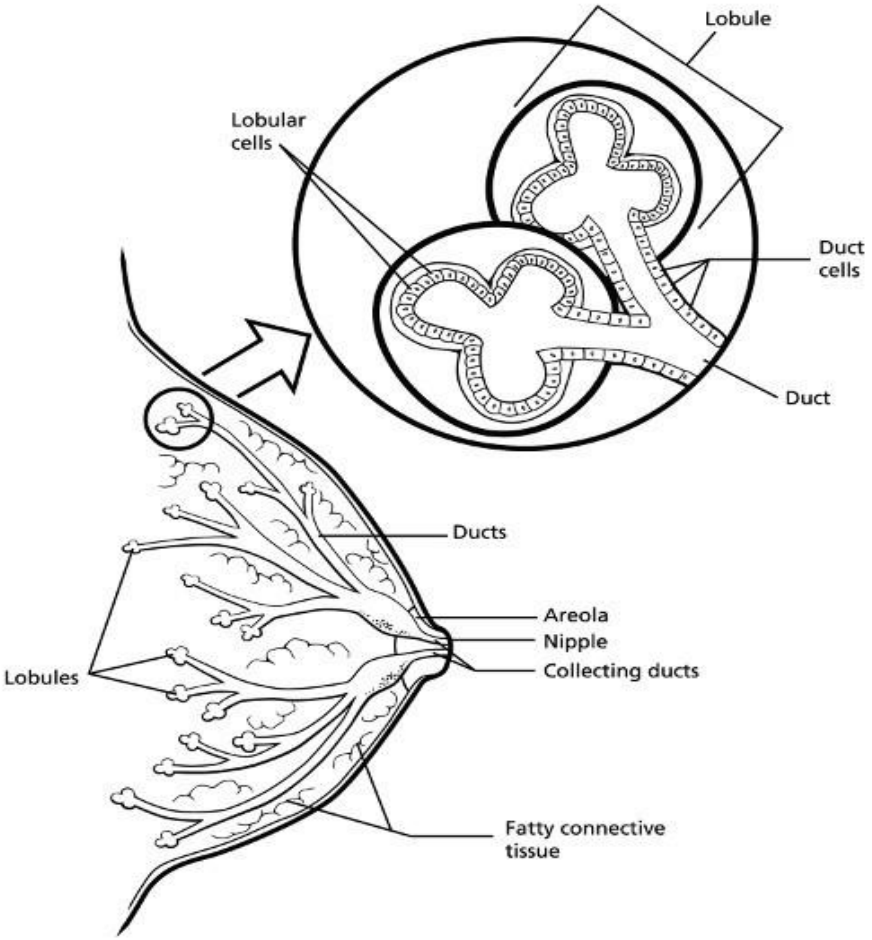

Figure 2.5: Breast dense tissue structure

Breast density: Mammograms is still the only diagnostic medium for breast density. The fact that some breasts are firm does not necessarily mean they are dense. For firmness of breast does not determine breast density. Breast density is never related to breast firmness or size.

The doctors who read x-rays like mammograms are known as radiologist. Radiologists check the mammogram to ascertain breast density as well as abnormalities. In general, breast density is categorized into four. They go from extremely dense tissue with very little fat to almost all fatty tissue as shown in figure 2.6. From Those four categories, radiologist decides which best describes how dense a breasts is.

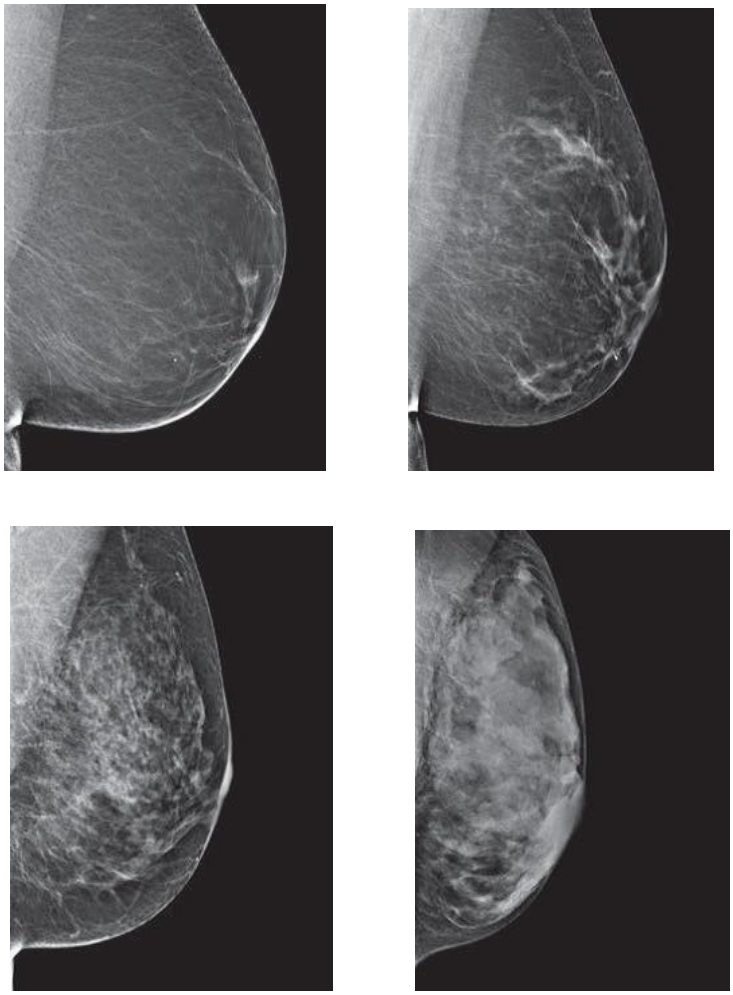

Figure 2.6: Breast density categories including breast with almost all fatty tissue, breast with scattered areas of dense fibrous and glandular tissue, breast with dense fibrous and glandular tissue and extremely dense breast

Importance of Breast Density: Ladies with less thick breast tissue have a marginally less danger of breast growth contrasted with ladies with thick breast tissue. Unmistakably thick breast tissue makes it harder for radiologists to see disease yet indistinct why thick breast tissue is connected to breast tumor chance. Dense breast tissue on mammograms has white looks. Also,

breast tumours or masses has white look, so tumours are easily hidden by the dense tissue. Conversely, fatty tissue has almost black looks. It is easier to see a tumor that looks white on a black background. Conclusively, in women with dense breasts, mammograms are less accurate.

2.6 Connective Tissue

Ligaments and connective tissue gives the breast its shape as well as provide support to the breast. The breast derives its sensation from the nerves. Also, the breast contains lymph nodes, lymph vessels and blood vessels. Connective tissue consisting of blood vessels, fat and muscles happens to be the beginning point of breast cancer. Sarcoma is the cancer that begins in the connective tissue. Sarcomas of the breast are rare. Figure 2.7 is a side view of mamma showing connective tissue.

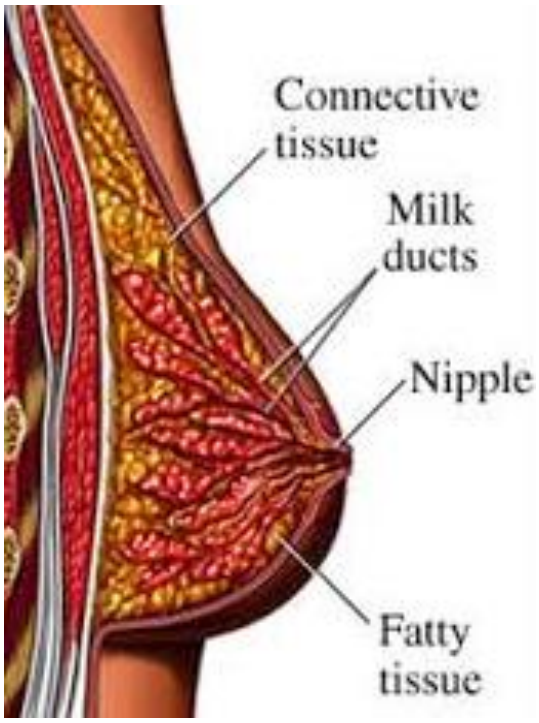

Figure 2.7: Woman's breast displaying connective tissue

Angiosarcoma: The form of cancer that starts from cells lining lymph vessels or blood vessels is known as the angiosarcoma. Detection of angiosarcoma in the breast is some worth rare. When it does, it is as a result of complications from previous treatments of radiation. Women who develop lymphedema as a result of radiation therapy or lymph node surgery to treat breast cancer

could easily be diagnosed of angiosarcoma within their arm. Such cancers types tend to spread and grow quickly. Angiosarcoma treatment is generally the same as for other sarcomas.

2.7 Adipose Tissue

Breast cancer and adipose tissue are dynamic duo known to be dangerous. Adipose tissue particularly surrounds the mammary glands and is abundant in the breast. Along these lines, it comes into coordinate contact with fatty tissue when breast disease winds up noticeably obtrusive. In the 20s, various looks into have demonstrated that progenitors and adipocytes advance breast malignancy forcefulness by multiplication incitement and, particularly, attack by genius provocative cytokines, discharging proteases, and by tumor cell digestion balance.

Considering obesity scenarios, the number and size, as well as adipocytes emissions are affected significantly. Patients having both breast malignancy and obesity show at determination more forceful tumors and malady movement are set apart by a substantially higher rate of mortality. In this way, in medication, considering the impact of fat cells on illness movement is of a noteworthy intrigue, particularly for obese patients' treatment. As delineated in figure 2.8, early neighborhood tumor intrusion in breast malignancy brings about quick nearness of cancerous cells to adipose tissue. The figure likewise demonstrates the obtrusive breast tumor histological examination after H&E recoloring unique amplification X 200 with arrows showing tumour.

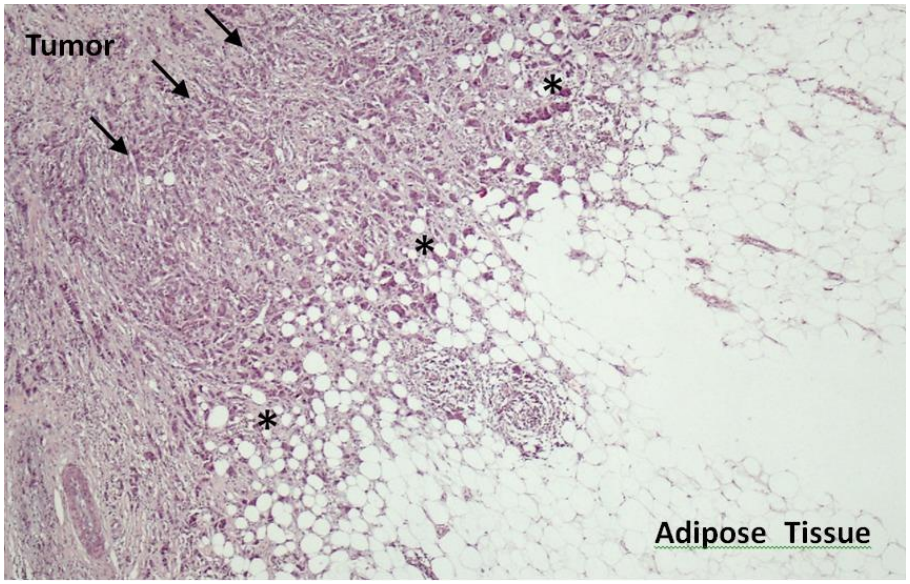

Figure 2.8: The tumour and adipose tissue (http://blogs.biomedcentral.com/on-medicine/2015/04/29/breast-cancer-adipose-tissue-a-bulky-neighbor-causing-trouble/)

CHAPTER 3
RELATED RESEARCH

3.1 Overview of Related Works

Background information relevant for some research works recommended in modeling mammography are presented in this chapter. The chapter further presents the breast tissue concept analysis and the implored analytical methodologies in modeling breast tissues. Artificial neural networks implored in this thesis, focusing on the artificial intelligence techniques will be introduced in this chapter. Academic papers previously published in breast tissue modeling concept with much focus on malignant tissues will be discussed. Many reviews have demonstrated the importance of EIS for the recognition of breast tissue/cancer. Many of these reviews are investigated in the section below.

Since 1926, specialists have been carrying researches on breast tumors electrical properties (Fredicke and Morse, 1926). Because of shifting outcomes that have been in existence, the agreement has been that breast tumors electrical properties do contrast from healthy breast tissue. Surowiec and partners in the year 1988 (Surowiec et al., 1988) demonstrated some vitro tests in order to decide the fluctuation of some properties between healthy tissue samples, a mix of carcinoma with healthy tissue samples and finally, samples of breast carcinoma including the apparent limit of lesion with samples of healthy tissue as it were. The group reasoned conductivity of cancerous tissues and dielectric constants contrasted between sample groups with frequency measurements from 20KHz to 100MHz, albeit significant variability that existed between data to be measured.

In (Morimoto et al., 1993) and (Morimoto et al., 1990), Morimoto and associates exhibited the measurement of breast tumors electrical impedance in vivo. Execution of the task was made through injecting fine-needle electrode into the tumor utilizing three-electrode technique. They group the membrane capacitance, intercellular resistance and extracellular resistance in light of measured complex impedance and model circuit. Extracellular resistance with a combination of series of the capacitance and the intracellular resistance are within the model circuit. Frequency range of 0 to 200KHz was used to obtain the estimations. The group inferred that there are factually noteworthy contrasts amongst pathology and normal tissue. In any case, degree of

values ascertained usually overlap for every tissue type.

In (Jossinet, 1996), (Jossinet, 1998) and (Jossinet and Schmitt, 1999) Jossinet used frequency range of 488Hz-1MHz to study the six sets of breast tissue impedance measurement. From 64 patients, 120 samples of impedance spectra were obtained, having specimen sets grouped in three categories of typical tissue of the woman breast; carcinoma and two types of benign tissue. Every one of the three articles introduces studies using similar data. In Jossinet's first research work, he examined how data with impedance properties varies between gropus, examining reduced standard deviation as well as standard (Jossinet, 1996).

Jossinet's second research was directed towards computing the Cole-Cole parameters by plotting intricate impedance against frequency. His research included computing parameters that would separate other samples from carcinoma samples (Jossinet, 1998). This research recommends that at frequencies more than 125KHz, there is a much distinction in attributes of cancerous tissues. Schmitt was invited by Jossinet in his third research where they tried to characterize a new set of eight parameters by which other tissues can be separated from cancerous tissue (Jossinet and Schmitt, 1999). Both infer that tissue characterizations are appropriate for several parameters spanning a range of frequencies.

In (Chauveau et al., 1999), Chauveau and associates using a range of frequency values calculated bio-impedance parameters. Considering samples ex-vivo of both normal and pathological tissues, estimations were gotten for frequency range from 10KHz to 10MHz. In view of the measurements above and a model that incorporates a constant phase element, membrane capacitance, intracellular and extracellular resistances were computed from the measurement and a model that incorporates a consistent phase element. Considering these values, three indices were characterized for pathological tissue classification. Experimental observations in (Chauveau et al., 1999) differentiated tissues with fibrocystic changes and normal tissues from cancerous tissues.

In (Zhao et al., 2012), in a bid to enhance the spatial resolution of impedance images built a trans-admittance mammography system. A system with an array of 60 x 60 electrodes that look like an X-ray mammographic set-up was developed to accomplish their idea. In (Kim, 2012), Kim used the electrical impedance scanning probe to carry out a research on the frequency

dependent conduct of induced current. Along these lines, relationship between the difference in the values of conductivity between the encompassing tissues, tissues with cancer and that of the current measurement was obtained. From the above data, they proposed a breast tumor size formula in view of the current measurement.

Considering (McGivney et al., 2012), the group recommended electrical impedance spectroscopy as a highly ill-posed and a regularized inverse problem due to prior knowledge from modeling error and mammogram images. They introduced and properly analyzed the computational techniques for solving tissue classification methods and EIS inverse issue for the breast.

In (Perlet et al., 2000), Perlet and partners explored the dependability of impedance measurement of breast tissue in a healthy state. Ones in every week, they recorded measurements over two successive menstrual cycles to figure out if electrical impedance spectroscopy images rely on hormones. They presumed that impedance is reliant on hormone levels since the images got differed all through the cycles with some consistency.

Soft computing, genetic algorithms and machine learning are examples of artificial intelligent methodologies. Artificial neural networks (ANN) inside these techniques are most ordinarily utilized as a part of medicinal expectations as presented by David and Joseph in (David and Joseph, 2006), (Lisboa et al., 2006) and (Yardimci, 2009). Neural systems work by distinguishing designs in data as demonstrated by David and Joseph's capacity to learn through understanding; learning from the connections and adjusting to them. To anticipate the result for new sets of data, the learning information is then used.

James W. F. Catto and associates as observed in (Abbod et al., 2005) and (Catto et al., 2009, 2003, 2006), did research on the utilization of neuro-fuzzy modelling (NFM) in bladder tumor. ANN has been contrasted with straight relapse and NFM with foresee the exactnesses of relapse time of bladder cancer patients and tumor relapse in (Catto et al., 2003). At First, patients' arrangement were made considering whether their tumor would relapse or not and after that 'opportunity to backslide' expectations were made for the relapse patients. Statistical methodologies demonstrated poorer outcomes as compared with artificial intelligent methods, with ANN appearing poorer than the NFM at predicting the time to relapse. Based on cancer prediction studies, the authors claimed that, this turns out to be the first research work on neuro

fuzzy modeling. To anticipate backslide in a similar malignancy area, the prescient capacity of these three models has been utilized with various quantities of data and included various examinations with extra information factors contains traditional clinicopathological and molecular biomarkers (Abbod et al., 2004, 2005). Besides, to anticipate the movement of transitional cell carcinoma with NFM and seemed better than ANN, the same predictive model was used as shown in (Catto et al., 2006).

3.2 Dielectric Properties of Breast Tissues

To decide the weakening of a signal through a medium and the reflections caused by a medium, the relative permittivity, conductivity and dielectric properties are utilized, allowing the separation between various sorts of tissue inside the breast at microwave frequencies. Observing the ex vivo and in vivo dielectric properties, a few authentic investigations have been performed considering the normal and cancerous breast tissues specifically, and these are analyzed in detail in the accompanying sections.

In 1984, (Chaudhary et al., 1984) first inspected ex vivo of breast tissue samples expelled amid cancer surgeries. Amongst cancerous and normal tissues, a noteworthy dielectric differentiation was found over the frequency scope of 3MHz to 3GHz, at 25°C. Chaudhary exhibited that critical contrasts existed in the cancerous and normal tissues dielectric properties of the ladies breast, with the best dielectric distinction happening at frequencies underneath 100MHz. Contrast ratio found for conductivity relative permittivity was 4.7:1 and 5:1, respectively. Figure 3.1 demonstrates properties of dielectric variation of malignant and normal tissue with frequency reported in his study.

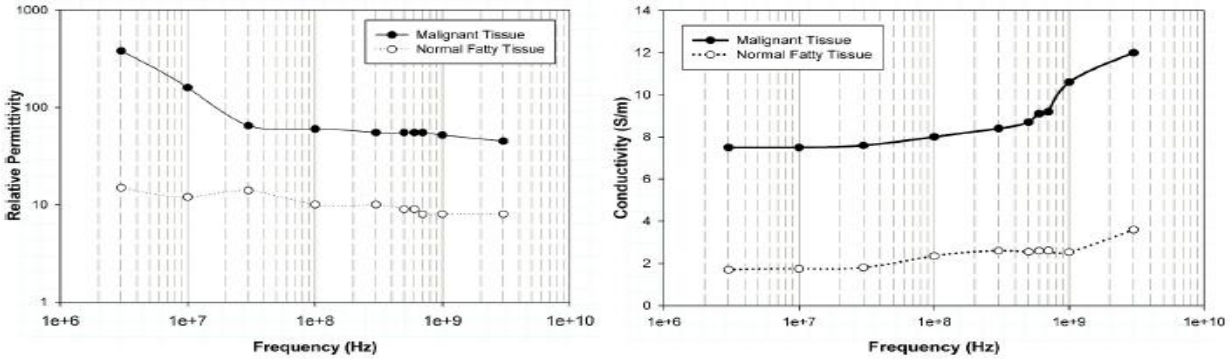

Figure 3.1: The conductivity (right) and relative permittivity (left) variation of malignant and normal tissue between 3 GHz and 3 MHz as depicted in (Chaudhary et al., 1984)

In 1988, (Surowiec et al., 1988) performed conductivity and the relative permittivity of breast carcinoma penetration, where the fringe tissue and the encompassing tissue at frequencies in the vicinity of 20KHz and 100MHz. The instances of ex vivo were taken from a populace of seven patients and put away in physiological saline.

Three measurements were made in three locations: the peripheral tissue, tumour central part and the tissue directly surrounding the tumour approximately 2cm away from the center of the tumour. Their results, shown in figure 3.2, may suggest that there are increased dielectric properties even at the edge of the tumour due to tumour cell proliferation, and that smaller tumours may still be detected using a UWB radar.

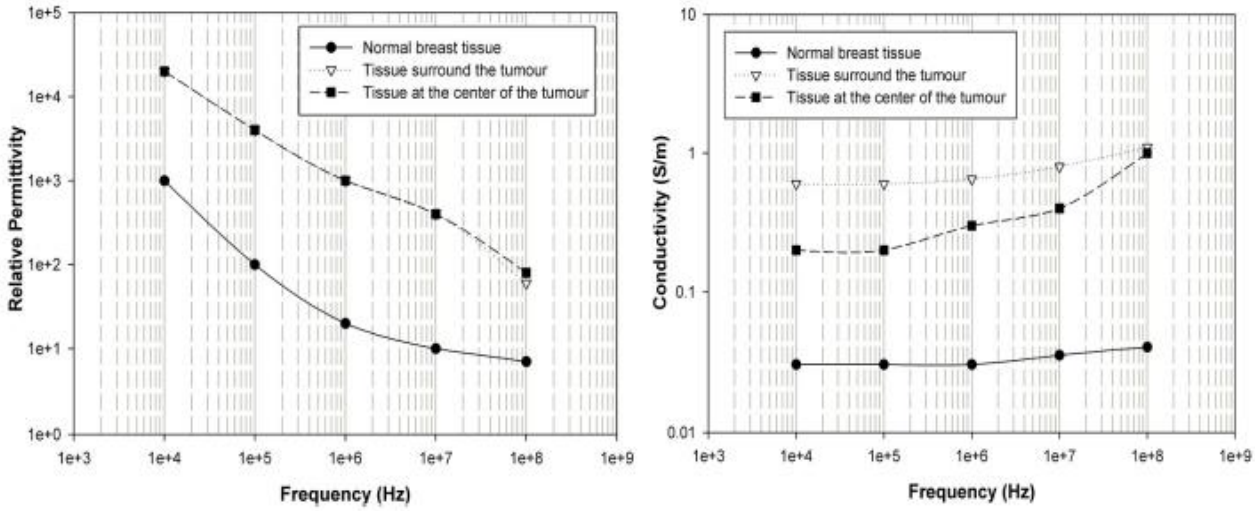

Figure 3.2: The conductivity (right) and relative permittivity (left) variation of surrounding tissue, tumour tissue and peripheral tissue across the frequency band of 0.02MHz and 100MHz, as presented in (Surowiec et al., 1988)

In 1992, Campbell and Land (Campbell and Land, 1992) on the properties of dielectric of ex vivo ladies breast tissue at 3.2GHz, gave itemized data to microwave thermography applications. In this study, the properties of dielectric in four different types of tissue were measured by a resonant cavity technique: normal tissue, fat tissue, malignant breast tumour and benign breast tumour. Table 3.1 demonstrates their results. Where they found that there is an overlap in the dielectric properties for malignant and benign tumour tissues and also observed a much greater properties of dielectric degrees for normal breast tissue suggesting that, both malignant and benign tissue and normal tissue may be difficult to differentiate solely based on their properties of dielectric.

Table 3.1: The dielectric properties of female breast tissue at 3.2GHz as presented by Land and Campbell in (Campbell and Land, 1992)

Tissue Type	Relative permittivity	Conductivity (Sm^{-1})	Water content (%)
Fat tissue	2.8-7.6	0.54-2.9	11-31
Normal tissue	9.8-46	3.7-34	41-46
Benign breast tumour	15-67	7-49	62-84
Malignant breast tumour	9-59	2-34	66-79

In 1994, analysts in (Joines et al., 1994) utilized a range commonly use for microwave-prompted hyperthermia to quantify tests of ex vivo at frequencies between 50MHz to 900MHz. Tissue tests taken from 12 patients were dissected and comes about demonstrated critical contrasts amongst cancerous and normal tissues for the mammary organ, with a distinction proportion of 6.4:1 and 3.8:1 for the relative permittivity and conductivity separately, which is accounted for to be when all is said in done concurrence with the outcomes detailed by (Chaudhary et al., 1984). Their outcomes are plotted in figure 3.3.

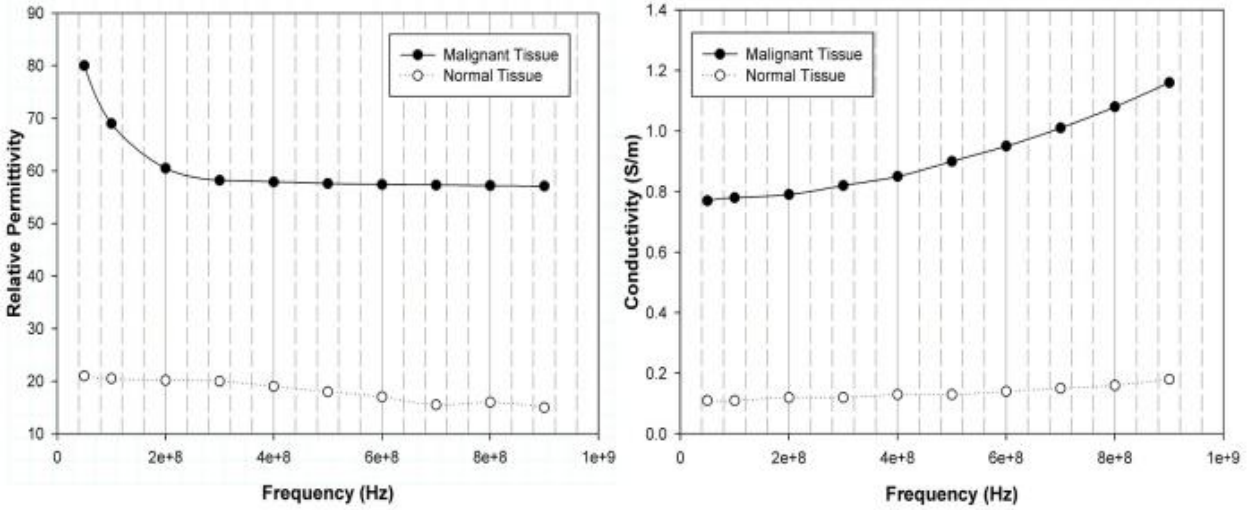

Figure 3.3: The conductivity (right) and relative permittivity (left) variation of malignant and normal tissue between 50MHz and 900MHz, as presented in (Joines et al., 1994)

Also in the year 1994, (Choi et al., 1994) examined the metastasized lymph nodes and normal lymph nodes properties of dielectric, along with the breast cancer tissue dielectric properties in a frequency range from 0.5 to 30GHz. The outcomes are depicted in figure 3.4, and it is noted that both breast cancer tissue and metastasised lymph nodes differ significantly from normal lymph nodes.

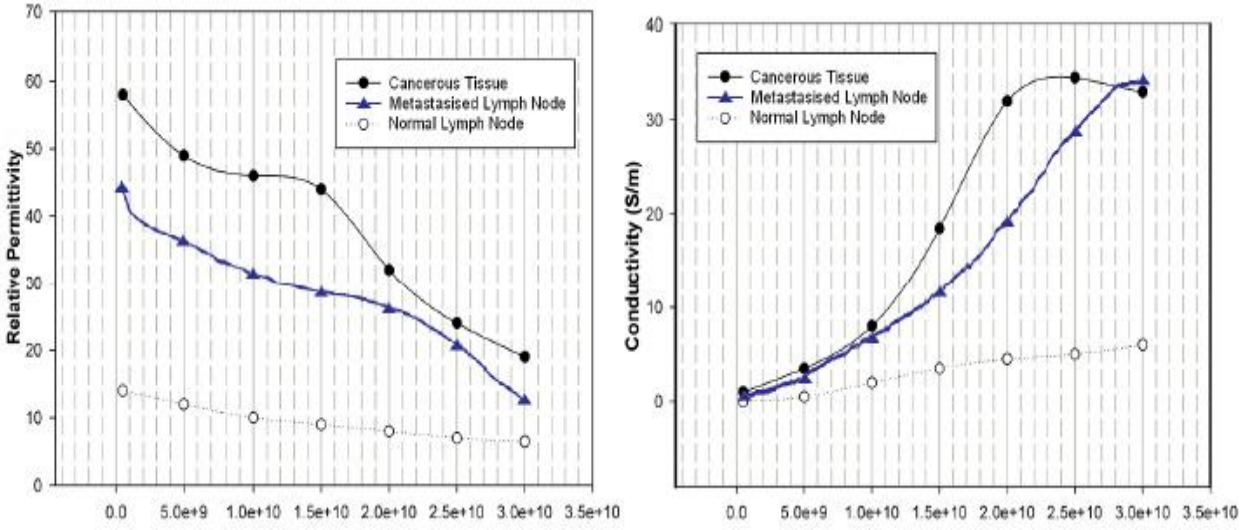

Figure 3.4: The conductivity (right) and relative permittivity (left) variation of malignant and normal tissue between 0.5GHz and 30GHz, as demonstrated in (Choi et al., 2004)

Meaney and partners in (Meaney et al., 2000) utilizing a model of microwave imaging framework, played out the primary clinical examination in vivo in the year 2000. In their examination in a tomographic microwave imaging framework in the vicinity of 300 and 1000MHz, a 16 component monopole reception apparatus exhibit was utilized. In table 3.2, outcomes at 900MHz are displayed and it can be noted that the normal relative permittivity esteem is fundamentally higher, around in the values of 31 and 36, than that distributed in Joines et al.'s study. In the course of this study, no malignant tissue was examined and so a direct comparison cannot be made to the previous ex vivo studies.

Table 3.2: Female breast tissue average dielectric properties at 900MHz measured in vivo using an active microwave imaging system developed and presented in (Meaney et al., 2000)

Patient	Age	Average relative permittivity (%)	Average conductivity (Sm^{-1})
1	76	17.22±11.21	0.5892±0.3547
2	57	31.14±4.35	0.6902±0.3650
3	52	36.44±6.24	0.6869±0.3156
4	49	35.43±3.93	0.5943±0.3841
5	48	30.85±7.22	0.6250±0.3550

All the more as of late, analysts (Lazebnik et al., 2007) concluded a standout amongst the most extensive examinations to date on the breast properties of dielectric. The primary investigation (Lazebnik et al., 2007) concentrated on the normal tissue properties of dielectric and the second examination (Lazebnik et al., 2007) concentrated on the cancerous and normal breast tissues

dielectric differentiation. On the greater part of their investigations, the Cole-Cole portrayals were mapped to the information so as to help on the properties of dielectric estimations. Lazebnik with the would like to enhancing a considerable lot of the obvious shortcomings of past explores, for example, gaps in the frequency bands and small patient specimen sizes inspected and studied histopathologically a substantial pool of naturally extracted breast tissue from patients and separated normal tissue instances into 3 sets, recognizing each by the rate of glandular, adipose and fibro-connective tissue contained in the specimen before getting the qualities for the properties of dielectric. Definition of the three sets is given below:

- All samples from 0-30% adipose tissue are contained in category 1
- All samples from 31-84% adipose tissue are contained in category 2
- All samples from 85-100% adipose tissue are contained in group 3.

Major findings in their first research in (Lazebnik et al., 2007) were that breasts with low fibro-glandular and high adipose contents presented dielectric properties in lower average, whereas breasts with high fibro-glandular and low adipose tissues presented higher dielectric properties, which suggested that, within healthy breasts, a wide range of properties of dielectric is possible. Results are summarized in figure 3.5.

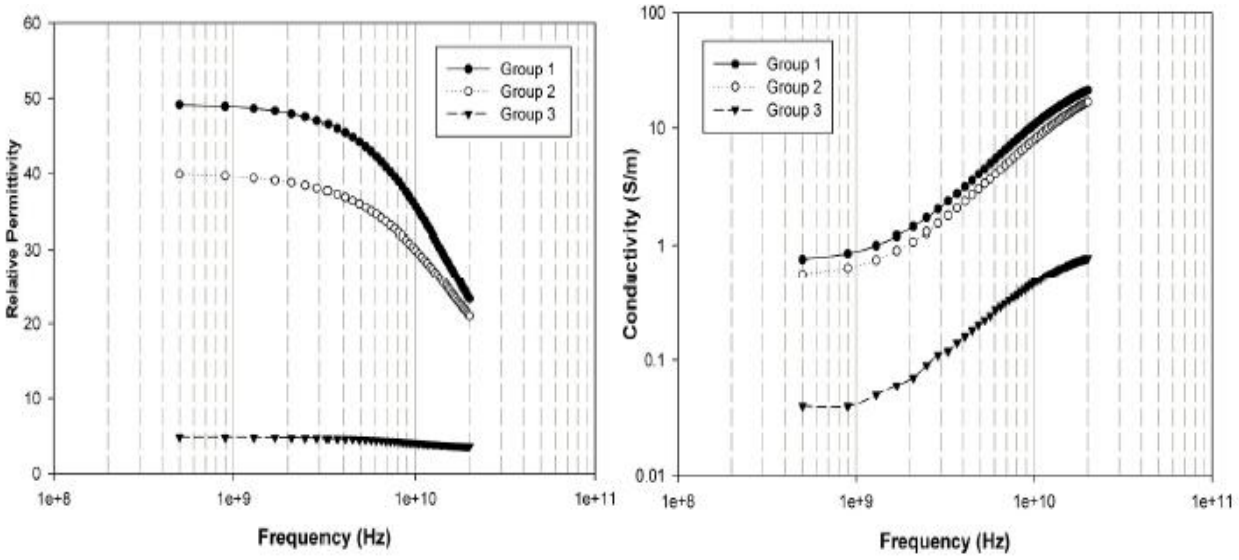

Figure 3.5: The conductivity (right) and relative permittivity (left) of normal breast tissue as presented in (Lazebnik et al., 2007) over the frequency band 0.5GHz to 20GHz. Group 1 assumes 0-30% adipose tissue, group 2 assumes 31-84% adipose and group 3 assumes 85-100% adipose tissue

Lazebnik's main conclusions in comparing these results to previous dielectric studies are as follows:

- Normal tissue dielectric properties in the specimens of Group 3; the most elevated in adipose substance, were lower than any past investigations.
- Normal tissue properties of dielectric in the specimens of Group 1; the most astounding in fibro-glandular substance and least in adipose substance, were higher than any past examinations.
- Data of dielectric crossed a considerably more prominent scope of qualities than those displayed in past studies, with an exemption to Land and Campbell's exploration (Campbell and Land, 1992).

In general, as previously noted in (Campbell and Land, 1992), Lazebnik accounted these changes to the large heterogeneity in normal breast tissue and acknowledged the relation found between the content of tissues within the breast; more or less adipose, and the measured dielectric properties; lower or higher, respectively. In their second research in (Lazebnik et al., 2007), they further addressed the differences between malignant, benign and normal, tumours across a degree of frequencies; 0.5 to 20GHz. Adipose, fibro-connective and glandular tissues are presented as normal breast tissue. Cysts and fibro-adenoma are presented as benign tumour tissue and finally, lobular and ductal carcinomas are classified as malignant tumour (IDC, DCIS, ILC and LCIS). The results are shown in figure 3.6.

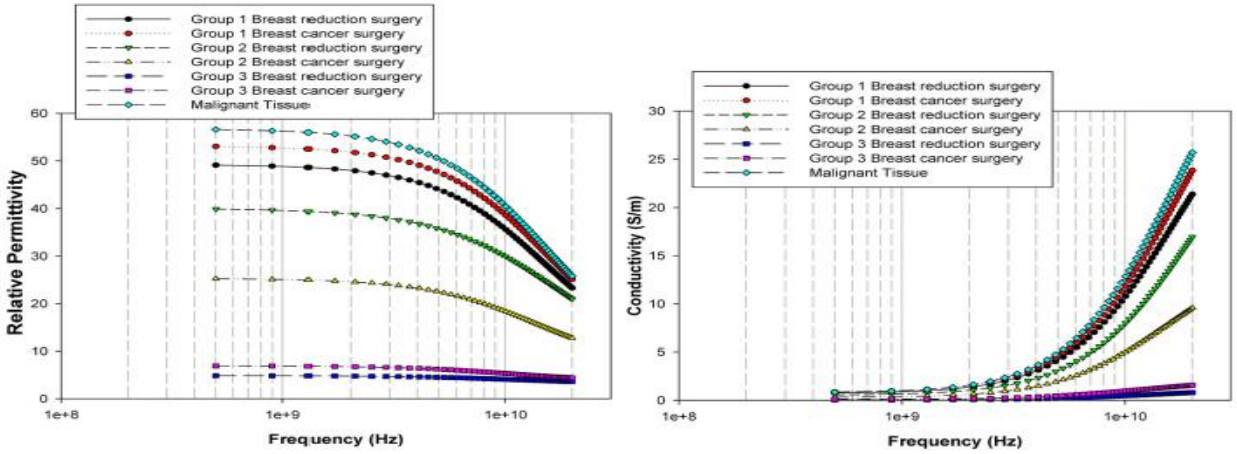

Figure 3.6: The conductivity (right) and median relative permittivity (left) Cole-Cole curves for groups 1, 2 and 3 for normal tissue obtained from cancer surgeries and reduction surgeries. The median relative permittivity curve of the dielectric properties of samples that contained at least 30% malignant tissue content is also shown for comparison. All results correspond to the 50th percentile (Lazebnik et al., 2007)

From the exploratory outcome above, Lazebnik and partners in (Lazebnik et al., 2007) recorded that estimation of dielectric esteems for cancerous tissue were total terms with studies in (Chaudhary et al., 1984), (Surowiec et al.,1988) and (Joines et al., 1994). Additionally, (Lazebnik et al., 2007) altogether advocated the contrasts between the curves for category 2 with a test error because of the nearly little size of instance utilized breast reduction surgery study into contrasted with the examination of tumor surgery, which changed from 16 to 84 instances. Besides, in (Lazebnik et al., 2007), it was recognized that normal tissues properties of dielectric acquired through breast malignancy surgery were lower than those gotten in surgery of breast lessening and noticed this is because of the way that in glandular tissue, tumors normally created and thus the non-influenced tissues evacuated had relatively higher substance of adipose.

In (Lazebnik et al., 2007), by altering the substance of adipose in the instances, it is discovered that there existed just a 10% distinction between the conductivity of cancerous tissue and normal tissue, and an inexact 8% contrast in relative permittivity at 5GHz. Additionally, by altering for the substance of both fibro-connective and adipose tissues in the specimens, they found no measurable distinction between cancerous tumor tissues and normal fibro-glandular in the breast. Inside the breast, the high fibro-glandular/normal tissue properties of dielectric cover those of cancerous tissue, and therefore emerged as 'false positive' results and and as previously observed, create a much more difficult imaging scenario.

In general, researches presented in (Lazebnik et al., 2007) highly added to breast tissue dielectric properties knowledge by characterizing the tissues over a wide frequency band in the range of 0.5GHz and 20GHz (Lazebnik et al., 2007), by having separate examination relying upon the extent of various sorts of tissues inside the breast and by fundamentally expanding the populace measure, an extremely point by point properties of dielectric database in view of Cole-Cole parameters was set up for each tissue sort, which is critical to satisfactorily actualize a numerical breast apparition.

Moreover, in the year 2009, Halter and associates in (Halter et al., 2009), from a clinical report with fewer patients in which estimations of cancerous breast tissue properties of dielectric were acquired in three distinct situations, exhibited the underlying outcomes. Conclusively, the dielectric properties were taken thus; (i) coordinated estimation in ex vivo breast cancer samples

with both MIS and EIS tests, (ii) coordinated estimation in vivo breast tumor with both microwave impedance spectroscopy (MIS) and electrical impedance spectroscopy (EIS) tests lastly, (iii) estimation means of electrical impedance tomography imaging. While some limitations in such particular research may be due to the fewer patients and inaccurate simultaneous measurements for scenarios (i), (ii) and (iii) for each considered lesion, it was noted in this study that there may also be some limitations related to *ex* vivo measurements such as those in (Lazebnik et al., 2007). In (Hatler et al., 2009), observations were made that, the normal breast tissue dielectric properties reported in scenario (i) is in terms with the previous studies; (Lazebnik et al., 2007). Conclusively, the estimations in scenario (iii) dielectric properties is also in terms with early ex vivo researches; (Surowiec et al., 1988), (Campbell and David, 1992), (Jossinet and Schmitt, 1999) and (Lazebnik et al., 2007). Nevertheless, observations from (Haemmerich et al., 2002) show that there is an adjustment in a few parameters of dielectric subsequent to extracting tissue and credited those progressions to varieties of ischemic impacts, tissue lack of hydration and temperature. They noticed that in some seconds, those progressions takes place after extraction of tissues and afterward may balance out for a considerable length of time. At long last, Halter's exploration in situation (ii) demonstrated that results for relative permittivity and conductivity of breast cancer were particularly more noteworthy than those exhibited in perceptions (iii) or (i).

CHAPTER 4
MACHINE LEARNING TECHNIQUES FOR BREAST TISSUE CLASSIFICATION

4.1 Machine Learning

A subfield of computing that gives computer-aided device the capacity to learn without being unequivocally modified is known as machine learning. From the investigation of computational learning hypothesis in artificial intelligence and pattern recognition, machine learning investigates the algorithmic construction and research that can learn from and make predictions on data (Ron, 1998). Static programming instructions in making data-driven decisions or predictions to strictly follow such algorithms are by constructing a model for input datasets. In a range of computing tasks, Machine learning is employed where outlining and programming express calculations with great execution is troublesome or unfeasible; case of such applications incorporate computer vision, email sifting, figuring out how to rank and identification of system gatecrashers or vindictive insiders working towards an information rupture and optical character recognition (OCR).

With computational insights, machine learning regularly covers and is firmly related which additionally concentrates on forecast making using computers. Machine learning has solid binds to scientific enhancement, which conveys hypothesis, strategies and application areas to the field. Data mining is now and then conflated with machine learning, where the last subfield concentrates more on exploratory data investigation and is called unsupervised learning. Likewise, machine learning can be utilized to learn and set up gauge behavioral profiles for different substances and after that used to discover significant oddities and consequently named unsupervised.

Machine learning inside the field of data investigation is a technique used to devise complex models and calculations that loan themselves to forecast. This is known as predictive analytics in commercial use. It is an investigative model that permits data researchers, scientists, investigators and designers to create repeatable choices, comes about, dependable and reveal concealed bits of knowledge through learning from patterns in the data and also verifiable connections.

Contingent upon the idea of the learning sign or criticism accessible to a learning framework, machine learning undertakings are commonly grouped into three general classifications:

4.1.1 Supervised Learning

Here, involving a teacher, the system is provided with sample of inputs and their desired outputs and the goal is to learn a general rule that maps inputs to targets. In other words, supervised learning comprises of algorithms which rely on labelled data for learning. Labelled data are data which capture input data and corresponding desired output (target) data. Basically, such algorithms when supplied with both input and desired output data, perform some computations, compare the output of the algorithm to the desired output, and then use errors accrued for modifying the parameters of the model accordingly (Caruana and Niculescu, 2006). Supervised learning is used in areas such as classification, regression, data compression, etc. However, one drawback of supervised learning will be seen in that supervised learning is only usable when data are labelled; unfortunately, most data in real life are not labelled or that the cost of labelling data is expensive. Figure 4.1 shows the supervised learning paradigm.

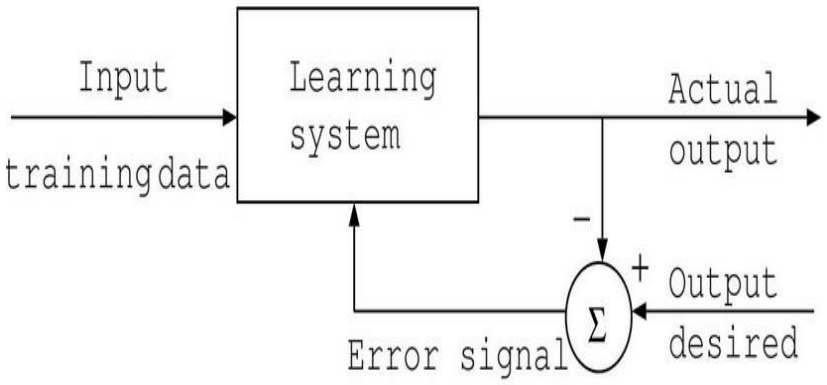

Figure 4.1: Supervised learning paradigm (Caruana and Niculescu, 2006)

4.1.2 Unsupervised Learning

In an unsupervised learning, the learning algorithm has no labels, leaving it on its own to find structure in its feature vectors/input. In itself, it can be a goal; finding a means towards an end (feature learning) or hidden patterns in data. This class of algorithms do not relying on labelled data. The only data that is required are the input data. There is no desired output supplied to such algorithms (Caruana and Niculescu, 2006). The learning algorithm is tasked with exploring data

for discovering interesting patterns and structures which form some sort of clusters based on data features. Figure 4.2 shows the unsupervised learning paradigm.

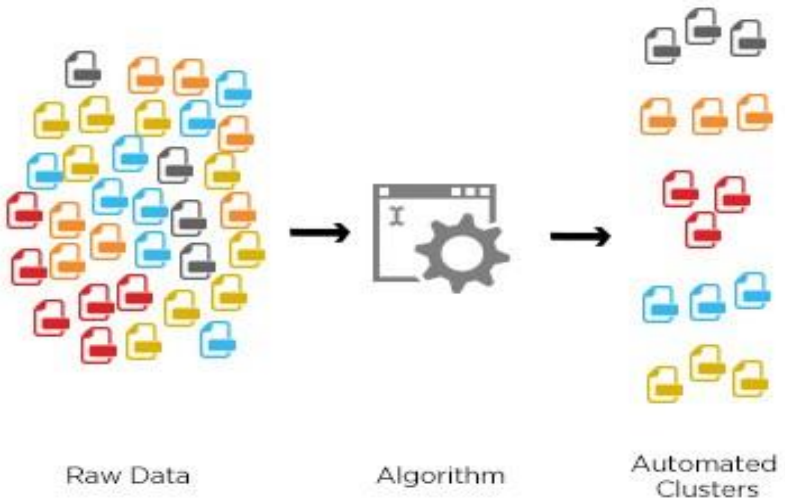

Figure 4.2: Unsupervised learning algorithm paradigm (Caruana and Niculescu, 2006)

4.1.3 Reinforcement Learning

In reinforcement learning, data are unlabelled and there are explicit desired output data, just as in the case of unsupervised learning. However, such algorithms are provided with positive and negative reward signals which serve as some sort of guide towards some optimal solution. Figure 4.3 shows the reinforcement learning paradigm.

In this algorithm, to perform a certain goal such as playing a game against an opponent or driving a vehicle, computer program interacts with a dynamic environment, where the program is provided feedback in terms of rewards and punishments as it navigates its problem space.

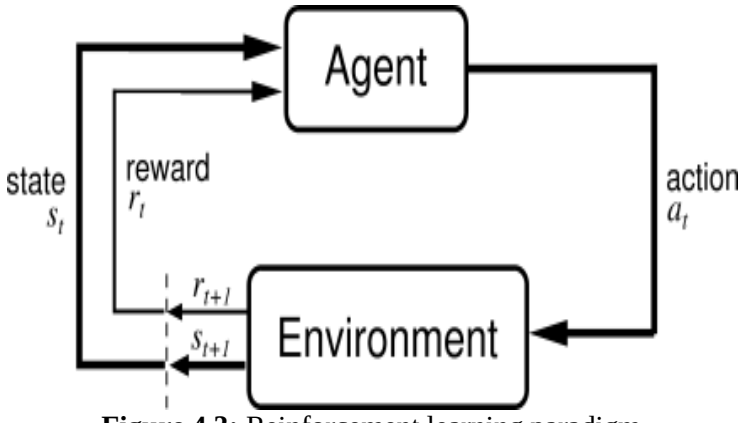

Figure 4.3: Reinforcement learning paradigm

Recently, the employments of different machine learning (ML) systems possibilities in prescription have turned out to be well known in the field of research. Interestingly, this study is motivated towards investigating whether the utilization of machine learning techniques could enhance the predictive strength of established models in legitimate characterization of breast tissues. An extensive variety of machine learning models including; genetic algorithms, fuzzy systems, artificial neural networks, support vector machines, decision trees, and various combinations of intelligent technologies have been used in medical prognosis modeling.

4.2 The Explored Machine Learning Techniques

This study examined three machine learning techniques for the classification of six classes of breast tissues deduced from electrical impedance spectroscopy. The three techniques include; Radial Basis Function Network (RBFN), Naïve Bayes (NB) and Back propagation Neural Network (BPNN).

4.2.1 Radial Basis Function Network (RBFN)

RBFN is a particular kind of neural system. In the most part, when people talk about mimicked neural networks or neural networks, they are implying the multilayer perceptron (MLP). In a MLP, each neuron takes the weighted aggregate of its data esteems. That is, each of the information regard is expanded by a coefficient, and the results are summed. A single MLP neuron is a clear straight classifier; however complex non-direct classifiers can be worked by merging these neurons into a network (Strumillo and Kaminski, 2003).

Generally, the approach in RBFN is more natural than that of the MLP. A RBFN performs gathering by measuring the information's closeness to cases from the readiness set. Each RBFN neuron stores a "model", which is just a single of the cases from the arrangement set. When we have to orchestrate data, each neuron forms the methodical detachment between the data and its model. Also, if the data all the more almost resembles the class A models than the class B models, it is named class A. This is shown in figure 4.4 below.

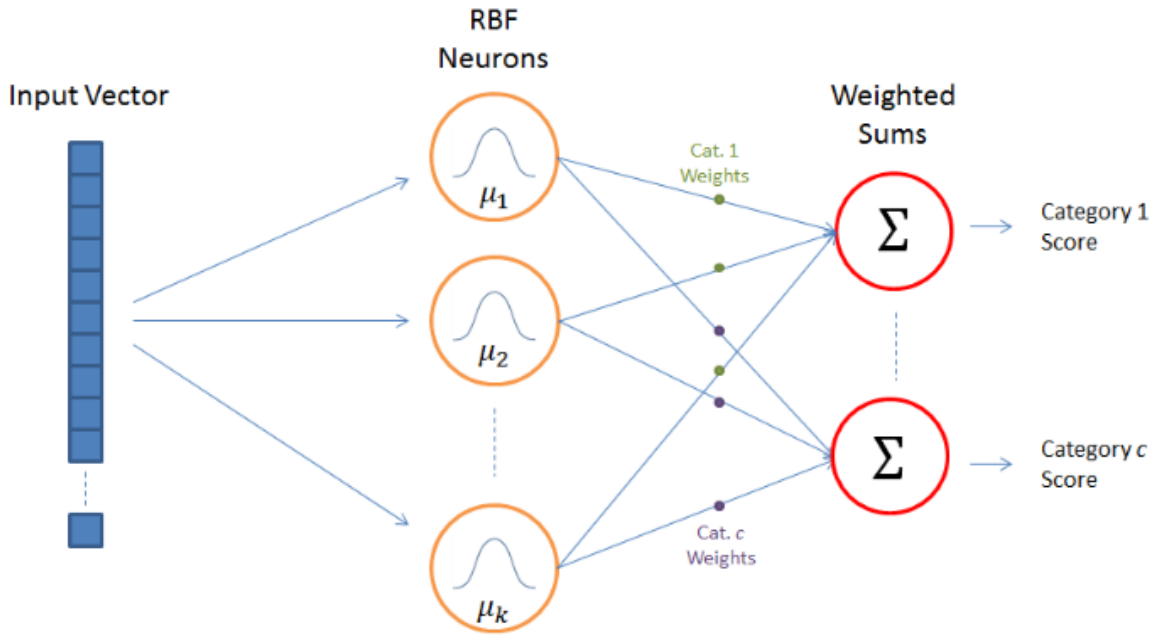

Figure 4.4: RBF Network architecture (Strumillo and Kaminski, 2003)

The above delineation demonstrates the common engineering of a RBFN model. It comprises of input vector, an RBF neurons layer and a target layer having a node for each classification or group of data.

The Input Vector: Satisfactorily, the n-dimensional vector is the input vector in which attempt is made to classified. The whole input vector is appeared to each neurons of the RBFN.

The RBF Neurons: Here, each neuron of the RBF stores a "model" vector which is only one of the vectors from the preparation group. Each neuron of the RBF looks at the input vector to its model, and outputs an incentive in the vicinity of 0 and 1 which is a measure of closeness. On the off chance that the input is equivalent to the model, then the neuron of the model output would be 1. When the separation between the input and model develops, the reaction exponentially tumbles off towards 0. The state of neuron of the RBF's reaction is a chime bend, as outlined in the network design chart.

The peak reaction of the neuron is additionally called its "actuation" score. The model vector is additionally frequently known as neuron's "middle", for it is the incentive at the focal point of the chime bend.

The Output Nodes: The output of the network comprises of an arrangement of nodes, one for each classification that we are attempting to be classified. Every node at the output layer processes a kind of score for the related classification. Regularly, the choice of the classification is made by mapping out the input to the class with the most outstanding score.

The score is prepared by taking a weighted sum of the established figures from every neuron of the RBF. By weighted sum we suggest that a target node relates a weight a motivation with each of the neurons in RBF, and copies the neuron's order by this weight before adding it to the total response.

Since each target node is figuring the score for an alternate classification, each node at the target layer has its own specific weights course of action. The output of the node would conventionally give a positive weight to the neurons of the RBF that have a place with its classification, and a negative weight to the others.

4.2.2 Naïve Bayes Algorithm

A naïve Bayesian network (NB), is a crucial case of a Bayesian technique that comprises of one class variable C which is conditional on a set of feature variables F: $\{F_1, \ldots, F_n\}$. All variables in F are assumed to be conditional independent from each other given C, meaning that $P(F_i|C, F\backslash F_i) = P(F_i|C)$ for each i. The arcs are going from the class node C to all feature nodes $F \epsilon F$ as shown in figure 4.5. Despite their over-simplified and naïve design independence assumptions, naïve Bayes classifier usually works well in several real-world situations that seemed complex.

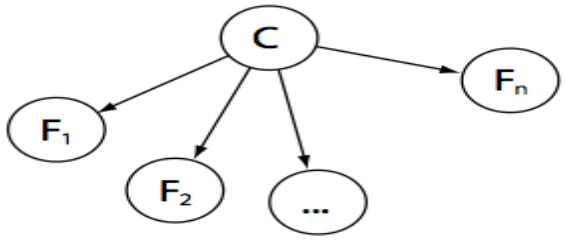

Figure 4.5: Naïve Bayes network

4.2.2.1 Tree Augmented Naïve Bayes Network

The real disadvantage of naïve Bayes is that it expects all element factors to be restrictively autonomous given the class variable. In practice, these variables are often strongly related to

each other. The tree augmented naïve Bayes, or TAN, algorithm sustains the fundamental structure of naïve Bayes, additionally allows each element node to have at most one other component node as a parent. This allows the model to capture dependencies between the feature nodes.

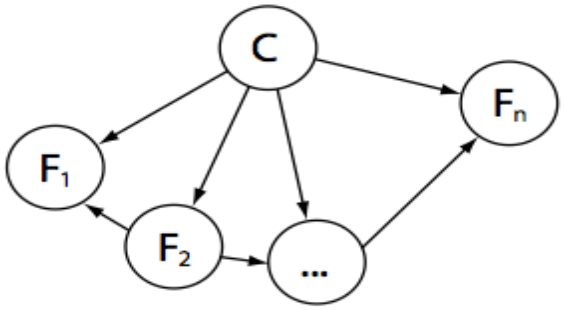

Figure 4.6: Tree augmented naïve Bayes network

4.2.2.2 Bayesian Classifiers

Bayesian networks can be used to infer probabilities. For instance, this thesis implored the use of Bayesian networks to inferring the probability of accurate classification of six classes of breast tissue.

A classifier is a mapping from discrete or continuous values to a set of labeled classes. This research explores the mapping of the output of a Bayesian network to the six classes of freshly excise breast tissue. By applying different threshold values the probabilities can be mapped into the different labeled classes of breast tissue.
With a binary classifier (two classes, e.g. negative (n) and positive (p)), there are four possible outcomes. If the result of the classifier is p and the actual value is p as well, this is known as true positive (TP) or hit. Moreover, if n is the actual value, this is known as false positive (FP) or 'false alarm'. When n is the actual value and n is the predicted value as well, this is called true negative (TN) or correct rejection. A false negative (FN) or miss signifies; the actual value is p.

With multiple cases and a classifier we can calculate model's sensitivity and specificity. The sensitivity is the number of correctly classified breast tissue instances divided by the total number of breast tissue instances.

$$Sensitivity = \frac{TP}{TP+FN} \quad 4.1$$

The specificity is the number of correctly classified non-cancerous instances divided by the total number of non-cancerous instances.

$$Specificity = \frac{TN}{TN+FP} \quad 4.2$$

4.2.3 Back propagation Neural Network

One of the most popular multilayer network models is the back propagation neural network (BPNN), explored in this thesis for modelling breast tissue classification. Figure 4.7 shows a simple architecture of BPNN. Also, in many literatures, it is not uncommon to find back propagation neural network referred to as multilayer perceptron (MLP).

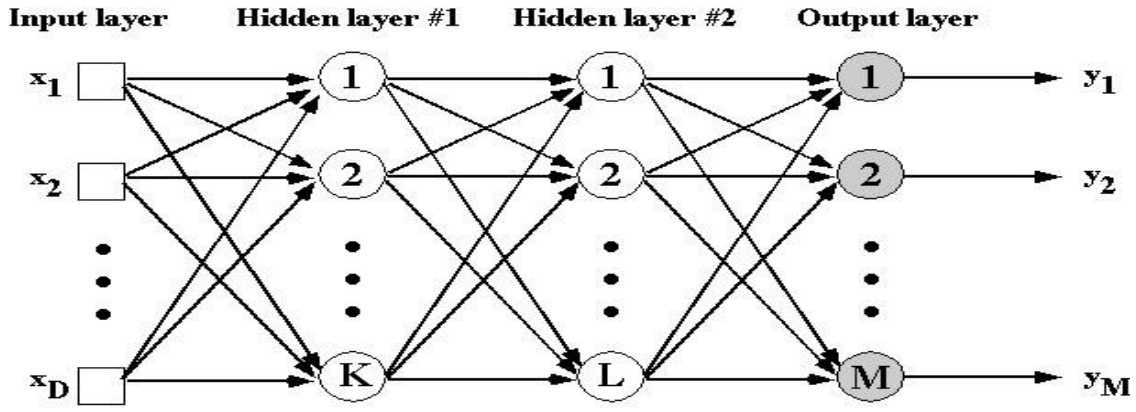

Figure 4.7: Back propagation neural network (BPNN) architecture (Eluyode and Akomolafe, 2013)

Furthermore, BPNN relies on learning scheme referred to as supervised learning for learning tasks. The supervised learning scheme is a situation where a model is supplied inputs and corresponding desired outputs (or targets). The back propagation neural network is basically a stacked of artificial neurons as layers (Eluyode and Akomolafe, 2013) Back propagation neural networks have at least three layers; the input, hidden and output layers.

The input layer is where independent (input) variables are supplied to the network, the hidden layer is primarily where the abstract features (associations) between the independent and dependent variables are extracted (or learned) and the output layer is where the computed dependent and target dependent are used to obtain network error for iteratively updating the parameters of the network. Note that back propagation neural network can have more than one

hidden layer; however, one hidden layer is sufficient for learning most tasks. The back propagation neural network is displayed in figure 4.7.

From Figure 4.7, the input layer is shown having input variables (attributes) x_1, x_2....x_D, hidden layer 1 having K neurons, hidden layer 2 having L neurons and target layer with M neurons. Note that the suitable number of neurons in hidden layers is determined heuristically, while the number of neurons in target layer depends on the task. Generally, for regression problems, the number of dependent variables is the number of neurons in the output layer. Back propagation basically operates in two phases which are forward pass and backward pass. In the forward pass, inputs attributes supplied to the network are propagation from the input layer to the output layer and the network outputs are computed. Differences between the desired outputs and actual (computed) outputs of the network are used to obtain error terms. During backward pass, the errors are propagated from the target layer to the input layer for updating network weights. Specifically, a cost function is defined for defining the error terms. The aim of learning in back propagation neural network is to minimize this cost function as learning progresses (i.e. iteratively) with weights update. At the end of a successful learning or training, the internal parameters of the model should demonstrate a smooth mapping function of inputs to corresponding outputs. Equation 4.3 shows a typical cost function for back propagation neural network, mean squared error (MSE) function.

$$MSE = \sum_{m=1}^{M} \left(t_m - y_m\right)^2 \tag{4.3}$$

Where, y is the actual (computed) output and t is the desired (target) output, m indexes output neurons and the number of output neuron is M respectively.

Weights of the output-hidden layer, w_{ml}, and corresponding bias weights, w_{mb} are updated using equations 4.4 and 4.5, respectively.

$$w_{ml}(i+1) = w_{ml}(i) + \eta \Delta_m O_l + \alpha[\delta w_{ml}(i)] \tag{4.4}$$

Where, learning rate is η, Δ_m is the hidden-output layer error signal for neuron m, output of hidden neuron l is O_l, α is the momentum rate, δw_{ml} is the previous weight change and i is the iteration index.

$$w_{mb}(i+1) = w_{mb}(i) + \eta\Delta_m O_b + \alpha[\delta w_{mb}(i)] \quad (4.5)$$

Where, the hidden-output bias neuron, O_b, is set to 1; δw_{mb} is the previous weight change for the hidden-output bias neuron. Note that the hidden-output error signal for neuron m, Δ_m, is calculated using equation 4.6.

$$\Delta_m = O_m.(1-O_m)(t_m - O_m) \quad (4.6)$$

The hidden-hidden layers weights, w_{lk}, are updated using 4.7

$$w_{lk}(i+1) = w_{lk}(i) + \eta\Delta_l .{}_k + \alpha[\delta w_{lk}(i)] \quad (4.7)$$

From (4.7), error signal of hidden layer neuron l is Δ_l, output of preceding hidden layer neuron k is O_k, δw_{lk} is the previous weight change for the hidden-hidden bias neuron. Note that Δ_l can be obtained using equation 4.8.

$$\Delta_l = O_l(1-O_l)\sum_{l=1}^{L} w_{ml}\Delta_m \quad (4.8)$$

The input-hidden layer weights, w_{kd}, and corresponding input-hidden bias weights, w_{kb} , are updated using equation 4.9 and 4.10.

$$w_{kd}(i+1) = w_{kd}(i) + \eta\Delta_k x_d + \alpha[\delta w_{kd}(i)] \quad (4.9)$$

Where, x is the network input, d indexes input variables (attributes), error signal of neuron k in the hidden layer following the input layer is Δ_k.

$$w_{kb}(i+1) = w_{kb}(i) + \eta\Delta_k O_b + \alpha[\delta w_{kb}(i)] \quad (4.10)$$

Where, the input-hidden bias neuron, O_b, is set to 1; δw_{kb} is the previous weight change for the input-hidden bias neuron. Note that the error signal for input-hidden neuron k, Δ_k, is calculated using equation 4.11.

$$\Delta_k = O_k(1-O_k)\sum_{k=1}^{K} w_{lk}\Delta_l \quad (4.11)$$

CHAPTER 5
SYSTEM DESIGN AND EXPERIMENTAL RESULT ANALYSIS

5.1 Overview

Algorithm is the basic component of any intelligent system. An algorithm process and generates knowledge from data. Creating systematic algorithmic models encompasses many steps. The first step is data preprocessing which is an essential step in data processing; filters and makes dataset ready for operations. Analyzing unfiltered data can generate inappropriate model or misleading results. Hence, the representation and quality of data is first and foremost before further analysis and classification. The second step as featured in this thesis is feeding the processed data onto the classifiers; radial basis function network (RBFN), naïve Bayes technique (NB) and back propagation neural network (BPNN).

5.2 Dataset Analysis

For high performance accuracy to be obtained, the dataset were normalized/rescaled to the range of [0, 1]. Table 5.1 summarized the breast tissue dataset used for the classification task.

Table 5.1: Breast tissue class description

Class code	Classes	Number of instances
1	Connective tissue	14
2	Adipose tissue	22
3	Glandular tissue	16
4	Carcinoma	21
5	Fibro-adenoma	15
6	Mastopathy	18

5.2.1 Feature Selection Method

Feature extraction method as implemented in the Naïve Bayes technique of this study, is often a crucial data processing stage before learning algorithmic implementation. It is an act of removing redundant and irrelevant information to improve the performance of machine learning algorithm. This thesis considered Sequential Feature Selection (SFS) method in the naïve Bayes algorithm to boost performance accuracy as well as execution time since naïve Bayes algorithm is spotted

to spending much time on data mapping before classification. The application of SFS to the breast tissue dataset reduced the dataset dimensionality by selecting just the subset of feature vectors (predictor variables) to perform the classification. The criteria for the selection minimized the predictive error of the different selected subsets. The encapsulated techniques searched for a subset of indicators that ideally models measured reactions, subject to limitations, for example, excluded or required elements and the span of the subset. Since the first significance and units of elements are imperative and the displaying objective is to recognize a persuasive subset, highlight choice is consequently desirable over component change. In the dataset, numerical transformations are inappropriate and categorical features are present, hence the choice of SFS as a means of dimension reduction.

Two components of SFS include: first is called the criterion also referred to as the objective function which minimizes all feature subsets that are feasible. Some of the common criteria include; regression models (MSE) and classification models (misclassification rate). The second part is the model for successive inquiry (sequential search algorithm), this removes or includes highlights from a competitor subset assessing the standard. consecutive ventures move in just a single course, continually contracting or continually developing the hopeful set when a comprehensive examination of the foundation esteem at all 2n subsets of n-attribute data collection is normally infeasible (contingent upon the span of n and the cost of target calls).

Two variants of sequential feature selection:

- The first explores a scenario where attributes are successively added to a vacant competitor set until the point that the option of further components does not diminish the criterion and it is known as Sequential forward selection (SFS).
- In the second scenario, attributes are successively expelled from a full applicant set until the point when the evacuation of further attributes increment the rule and this is called Sequential in backward selection (SBS).

The method examined a consecutive element determination system composed particularly for least-squares fitting known as stepwise regression. Where the functions (stepwisefit and stepwise) uses enhancements that are just conceivable with least-squares rules. Stepwise

regression may expel attributes that have been included or include attributes that have been evacuated, contrarily to generalized sequential feature selection.

"Sequentialfs" is the machine learning tool compartment criterion and Statistics does SFS. Here, input contentions are response data and target and reaction data and a function handle to a document that executes the measure work. Discretionary input enable the classifier to determine SBS or SFS, excluded or required elements, finally, the feature subset size. Criterion at different candidate sets is evaluated by calling the functions; cvpartition and crossval.

5.2.2 Cross Validation

In all classification and learning techniques, the datasets should be separated into the training set, testing set and validation set. In accordance with the number of datasets, the division of data into these three groups varies. If a large sample size is provided, 60% of the datasets should be considered training set, 20% of the dataset as the testing set while the final 20% of the dataset as the validation set. If a medium sample size is provided to the classification algorithm, 60% of the dataset should be allocated training set while 40% of the dataset as the test set. Finally, if the sample size is too small, having the sets of training with testing may not be a good idea. In such situation the use of cross validation technique would be presumed.

In cross validation before the training starts, a portion of the dataset is evacuated and if training is completed, the prior expelled dataset can be utilized to test the execution of the learned model on the new dataset. An entire class of model assessment strategies known as cross validation utilizes this fundamental thought. A few methods of cross validation incorporate; the K-fold cross validation method, leave-one-out method and finally holdout method.

One of the least complex sorts of cross validation strategies is the holdout strategy. With the holdout technique, the dataset is divided into two sets; the training and testing. Utilizing the training set, the approximator criterion fits a criterion. The target value for the dataset in the testing phase is then predicted by the function approximator. To evaluate the model, errors committed earlier are collected to give the mean total test set error. It requires no longer time to register and generally desirable over remaining strategy and this turns out to be the major advantage of this method. Moreover, it has a high variance evaluation. Determining which information focuses wind up in the testing set and which wind up in the training set vigorously

rely upon the model assessment of the current dataset, and along these lines the assessment might be altogether extraordinary relying upon how the dataset division is made.

One approach to enhance over the holdout technique is known as the k-fold cross validation. Where the dataset is separated into k subsets and a method called the holdout technique is iterated K times. Regularly, K-1 subsets are put together forming training set while one of the K subsets is utilized for testing. After this, the accumulated error over all K trials is figured. Ways in which the data are divided matters less, which is one of the advantages of this method. Each of the data point will appear in training phase K-1 times and will appear in a test phase exactly ones. If K is appreciated, the variance of the resulting estimate is reduced. Moreover, the training model must be rerun starting from the beginning K times, which implies it takes K times as much programming to make an assessment and this is the real impediment of this technique. Random division of the dataset to training and testing in K various times usually stand out as a variant of this method. It is observed that, one can freely pick how substantial each test set is and what number of trials he computes hence, flexible nature and advantage of using this method.

Another K-fold cross validation method is the Leave-one-out (LOO) method, with K equivalent to the quantity of data locations in the set (N). Meaning that in N isolate times, the approximator criterion is prepared on every one of the data aside from one point and an expectation is made for that point. Evaluating the model leads to the computation of average error. Evaluation error of LOO-XV is great be that as it may, at first pass, it appears to be exceptionally dreary to process.

Interestingly, LOO predictions made by locally weighted learners are simply as they make standard expectations. Meaning, LOO-XVE computation sets aside no greater opportunity to compute the lingering error which is a vastly improved approach to assess models.

In this thesis, two cross validation methods; K-fold and holdout are used apparently in all the models. But specifically, the holdout cross validation which is the simplest method was applied to the naïve Bayes algorithm such that, random sampling of the training as well as testing datasets would minimize model’s biasness.

5.3 Breast Tissue Classification

This thesis utilized radial basis function network (RBFN), Naïve Bayes (NB) and back propagation neural network (BPNN) techniques for the classification implementation of breast tissue. Experimental outcomes of the RBFN, NB and BPNN models demonstrated that RBFN outperformed NB and BPNN techniques with a remarkable margin. Figures 5.1 and 5.2 depict the flowchart diagram of the framework and the neural network topology respectively.

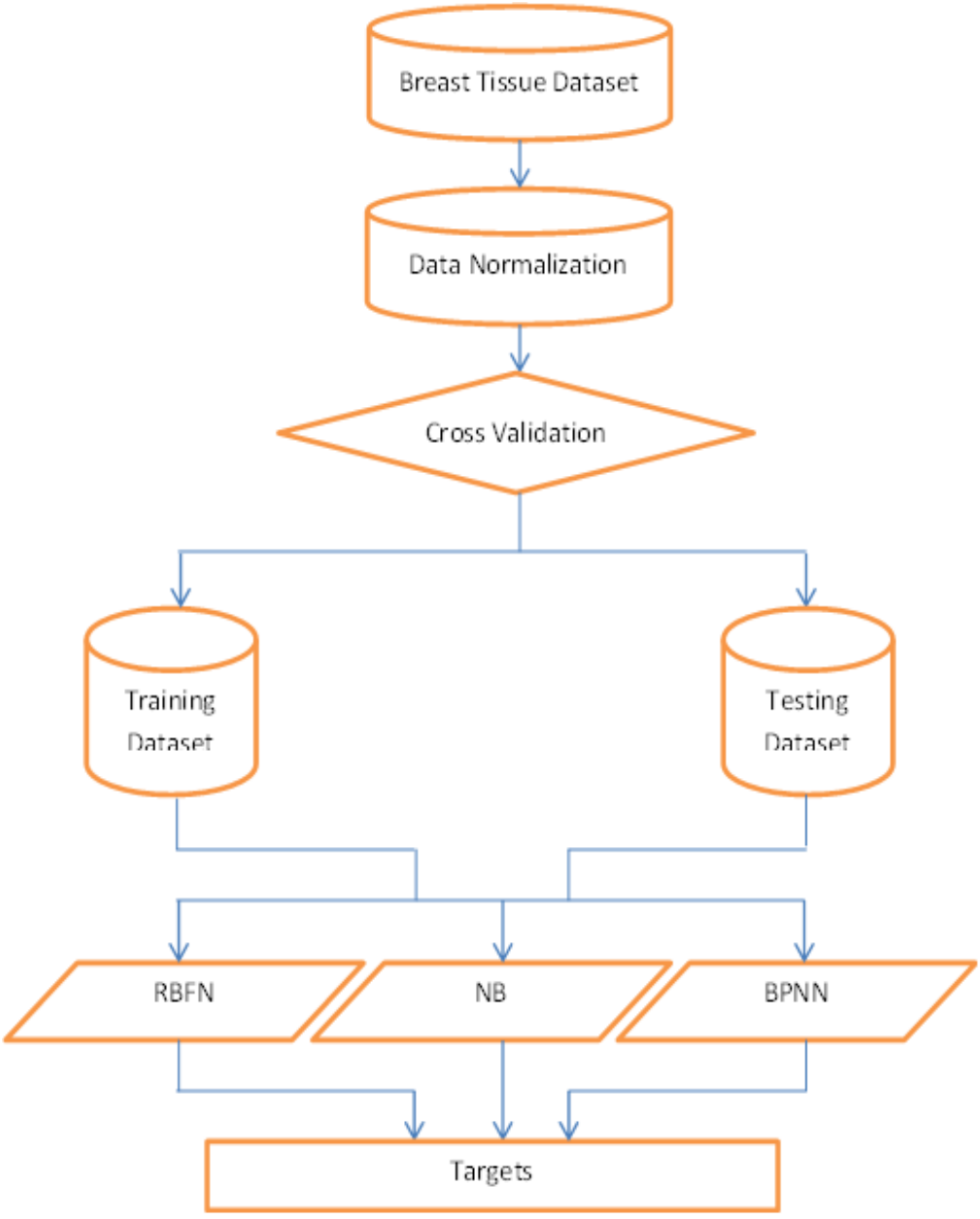

Figure 5.1: Flowchart diagram of the framework

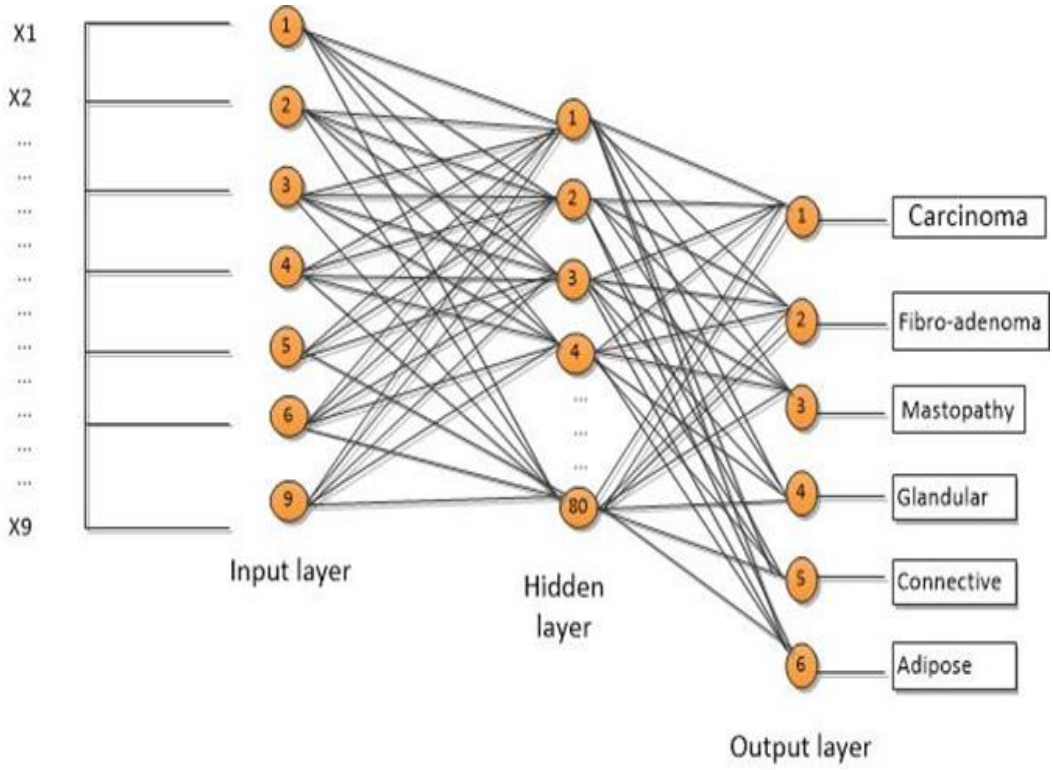

Figure 5.2: Topology of the neural network learning techniques

5.4 Classification Using RBFN

RBFN networks are very much related to back propagation networks topology. Basic irregularities are in the analogy behind weight computation. The activation function used at the neurons' outputs basically has one hidden layer. The hidden layers with regards to a neural system give a group of "criterions" (radial-basis functions) constituting arbitrarily the "basis" for input designs when they are ventured into hidden layer. (Helwan and Tantua, 2016).

The motivation behind RBFN and some other neural network classifiers is based on the knowledge that pattern transformed to a higher-dimensional space which is nonlinear is probably more to be linearly separable compare to that in the low-dimensional vector representations of same patterns (cover's separable theorem on patterns). The output of neuron units are calculated using k-means clustering similar algorithms, after which Gaussian function is applied to provide the unit final output. In the training phase, the hidden layer neurons are usually centered randomly in space on subsets or all of the training patterns space (dimensionality is of the training pattern) (Helwan and Abiyev, 2016). After this, the Euclidean distance between each neuron and training pattern vectors are calculated, then the RBF (also referred to as a kernel)

applied to calculated distances. Since the radius distance is the focal point to the function, hence the name; radial basis function (Helwan and Abiyev, 2015) as shown in equation 5.1

$$Weight = RBFN\ (distance) \tag{5.1}$$

Other functions such as logistic and thin-plate spline can be used in RBFN networks but in this thesis, the Gaussian function was utilized to execute the grouping. In training, radius of Gaussian function is usually chosen and this affects the extent to which neurons influences the considered distance. The multiplication of both the output values summation of the RBFs and weights computed for each neuron leads to the best predicted value for the new point as shown in (Helwan and Abiyev, 2016). The equation relating Gaussian function output to the distance from data points (r>0) to neurons center is given by:

$$\varphi(r) = e^{-r^2/2\sigma^2} \tag{5.2}$$

Where, the smoothness of the interpolating function is controlled using σ (Helwan and Tantua, 2016) and r is the Euclidean distance from a neuron center to the training data position.

5.4.1 RBFN Training

In order to obtain a stable and reliable result, three experiments were performed using radial basis function networks (RBFN1, RBFN2 and RBFN3) with different values of hidden neurons and spread constant trained on two different datasets; 70 instances (about 70%) for training and 36 instances (about 30%) for testing. This aims to observe the networks' performances when trained with different values of spread constant. Table 5.2 summarizes the training parameters of three RBFNs used in simulation.

Table 5.2: RBFNs training parameters

Network parameters	RBFN1	RBFN2	RBFN3
Number of training samples	70	70	70
Number of hidden neurons	30	50	80
Spread constant	0.14	0.5	1.0
Maximum epochs	50	50	50
Training time (secs)	10	7	9
Mean Square Error	0.0320	0.0309	0.0319

From the three RBFN experiments performed, the least/lowest mean square error obtained was at epochs 50 in experiment 2 (RBFN2). Figures 5.3, 5.4 and 5.5 demonstrate the learning curves of networks RBFN1, RBFN2 and RBFN3 respectively.

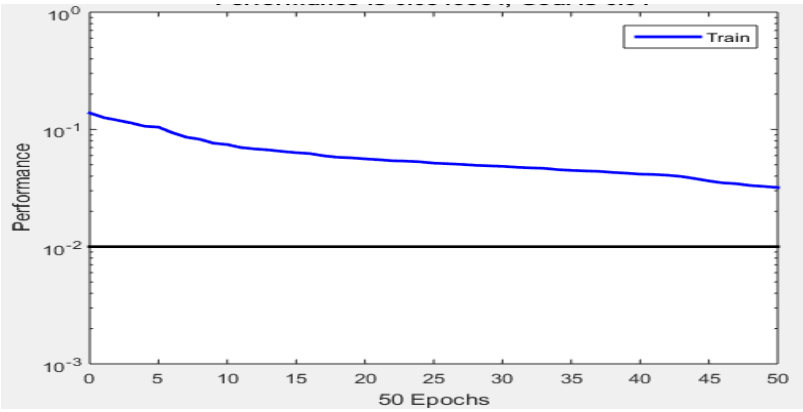

Figure 5.3: RBFN1 learning curve

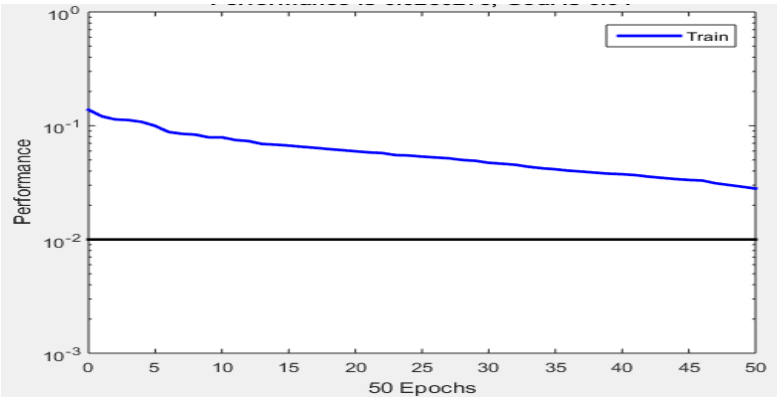

Figure 5.4: RBFN2 learning curve

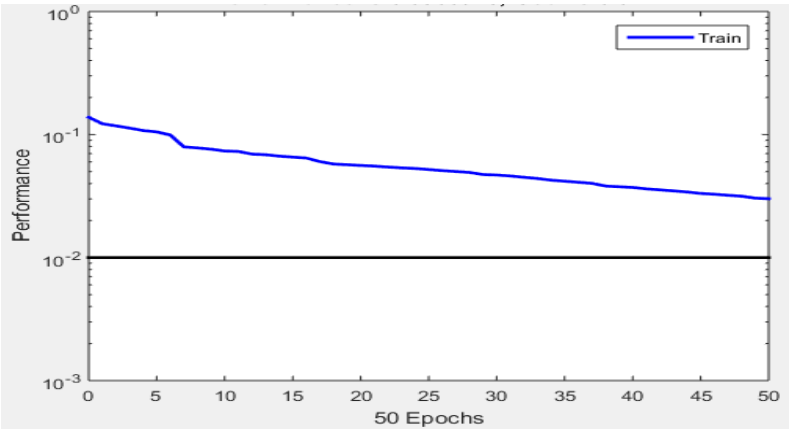

Figure 5.5: RBFN3 learning curve

5.4.2 RBFN Testing

Similarly, the Radial Basis Function networks (RBFN1, RBFN2 and RBFN3) were also tested using same configurations/parameters. The testing phase was assigned about 30% of the dataset. As shown in table 5.3, RBFN2 achieved the highest recognition rate (91.66%) amongst the three networks when tested on 30% of the data. Note that this network (RBFN2) did not outperform the other networks (RBFN1 and RBFN3) in the training phase, as it achieved 95.5% recognition rate while RBFN1 and RBFN3 obtained 97.6% and 97.28% respectively.

Table 5.3: RBFNs training and testing results

Network parameters	RBFN1	RBFN2	RBFN3
Number of training samples	70	70	70
Correctly classified training instance	69	67	68
Training recognition rate	97.6%	95.5%	97.28%
Number of test instances	36	36	36
Correctly classified test samples	30	33	31
Recognition rate on testing	83.33%	91.66%	86.11%
Overall recognition rate	93.39%	94.33%	93.39%

5.5 Classification Using Naïve Bayes Technique

Using Bayes theorem of posterior probability, this thesis also proposed a Naïve Bayes classification algorithm. Given the instances of breast tissue, the conditional probabilities of the classes are computed by the algorithm and it picks the class with the highest posterior. The Naïve Bayes (NB) classification assumes that attributes/features are independent. By traditional counting, estimation of nominal attributes is by probabilities, while the assumption of all normal distribution for each class and attribute estimates continuous attributes. The algorithm is programed to simply skip unknown attributes.

5.5.1 Naïve Bayes Training

Considering Naïve Bayes technique training phase, Cross Validation Approach (CVA) was explored and applied to the breast tissue dataset having 106 observations/instances. Execution of the CVA automatically partitioned the dataset into two; first is training set having 96 observations and the second is test set having 10 observations. The CVA was incorporated into the proposed Naïve Bayes technique and result of the Naïve Bayes machine learning technique without the use of Sequential Feature Selection (SFS) algorithm yielded the total performance accuracy of 80%. And when implemented with SFS inclusion, the combined algorithms yielded the total performance accuracy of 70%. This implies that SFS inclusion to the Naïve Bayes algorithm did not boost the performance of the system. But rather reduced the performance accuracy by 10%. Observations from this experimental result proved that it is totally not advisable to use SFS on datasets with fewer features since further reduction of the feature vectors would result to misclassification and possibly poor performance accuracy as shown in tables 5.4 and 5.5. The two tables depict the tabular experimental result representation of the Naïve Bayes machine learning technique respectively.

Table 5.4: Sequential feature selection data analysis

Computed parameters	Tagged columns	Criterion values
Step 1	1	0.3
Number of connected workers	0	NILL
Number of feature vectors included	1	NILL

From table 5.4 above, it is clearly noted that the sequential feature selection method dimensionally reduced the breast tissue dataset to just one subset within column 1 with respective criterion value; 0.3. The feature extraction process could not continue after reaching the minimum criterion value of 0.3. The one automatically generated feature subset was fed into the Naïve Bayes learning algorithm for further classification. Table 5.5 below shows the Training parameters and performance of Naïve Bayes (NB) technique.

Table 5.5: Training parameters and performance of Naïve Bayes (NB) technique

Learning parameters	NB training with SFS	NB training without SFS
Number of training observations	96	96
Number of testing observations	10	10
Number of connected workers	NILL	NILL
Number of feature vectors included	1	0
Recognition rate	70%	80%

Table 5.5 clearly shows the poor performance of the Naïve Bayes algorithm when used with sequential feature selection method. The model recorded such poor performance because it is unethical to use feature extraction method on datasets with fewer features/attributes.

5.6 Classification Using BPNN

The back propagation neural network (BPNN) uses a feed forward process, a back propagation updating method, and supervised learning topology. This algorithm was the reason of neural networks development in the 80s of the last century. Back propagation is a general purpose learning algorithm. Although it is very efficient, it is costly in terms of processing requirements for learning. Back propagation network with a given hidden layer of features can simulate any function to any accuracy level (Helwan and Tantua, 2016). These networks have been extensively used for solving different problems (Helwan and Abiyev, 2016) to (Mamedov and Abiyev, 2001). These are related breast cancer identification (Helwan and Abiyev, 2016), (Helwan and Abiyev, 2015), face recognition (Abiyev, 2014), iris recognition (Abiyev and Altunkaya, 2009) to (Abiyev and Altunkaya, 2007), for control purpose ()Abiyev and Altunkaya, 2009, for channel equalization (Abiyev and Aliev, 1994).

The back propagation algorithm is still as simple as it was in its first days. That is due to its simple principle and efficient algorithm. At the first layer of the network, the input set of training data is presented and the input layer passes this data to the next layer where the processing of data happens. The results after being passed through the activation functions are then passed to the output layers. The result of the whole network is then contrasted with a desired target. And the error is used to make a one update of the weights preparing for a next iteration. After the adjustment of the weights, the inputs are passed again to the input, hidden, and output layers and a new error is calculated in a second iteration and vice versa (Mamedov and Abiyev, 2001).

The back propagation is an algorithm that uses the theory of error minimization and gradient descent to find the least squared error. Finding the least squared error imposes the calculation of gradient of the error for each of the iterations. As a result, the error function must be a continuous derivable function. These conditions lead to the use of continuous derivable activation functions as they are the precedents of error calculation (Helwan et al., 2016). In most cases, the tangent or logarithmic sigmoid functions are used. The pseudo-code algorithm for BPNN is to choose the initial weights randomly while error is too large. It is further expressed thus;

For random order presentation of each training pattern:

- Inputs application to the network
- For every neuron, output is calculated from input unit through the hidden unit(s), to target unit
- Error at the targets is calculated
- Error signs for pre-target units is computed using output error
- Weight adjustments are computed using error signals
- weight modifications is the applied

Evaluate the network performance periodically by:

- Each input parameter's value is applied to each input node
- Identity function is computed only by input nodes

5.6.1 BPNN Training

The breast tissue classification system was trained with 70 cases (about 70% of the data) and tested with 36 cases (about 30% of the data). The developed BPNN network for this classification comprises of 9 input neurons in its input layer to accommodate the 9 features. The output layer has 6 output neurons to accommodate the 6 classes in the breast tissue dataset; Fibro-adenoma, Carcinoma, Glandular, Mastopathy, Adipose and Connective tissue. Appropriate number of hidden layer neurons is found experimentally while training the network. Figure 5.2 above shows the appropriate network topology. In figure 5.2, X1, X2, …., X9 represents the features used for each different tissue.

The developed back propagation neural network (BPNN) was trained on 70 cases of the 6 breast tissues as discussed above. For the purpose of network optimization, different values of hidden neurons, learning rate and momentum rate were considered. Thus, three networks (BPNN1, BPNN2 and BPNN3) were compared, each with different values of hidden neurons, learning rate and momentum rate. Table 5.6 shows the training parameters of the three networks.

Table 5.6: BPNNs training parameters

Network parameter	BPNN1	BPNN2	BPNN3
Number of training samples	70	70	70
Number of hidden neurons	20	50	80
Learning rate (η)	0.03	0.13	0. 3
Momentum rate (α)	0.55	0.65	0.85
Maximum epochs	3000	3000	3000
Training time (secs)	49	52	40
Mean Square Error	0.03558	0.03452	0.03352

It can be seen in Table 5.6 that BPNN3 with 80 hidden neurons, learning rate (0.3), and momentum rate (0.85) achieved the lowest mean square error (MSE) of 0.03352s. This MSE was reached in 40 seconds. The learning curve that shows the convergence of the network is a plot of the mean square error versus the increase of iteration number as shown in figures 5.6, 5.7 and 5.8 representing networks BPNN1, BPNN2 and BPNN3 respectively.

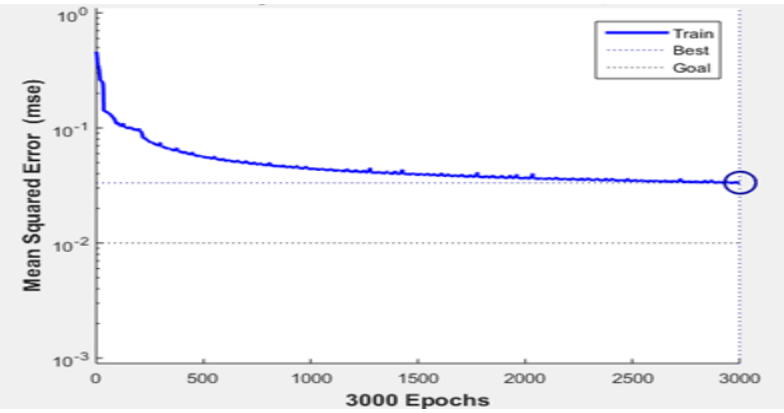

Figure 5.6: BPNN1 MSE plot

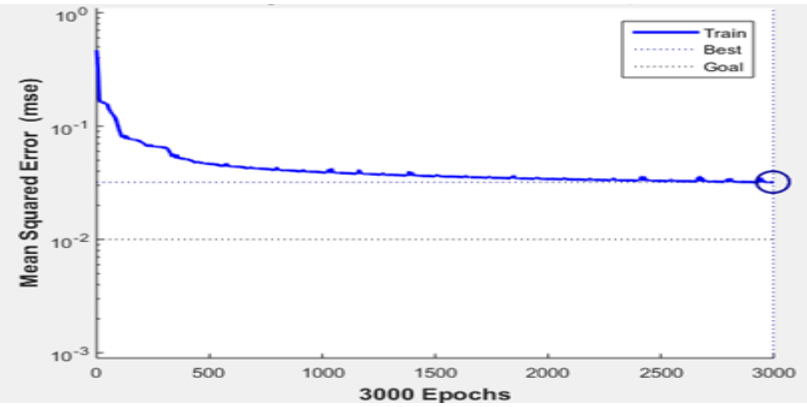

Figure 5.7: BPNN2 MSE plot

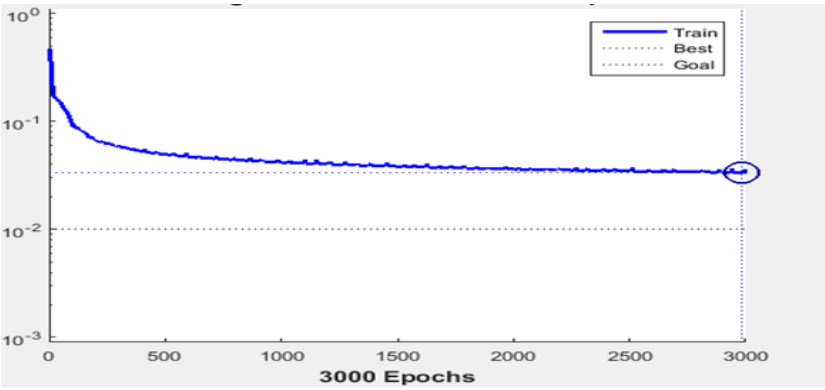

Figure 5.8: BPNN3 MSE plot

5.6.2 BPNN Testing

Both networks (BPNN1, BPNN2 and BPNN3) were tested on the available 30% of the data. The back propagation neural network models show a good generalization capability as shown in table 5.7. Although, overall recognition rates of the three implemented networks are found to be slightly different. It can be seen that the BPNN3 that uses 80 neurons at hidden layer reached the highest recognition rate of 93.39%. This means that this network has motivating generalization capabilities when unseen data of breast tissues are applied. Moreover, this network "BPNN3" outperformed BPNN1 and BPNN2 in terms of minimum error reached in the shortest time.

Table 5.7: BPNNs training and testing results

Network parameter	BPNN1	BPNN2	BPNN3
Number of training samples	70	70	70
Correctly classified training samples	65	66	66
Recognition rate on training	91.7%	92.6%	93.7%
Number of test samples	36	36	36
Correctly classified test samples	30	30	33
Recognition rate on testing	83.33%	83.33%	91.67%
Overall recognition rate	89.62%	90.56%	93.39%

5.7 Experimental Result Discussion and Comparison

The developed classification framework based on neural networks (RBFN, NB and BPNN) are shown to be capable of classifying breast tissues into 6 different classes; Fibro-adenoma, Carcinoma, Glandular, Mastopathy, Adipose and Connective tissues. The used networks in this thesis demonstrated promising performances as depicted in this chapter.

Best generalized (testing phase) classification accuracies obtained from the RBFNs, NBs and BPNNs are 91.66%, 80% and 91.67% respectively. It is remarked that the BPNNs achieved higher recognition rates on the test data. That is, better generalization capability as compared to

RBFNs and NBs. Also, considering RBFNs experiments, it is noted that the RBFN2 network's performance was the highest in the testing phase, but it was not the highest in the training phase. This proofs that a neural network can learn fast and accurately in the training phase, however it can be weak in generalizing; that is recognizing unseeing data. This may be due to training parameters values that allow the networks to get stuck in local optima.

Moreover, the mean square error reached for the RBFNs after convergence was less than that of BPNNs. Note that the difference of the error was not that high between both types of networks. However, the time taken for the networks to reach that error is what matters; since it was very high for BPNNs as compared to that taken for the RBFNs to reach its minimum error. Also it is important to spot on the training times for the BPNNs which are roughly 6 times those of the RBFNs and NB.

CHAPTER 6
CONCLUSION

This thesis presents an intelligent classification system for breast tissue dataset obtained using attributes of electrical impedance spectroscopy. Radial basis function network (RBFN), naïve Bayes (NB) technique and feed forward neural networks based back propagation algorithm (BPNN) were selected for this classification task. A comparison between these network models was made based on different parameters set during the training phase. Also, a comparison between the three types of neural networks was performed in order to assess accuracy of each as well as to discover network that performs best in the classification task.

The experimental observations demonstrated that a BPNN with more hidden layer performs better when trained and tested on unseen data. In addition, this network reached the least minimum square error in a shorter time than the other back propagation networks. For these networks (BPNN1, BPNN2 and BPNN3), the experiment that achieved the highest training recognition rate is the experiment that achieved the highest testing recognition rate.

In contrast, for the RBFNs (RBFN1, RBFN2 and RBFN3); the experiment that achieved the highest training recognition rate is different from the experiment the achieved the highest recognition in the testing phase. In other words, the network that reached the highest training recognition rate was not capable of achieving the highest recognition rate in the testing phase. This means that a network can be weak in generalization even if it performed well in the training phase. This is the major reason why different networks of the same type were utilized.

Furthermore, at the training phase of the naïve Bayes technique, sequential feature selection (SFS) method was examined. Interestingly, between the two experiments; NB algorithm without SFS and NB algorithm with SFS inclusion, it was recorded that NB without the use of SFS outperformed the experiment involving NB and SFS inclusion. This was due to the size of the dataset having fewer features. The experimental results show that SFS inclusion to the Naïve Bayes algorithm did not boost the performance of the system. But rather reduced the performance accuracy by 10%. Observations from this experimental result proved that it is totally not advisable to use SFS on datasets with fewer features since further reduction of the feature vectors would result to misclassification and possibly poor performance accuracy

Finally, general overview of the entire experiments depict that the back propagation neural network outperformed both the naïve Bayes technique and the radial basis function network for classifying six different breast tissues. This outperformance is in terms of accuracy, minimum error, maximum epochs and training time.

Future contributions to this problem would feature the repetition of the experiment using other machine learning techniques such as co-adaptive neuro-fuzzy inference system (CANFIS), extreme learning machines (ELMs), deep learning and support vector machines (SVMs) to obtaining more optimal results.

REFERENCES

Abbod, M. F., Catto, J. W. F., Chen, M., Linkens, D. A., & Hamdy, F. C. (2004). Artificial intelligence for the prediction of bladder cancer. *Biomedical Engineering: Applications, Basis and Communications, 16*(02), 49-58.

Abbod, M. F., Catto, J. W., Chen, M. Y., Linkens, D. A., & Hamdy, F. C. (2005, April). Artificial intelligence techniques for the prediction of bladder cancer progression. In *ESANN* (pp. 109-114).

Abiyev, R. H. (2014). Facial Feature Extraction Techniques for Face Recognition. *Journal of Computer Science*, *10*(12), 2360.

Abiyev, R. H., & Altunkaya, K. (2007, October). Neural network based biometric personal identification. In *Frontiers in the Convergence of Bioscience and Information Technologies, 2007. FBIT 2007* (pp. 682-687). IEEE.

Abiyev, R. H., & Altunkaya, K. (2009). Neural network based biometric personal identification with fast iris segmentation. *International Journal of Control, Automation and Systems*, *7*(1), 17-23.

Abiyev, R., & Altunkaya, K. (2008). Personal Iris Recognition Using Neural Networks. *International Journal of Security and its Application,* 2:12-24.

Abiyev, R., & Kilic, K. (2009). Adaptive Iris Segmentation. *Advances in Information Security and Assurance*, 90-99.

Aliev, R. A., Aliev, R. R., & Abiev, R. H. (1994). Synthesis of industrial neural controllers. In *Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on* (Vol. 3, pp. 1644-1649). IEEE.

Bassett, L. W., & Gold, R. H. (1987). Breast cancer detection: Mammography and other methods in breast imaging.

Berman, J. J. (2004). Tumor classification: molecular analysis meets Aristotle. *BMC cancer*, *4*(1), 10.

Bindu, G., & Mathew, K. T. (2007). Characterization of benign and malignant breast tissues using 2-D microwave tomographic imaging. *Microwave and Optical Technology Letters*, *49*(10), 2341-2345.

Bindu, G., Lonappan, A., Thomas, V., Aanandan, C. K., & Mathew, K. T. (2006). Active Microwave Imaging for Breast Cancer Detection. *In Electromagnetics Research* (pp. 149-169).

Bland, K., & Copeland, E. M. (2004). The Breast: Comprehensive Management of Benign and Malignant Disorders.

Bridges, J. E., Hagness, S., Taflove, A., & Popovic, M. (2002). *U.S. Patent No. 6,421,550*. Washington, DC: U.S. Patent and Trademark Office.

Brown, B. H., Barber, D. C., Morice, A. H., & Leathard, A. D. (1994). Cardiac and respiratory related electrical impedance changes in the human thorax. *IEEE Transactions on Biomedical Engineering*, *41*(8), 729-734.

Campbell, A. M., & Land, D. V. (1992). Dielectric properties of female human breast tissue measured in vitro at 3.2 GHz. *Physics in medicine and biology*, *37*(1), 193.

Caruana, R., & Niculescu *M. (2006).* An Empirical Evaluation of Supervised Learning in High Dimensions. In *SIGIR* (pp. 390–397).

Catto, J. W., Abbod, M. F., Linkens, D. A., & Hamdy, F. C. (2006). Neuro-fuzzy modeling: an accurate and interpretable method for predicting bladder cancer progression. *The Journal of urology*, *175*(2), 474-479.

Catto, J. W., Abbod, M. F., Linkens, D. A., Larré, S., Rosario, D. J., & Hamdy, F. C. (2009). Neurofuzzy modeling to determine recurrence risk following radical cystectomy for nonmetastatic urothelial carcinoma of the bladder. *Clinical Cancer Research*, *15*(9), 3150-3155.

Catto, J. W., Linkens, D. A., Abbod, M. F., Chen, M., Burton, J. L., Feeley, K. M., & Hamdy, F. C. (2003). Artificial intelligence in predicting bladder cancer outcome. *Clinical Cancer Research*, *9*(11), 4172-4177.

Chaudhary, S. S., Mishra, R. K., Swarup, A., & Thomas, J. M. (1984). Dielectric properties of normal & malignant human breast tissues at radiowave & microwave frequencies.

Chauveau, N., Hamzaoui, L., Rochaix, P., Rigaud, B., Voigt, J. J., & Morucci, J. P. (1999). Ex vivo discrimination between normal and pathological tissues in human breast surgical biopsies using bioimpedance spectroscopy. *Annals of the New York Academy of Sciences*, *873*(1), 42-50.

Choi, J. W., Cho, J., Lee, Y., Yim, J., Kang, B., Keun Oh, K., & Kwon, Y. (2004). Microwave detection of metastasized breast cancer cells in the lymph node; potential application for sentinel lymphadenectomy. *Breast cancer research and treatment*, *86*(2), 107-115.

Choi, J. W., Cho, J., Lee, Y., Yim, J., Kang, B., Keun Oh, K., & Kwon, Y. (2004). Microwave detection of metastasized breast cancer cells in the lymph node; potential application for sentinel lymphadenectomy. *Breast cancer research and treatment*, *86*(2), 107-115.

Cruz, J. A., & Wishart, D. S. (2006). Applications of machine learning in cancer prediction and prognosis. *Cancer informatics*, *2*, 59.

Davis, S. K., Van Veen, B. D., Hagness, S. C., & Kelcz, F. (2008). Breast tumor characterization based on ultrawideband microwave backscatter. *IEEE transactions on biomedical engineering*, *55*(1), 237-246.

Eluyode, O. S., & Akomolafe, D. T. (2013). Comparative study of biological and artificial neural networks. *European Journal of Applied Engineering and Scientific Research*, *2*(1), 36-46.

Enzinger, F. M., & Weiss, S. W. (1995). Soft Tissue Tumors. Third Edition. *Mosby - Year Book, Inc. Missouri.*

Fricke, H., & Morse, S. (1926). The electrical properties of tumors of the breast. In *J Cancer Res* (pp. 310-376).

Gorey, T. (2006). The Breast in Health and Illness: An Information Guide for Patients and Their Carers (DVD). *AstraZeneca Pharmaceuticals (Irl.) Ltd*.

Guliato, D., Rangayyan, R. M., Carvalho, J. D., & Santiago, S. A. (2008). Polygonal modeling of contours of breast tumors with the preservation of spicules. *IEEE Transactions on Biomedical Engineering*, *55*(1), 14-20.

Haemmerich, D., Ozkan, O. R., Tsai, J. Z., Staelin, S. T., Tungjitkusolmun, S., Mahvi, D. M., & Webster, J. G. (2002). Changes in electrical resistivity of swine liver after occlusion and postmortem. *Medical and Biological Engineering and Computing*, *40*(1), 29-33.

Hagness, S. C., Taflove, A., & Bridges, J. E. (1999). Three-dimensional FDTD analysis of a pulsed microwave confocal system for breast cancer detection: Design of an antenna-array element. *IEEE Transactions on Antennas and Propagation*, *47*(5), 783-791.

Halter, R. J., Zhou, T., Meaney, P. M., Hartov, A., Barth Jr, R. J., Rosenkranz, K. M., ... & Paulsen, K. D. (2009). The correlation of in vivo and ex vivo tissue dielectric properties to validate electromagnetic breast imaging: initial clinical experience. *Physiological measurement*, *30*(6), S121.

Heinitz, J., & Minet, O. (1995). Dielectric properties of female breast tumors. In *Electrical Bio-impedance* (pp. 356-359).

Helwan, A., & Abiyev, R. (2015). An Intelligent System for Identification of Breast Cancer. In *Biomedical Engineering* (pp. 17-20).

Helwan, A., & Abiyev, R., (2016). Shape and Texture Features for Identification of Breast Cancer. *Engineering and Computer Science* (pp. 19-21).

Helwan, A., & Tantua, D. P. (2016). IKRAI: intelligent knee rheumatoid arthritis identification. *International Journal of Intelligent Systems and Applications*, *8*(1), 18.

Helwan, A., Khashman, A., Olaniyi, E. O., Oyedotun, O. K., & Oyedotun, O. A. (2016). Seminal quality evaluation with RBF neural network. *Bulletin of the Transilvania University of Brasov. Mathematics, Informatics, Physics. Series III*, *9*(2), 137.

Henderson, R. P., & Webster, J. G. (1978). An impedance camera for spatially specific measurements of the thorax. *IEEE Transactions on Biomedical Engineering*, (3), 250-254.

Ho, W. T., & Lam, P. W. T. (2003). Clinical performance of computer-assisted detection (CAD) system in detecting carcinoma in breasts of different densities. *Clinical Radiology*, *58*(2), 133-136.

Joines, W. T. (1984). Frequency-dependent absorption of electromagnetic energy in biological tissue. *IEEE Transactions on Biomedical Engineering*, (1), 17-20.

Joines, W. T., Zhang, Y., Li, C., & Jirtle, R. L. (1994). The measured electrical properties of normal and malignant human tissues from 50 to 900 MHz. *Medical physics*, *21*(4), 547-550.

Jossinet, J. (1996). Variability of impedivity in normal and pathological breast tissue. *Medical and biological engineering and computing*, *34*(5), 346-350.

Jossinet, J. (1998). The impedivity of freshly excised human breast tissue. *Physiological measurement*, *19*(1), 61.

Jossinet, J. (1998). The impedivity of freshly excised human breast tissue. *Physiological measurement*, *19*(1), 61.

Jossinet, J. (1998). The impedivity of freshly excised human breast tissue. *Physiological measurement*, *19*(1), 61.

Jossinet, J., & Lavandier, B. (1998). The discrimination of excised cancerous breast tissue samples using impedance spectroscopy. *Bioelectrochemistry and bioenergetics*, *45*(2), 161-167.

Jossinet, J., & Schmitt, M. (1999). A review of parameters for the bioelectrical characterization of breast tissue. *Annals of the New York academy of sciences*, *873*(1), 30-41.

Jossinet, J., & Schmitt, M. (1999). A review of parameters for the bioelectrical characterization of breast tissue. *Annals of the New York academy of sciences*, *873*(1), 30-41.

Kerner, T. E., Paulsen, K. D., Hartov, A., Soho, S. K., & Poplack, S. P. (2002). Electrical impedance spectroscopy of the breast: clinical imaging results in 26 subjects. *IEEE transactions on medical imaging*, *21*(6), 638-645.

Kim, S. (2012). Assessment of breast tumor size in electrical impedance scanning. *Inverse problems*, *28*(2), 025004.

Kubicek, W. G., Patterson, R. P., & Witsoe, D. A. (1970). Impedance cardiography as a noninvasive method of monitoring cardiac function and other parameters of the cardiovascular system. *Annals of the New York Academy of Sciences*, *170*(1), 724-732.

Lazebnik, M., Popovic, D., McCartney, L., Watkins, C. B., Lindstrom, M. J., Harter, J., & Temple, W. (2007). A large-scale study of the ultrawideband microwave dielectric properties of normal, benign and malignant breast tissues obtained from cancer surgeries. *Physics in medicine and biology*, *52*(20), 6093.

Lazebnik, M., Popovic, D., McCartney, L., Watkins, C. B., Lindstrom, M. J., Harter, J., & Temple, W. (2007). A large-scale study of the ultrawideband microwave dielectric properties of normal, benign and malignant breast tissues obtained from cancer surgeries. *Physics in medicine and biology*, *52*(20), 6093.

Lemoine, K., & Nicholas R. (2001). Progress in pathology. *London: Greenwich Medical Media. p. 52. ISBN*, 9781841100500.

Lisboa, P. J., & Taktak, A. F. (2006). The use of artificial neural networks in decision support in cancer: a systematic review. *Neural networks*, *19*(4), 408-415.

Malich, A., Scholz, B., Kott, A., Facius, M., Fischer, D. R., & Freesmeyer, M. G. (2007). The impact of lesion vascularisation on tumours detection by electrical impedance scanning at 200 Hz. *Biomedical imaging and intervention journal*, *3*(4).

Mamedov, F., & Abiyev, R. (2001). Channel equalization using recurrent neural network. *Soft Computing and Computing with Words in System Analysis, Decision and Control* (pp. 327-330).

McGivney, D., Calvetti, D., & Somersalo, E. (2012). Quantitative imaging with electrical impedance spectroscopy. *Physics in Medicine and Biology*, *57*(22), 7289.

Meaney, P. M., Fanning, M. W., Li, D., Poplack, S. P., & Paulsen, K. D. (2000). A clinical prototype for active microwave imaging of the breast. *IEEE Transactions on Microwave Theory and Techniques*, *48*(11), 1841-1853.

Morimoto, T., Kimura, S., Konishi, Y., Komaki, K., Uyama, T., Monden, Y., & Iritani, D. T. (1993). A study of the electrical bio-impedance of tumors. *Journal of Investigative Surgery*, *6*(1), 25-32.

Morimoto, T., Kinouchi, Y., Iritani, T., Kimura, S., Konishi, Y., Mitsuyama, N., & Monden, Y. (1990). Measurement of the electrical bio-impedance of breast tumors. *European surgical research*, *22*(2), 86-92.

Nguyen, T. M., & Rangayyan, R. M. (2006, January). Shape analysis of breast masses in mammograms via the fractal dimension. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the* (pp. 3210-3213). IEEE.

Pepek, J. M., Chino, J. P., Marks, L. B., D'amico, T. A., Yoo, D. S., Onaitis, M. W., ... & Kelsey, C. R. (2011). How well does the new lung cancer staging system predict for local/regional recurrence after surgery?: A comparison of the TNM 6 and 7 systems. *Journal of Thoracic Oncology*, *6*(4), 757-761.

Perlet, C., Kessler, M., Lenington, S., Sittek, H., & Reiser, M. (2000). Electrical impedance measurement of the breast: effect of hormonal changes associated with the menstrual cycle. *European radiology*, *10*(10), 1550-1554.

Pethig, R. (1984). Dielectric properties of biological materials: Biophysical and medical applications. *IEEE Transactions on Electrical Insulation*, (5), 453-474.

Pool, T. B., Cameron, I. L., Smith, N. K. R., & Sparks, R. L. (1981). Intracellular sodium and growth control: a comparison of normal and transformed cells. *The Transformed Cell*, 398.

Rangayyan, R. M., El-Faramawy, N. M., Desautels, J. L., & Alim, O. A. (1997). Measures of acutance and shape for classification of breast tumors. *IEEE Transactions on medical imaging*, *16*(6), 799-810.

Ron, K., & Foster, P. (1998). Glossary of terms. *Machine Learning* (pp. 271–274).

Schwan, H. P. (1959). Alternating current spectroscopy of biological substances. *Proceedings of the IRE*, *47*(11), 1841-1855.

Schwan, H. P. (1989). Dielectric properties of tissues-a critical review. *CRC Crit Rev Bioeng*, *17*, 25-104.

Sha, L., Ward, E. R., & Stroy, B. (2002). A review of dielectric properties of normal and malignant breast tissue. In *SoutheastCon, 2002. Proceedings IEEE* (pp. 457-462). IEEE.

Strumillo P., & Kaminski W. (2003). Radial Basis Function Neural Networks: Theory and Applicatios. *Neural Networks and Soft Computing* (pp. 107-119).

Sun, Z., Aubry, M. C., Deschamps, C., Marks, R. S., Okuno, S. H., Williams, B. A., ... & Yang, P. (2006). Histologic grade is an independent prognostic factor for survival in non–small cell lung cancer: An analysis of 5018 hospital-and 712 population-based cases. *The Journal of thoracic and cardiovascular surgery*, *131*(5), 1014-1020.

Surowiec, A. J., Stuchly, S. S., Barr, J. R., & Swarup, A. A. S. A. (1988). Dielectric properties of breast carcinoma and the surrounding tissues. *IEEE Transactions on Biomedical Engineering*, *35*(4), 257-263.

Surowiec, A. J., Stuchly, S. S., Barr, J. R., & Swarup, A. A. S. A. (1988). Dielectric properties of breast carcinoma and the surrounding tissues. *IEEE Transactions on Biomedical Engineering*, *35*(4), 257-263.

Tachibana, S. H. U. N. R. O., Aguilar, J. A., & Birzis, L. U. C. Y. (1970). Scanning the interior of living brain by impedography. *Journal of applied physiology*, *28*(4), 534-539.

Travis, W. D., & Harris, C. C. (2004). *Pathology and genetics of tumours of the lung, pleura, thymus and heart*. Feance: IARC Press, 2004.

Vacek, P. M., Geller, B. M., Weaver, D. L., & Foster, R. S. (2002). Increased mammography use and its impact on earlier breast cancer detection in Vermont, 1975–1999. *Cancer*, *94*(8), 2160-2168.

Yardimci, A. (2009). Soft computing in medicine. *Applied Soft Computing*, *9*(3), 1029-1043.

Zhao, M., Wi, H., Kamal, A. H. M., McEwan, A. L., Woo, E. J., & Oh, T. I. (2012). High density trans-admittance mammography development and preliminary phantom tests. *Biomedical engineering online*, *11*(1), 75.

Zheng, B., Zuley, M. L., Sumkin, J. H., Catullo, V. J., Abrams, G. S., Rathfon, G. Y., ... & Gur, D. (2008). Detection of breast abnormalities using a prototype resonance electrical impedance spectroscopy system: A preliminary study. *Medical physics*, *35*(7), 3041-3048.

APPENDICES
SOURCE CODES

- **Data Normalization code**

```
%M=normalizing(a,c)
X=xlsread('Norm_BreastTissue.xlsx', 'sheet2');
% X=[1 2 3
%  ];
maxr = meshgrid(max(X),[1:size(X,1)]');
% minr = meshgrid(min(X),[1:size(X,1)]');
% maxr
%  data=X + min
normdata=X./maxr;
 xlswrite('normsara.xlsx', normdata,'sheet4');
%%%%%%%%%%%%%%%%%%%%%% to excel file
% filename = 'trainset.xlsx';
% writetable(M,filename,'Sheet',1,'Range','A')
```

- **Sequential Feature Selection, Cross Validation and Naïve Bayes Algorithm Code**

```
close all
clear all
% import dataset
BreastTissue = xlsread('NBBreastTissueDataset.xlsx', 'sheet1');
% coallation coefficient
covmat = corrcoef(BreastTissue);
% plotCovarianceMatrix(covmat);
% input and output variables
X = BreastTissue(:,1:end-1);
Y = BreastTissue(:,end);
% cross validation
c2 = cvpartition(Y, 'holdout');
Xtrain = X(training(c2, 1),:);
```

```
Xtest = X(test(c2, 1),:);
Ytrain = Y(training(c2,1));
Ytest = Y(test(c2,1));
% naive bayes classifier
%  baymodel = NaiveBayes.fit(Xtrain,Ytrain,'Distribution','kernel');
%  Ypredict = predict(baymodel,Xtest);
% confusion matrix showing the breakdown of observations
%  disp(1-(sum(Ytest~=Ypredict)/length(Ytest)));
%  confusionmat(Ypredict, Ytest)
% sequential feature selection
% if PARPOOL('size') == 0
%       PARPOOL open 2
%    end
  opts = statset('display','iter', 'useparallel', 'always');
  fun = @(Xtrain, Ytrain,Xtest,Ytest)...
      sum(Ytest~=predict(NaiveBayes.fit(Xtrain,Ytrain,'Distribution','kernel'),Xtest));
   fs = sequentialfs(fun,X,Y,'cv',c2,'options',opts);
% prediction using significant variables
  Ypredict = predict(NaiveBayes.fit(Xtrain(:,fs),Ytrain,'Distribution','kernel'),Xtest(:,fs));
  disp(1-(sum(Ytest~=Ypredict)/length(Ytest)));
  confusionmat(Ypredict, Ytest)
```

- **K-fold Cross Validation and RBFN Code**

```
X=xlsread('RBFNBreastTissueDataset.xlsx', 'sheet3');
Spread=1;
K_i=0;
 basisfunction='gaussian';
 goal=0.01;
t1=zeros(6,106);
a=t1(:,1:21);
b=t1(:,22:36);
```

```
c=t1(:,37:54);
d=t1(:,55:70);
e=t1(:,71:84);
f=t1(:,85:106);
a(1,:)=1;
b(2,:)=1;
c(3,:)=1;
d(4,:)=1;
e(5,:)=1;
f(6,:)=1;
train_target=[a,b,c,d,e,f];
Y=train_target;
% Setup Division of Data for Training and Testing
net.divideFcn, trainRatio = 70/100;
net.divideFcn, testRatio = 30/100;
%[W, phi]=train_rbf(X,Y,Xc,k_i,basisfunction)
%trains a radial basis function
%X is a N_p by N_dim matrix of training data
%Y is a N_p by N_dim matrix of training data
%Xc is a N_r by N_dim matrix of rbf centres
%basisfunction may be 'gaussian' or 'polyharmonicspline'
%k_i is a prescaler for 'gaussian' rbf and function order for
%'polyharmonicspline'. Set k_i(i)=0 for constant bias
%Outputs weight vector as well as unweighted RBF outputs for
%training data.
net = newrb(X,Y,goal,Spread);
%The network is simulated for a new input.
Y = sim(net,X)
%%%%%RECOGNITION RATE OF TRAIN DATA
%target max indices
[M,I_t]=max(train_target);% row vector
```

```
%dimensions of target matrix
[u,v]=size(train_target);
train_input=X;
%actual output matrix sim_net
sim_net=sim(net,train_input);
[N,I_sim_net]=max(sim_net);% row vector
%comparison of target and actual outputs
result = I_t==I_sim_net;% row vector
%sum of all elements,1s, to know how many corrects
corrects=sum(result);
%recognition rate,
s=double(corrects*100/v); %let recognition rate be w
fprintf('train recognition rate is %d\n',s);
```

- **K-fold Cross Validation and BPNN Code**

```
close all
clear all
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%Breast tissue
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%classification
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%using BPNN...Nb of
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%classes=6.
train_input=xlsread('RBFNBreastTissueDataset.xlsx', 'sheet3');
t1=zeros(6,106);
a=t1(:,1:21);
b=t1(:,22:36);
c=t1(:,37:54);
d=t1(:,55:70);
```

```
e=t1(:,71:84);
f=t1(:,85:106);
a(1,:)=1;
b(2,:)=1;
c(3,:)=1;
d(4,:)=1;
e(5,:)=1;
f(6,:)=1;
train_target=[a,b,c,d,e,f];
% CREATING AND INITIATING THE NETWORK
net = newff(minmax(train_input),[100 6],{'logsig','logsig'},'traingdx');
% Setup Division of Data for Training and Testing
net.divideFcn, trainRatio = 70/100;
net.divideFcn, testRatio = 30/100;
% TRAINING THE NETWORK
net.trainParam.goal = 0.01; % Sum-squared error goal.
net.trainParam.lr = 0.2;  % Learning Rate.
net.trainParam.epochs =3000;% Maximum number of epochs to train.
net.trainParam.mc = 0.7 % Momentum Factor.
[net,tr] = train(net,train_input,train_target);
ActualOutput=sim(net, train_input)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%
%%%%%RECOGNITION RATE OF TRAIN DATA
%target max indices
[M,I_t]=max(train_target);% row vector
%dimensions of target matrix
[u,v]=size(train_target);
%actual output matrix sim_net
sim_net=sim(net,train_input);
[N,I_sim_net]=max(sim_net);% row vector
```

```
%comparison of target and actual outputs
result = I_t==I_sim_net;% row vector
%sum of all elements,1s, to know how many corrects
corrects=sum(result);
%recognition rate,
s=double(corrects*100/v); %let recognition rate be w
fprintf('train recognition rate is %d\n',s);
```