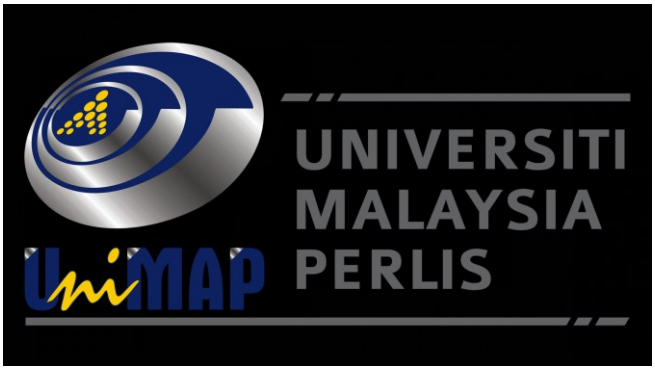


AN INTELLIGENT EARLY BREAST CANCER DETECTION USING STATISTICAL FEATURE GENERATION

by

VIJAYASARVESWARI A/P VEERAPERUMAL
(1540211677)

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

School of Computer and Communication Engineering
UNIVERSITI MALAYSIA PERLIS

2019

ACKNOWLEDGMENT

First of all thanks to God because of His blessings I could complete my research and prepare reports properly in the required timeframe. I would like to show my appreciation to the Universiti Malaysia Perlis (UniMAP) for giving me the opportunity and facilities to undergo my postgraduate. I would like to thank to the Dean of School of Computer and Communication Engineering, Associate Professor Dr. Azremi Abdullah Al-Hadi for providing excellent facilities to do the research smoothly. Not forget, I would like my sincere gratitude to my supervisor, Dr. Muzammil Jusoh for his valuable guidance and supervision for the research to shape the present work as below. Besides that, my deepest thanks to all my co-supervisors, Dr Thennarasan Sabapathy, Dr. Rafikha Aliana A. Raof and Prof. Dr. Sabira Khatun for giving full cooperation and persistent help. I would like to thank Dr. Allan Melvin Andrew for his guidance throughout the research period. All the guidance and assistance given by the supervisors are very much appreciated. In addition, do not forget my parents and my friends who give a lot of motivation and encouragement to me over the years. I would like to thank to Malaysia Education Ministry for providing financial support as fellowship. Financial support also is provided by the FRGS Grant 9003-00418.

Finally, once again, I would like to thank everyone involved either directly or indirectly throughout my research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ART	Adaptive resonance theory
ASR	Age-standardized rate
BR	Bayesian regularization
CFBNN	Cascade forward backpropagation neural network
CNN	Convolutional neural network
DAS	Delay and sum
DCT	Digital cosine transform
DT	Decision tree
DWT	Discrete wavelet transform
FCC	Federal Communications Commission
FFT	Fast Fourier Transform
GNN	Gray neural network
GUI	Graphical user interface
IDFT	Inverse discrete Fourier transform
MLP	Multilayer perceptron
MRI	Magnetic resonance imaging
MSE	Mean square error
MT	Microwave tomography
NB	Naïve bayes
NBRF	Narrowband radio frequency
PCA	Principal component analysis
PNN	Probabilistic Neural Network
PSO	Particle swarm optimization
RBF	Radial basis function network
RNN	Recurrent neural network
SMO	Sequential minimal optimization
SS	Spread spectrum
SVM	Support vector machine
TCL	Transmission co-reflection loss
UWB	Ultra-wideband
VNA	Vector network analyser

LIST OF SYMBOLS

d_{Tx}	Distance between transmitting antenna and breast model
d_{Rx}	Distance between receiving antenna and breast model
Tx	Transmitter antenna
Rx	Receiver antenna
S_{21}	Transmission co-efficient
μ	Mean

Pengesanan Kanser Payudara Awal Pintar Menggunakan Generasi Ciri Statistik

ABSTRAK

Kanser payudara adalah salah satu punca utama kematian wanita di seluruh dunia. Tumor payudara adalah peringkat awal kanser yang terletak di sel-sel payudara manusia. Pengesanan kanser payudara pada peringkat awal dapat meningkatkan peluang untuk diagnosis awal. Oleh sebab itu, ia adalah sangat penting untuk mencadangkan klasifikasi pintar untuk mengesan kanser payudara pada peringkat awal. Tesis ini mencadangkan penyelidikan awal untuk mengesan saiz dan lokasi kanser payudara pada peringkat awal. Isyarat UWB dihantar oleh antena dari satu sisi dan diterima dari sisi yang lain, dikawal oleh PC. Isyarat domain masa UWB ini ditukar menjadi isyarat domain frekuensi menggunakan "*Fast Fourier Transform*". Kedua-dua isyarat domain masa dan frekuensi ditukar kepada nilai digital. Untuk klasifikasi saiz, data akan melalui kaedah penjana ciri statistik yang dicadangkan. Pada mulanya, data tersebut telah melalui Analisis Principal Component dan kemudian, dinormalisasikan menggunakan sepuluh kaedah normalisasi data yang berbeza. Dimensi data dikurangkan. Daripada sepuluh kaedah normalisasi data, hanya lima dipilih secara statistik. Sepuluh ciri, gabungan masa (linear dan bukan linear) dan frekuensi (linear) ciri diekstrak daripada setiap set data yang dinormalisasi. Daripada 50 ciri yang diekstrak, sepuluh ciri dipilih berdasarkan ujian statistik. Ciri-ciri ini disatu menggunakan teknik penyatuan ciri. 10-HybridFeature, 8-HybridFeature dan 3-HybridFeature set data dibangunkan dan kemudian diuji dengan tiga klasifikasi penyeliaan. Di antara ketiga-tiga set data ini, set data 8-HybridFeature berfungsi lebih baik untuk ketiga-tiga klasifikasi. Eksperimen menunjukkan klasifikasi "*Support Vector Machine"* (94.07%) lebih baik berbanding dengan "*Naïve Bayes"* (91.98%) dan Rangkaian Neural Probabilistik (81.64%) dengan menggunakan set data 8-HybridFeature dari segi ketepatan. Kaedah yang dicadangkan lebih baik daripada kerja sebelumnya dengan meningkatkan 11.8% ketepatan. Untuk klasifikasi lokasi dari segi koordinat x, y, dan z, data dinormalisasikan menggunakan normalisasi binari dan diklasifikasikan menggunakan klasifikasi tanpa penyeliaan (k-means clustering). K-means clustering mempamerkan ketepatan purata 80.49%. Berdasarkan hasil klasifikasi, ciri dan klasifikasi terbaik dipilih untuk algoritma Pengesanan Kanser Payudara (BCD) yang dicadangkan. Algoritma BCD yang dicadangkan terdiri daripada dua bahagian. Bahagian pertama adalah untuk pengesanan saiz yang menggunakan klasifikasi penyeliaan dan bahagian kedua adalah untuk pengesanan lokasi menggunakan klasifikasi tanpa penyeliaan. Akhirnya, saiz dan lokasi yang dikesan digambarkan dalam persekitaran 2D dan 3D.

An Intelligent Early Breast Cancer Detection Using Statistical Feature Generation

ABSTRACT

Breast cancer is one of the main causes of women death worldwide. Breast tumor is an early stage of cancer that locates in cells of a human breast. Early breast cancer detection greatly increases the chances for early diagnosis. Towards this, it is very crucial to propose an intelligent classifier to detect breast cancer in the early stage. This thesis proposes a preliminary research to detect the size and location of the breast cancer in the early stage. UWB signals are transmitted by the antenna from one side of breast phantom and received from other side, controls by PC. These UWB time domain signals are converted into frequency domain signals using Fast Fourier Transform. Both time and frequency domain signals are converted into digital values. For size classification, data is analysed by using the proposed statistical feature generator method. Initially, Principal Component Analysis is performed on the data and then, is normalized into ten different data normalization methods. The data dimension is reduced by reducing the principal components. Out of ten data normalization method, only five are chosen statistically. Ten features, combination of time (linear and non-linear) and frequency (linear) features are extracted from each of the five normalized dataset. Out of 50 extracted features, ten features are selected based on the statistical test. These features are fused together using feature fusion techniques. 10-HybridFeature, 8-HybridFeature and 3-HybridFeature datasets are developed using the proposed features and then are tested with three different supervised classifiers for size classification. Among these three datasets, 8-HybridFeature dataset performs better for all three classifiers. Experiment shows Support Vector Machine_RBF (94.07%) classifier performs better compared to Naïve Bayes (91.98%) and Probabilistic Neural Network (81.64%) classifiers by using 8-HybridFeature dataset in terms of accuracy. The proposed method performs better than previous work by improving 11.8% of accuracy. For location classification in terms of x, y, and z coordinates, the data is normalized using binary normalization and classified using unsupervised classifier (k-means clustering). The k-means clustering classifier exhibits average accuracy of 80.49%. Based on the classification result, the best feature and best classifier are selected to proposed Breast Cancer Detection (BCD) algorithm. The proposed BCD algorithm consists of two parts. The first part is for size detection using supervised classifier and second part is for location detection using unsupervised classifier. Finally, the detected size and location are visualized in 2D and 3D environment.

CHAPTER 1 : INTRODUCTION

1.1 Overview

Breast cancer is one of the main reasons for the women's death worldwide. Breast tumour is an early stage of cancer that locates in human breast cells. Breast consists of skin, fatty tissues, glandular tissue, lobules and milk duct (Martini, 2016; Hipwell, Vavourakis, Han, Mertzanidou, Eiben, & Hawkes, 2016). In the human body, cells usually reproduce and replace the dead cells. When cells reproduce rapidly without control it might cause the abnormal growth. This causes the development of cancerous cells in breast tissue. Breast with a lump at the beginning is usually known as a benign tumor which is not harmful. However, if it grows after certain duration, it may press surrounding organ and causes pain. This may lead to the development of malignant tumors, which is cancerous and dangerous. Doctor needs to perform a biopsy to check the seriousness of the cancer (National Breast Cancer Foundation Inc; Smith, Andrews, Brooks, Fedewa, Manassaram-Baptiste, Saslow, & Wender, 2017). National Cancer Registry (NCR) states age-standardized rate (ASR) is 47.3 per 100 000 populations in the year 2003-2005 while in year 2012, ASR indicates 38.7 per 100 000 population (Lee, Mariapun, Rajaram, Teo, & Yip, 2017; Yip, Pathy, & Teo, 2014). This survey shows breast cancer cases in Malaysia is high and is recorded as second causes of the woman's death. The breast cancer risk rating increases as it has no symptom in the early stage where normally it is traced based on changes in the breast or in the nipple that can only be discovered in the late stage (Ozaki, Leppold, Tsubokura, Tanimoto, Saji, Kato, Kami, Tsukada & Ohira, 2016; Lim, Potrata, Simonella, Ng, Aw, Dali, Hartman, Mazlan & Taib, 2015). However, the numbers of surviving increase if cancer can be

detected in the early stage (Miller, Siegel, Lin, Mariotto, Kramer, Rowland, Stein, Alteri & Jemal, 2016; Saadatmand, Bretveld, Siesling, & Tilanus-Linthorst, 2015; Joy, Penhoet, & Petitti, 2005).

Researchers found various types of technique to detect breast cancer. The traditional techniques that are currently used in the clinics and hospitals are X-ray mammogram, ultrasound and magnetic resonance imaging (MRI) scans. These types of techniques can be used if some symptoms of breast cancer are identified, but it is too late already.

Besides these traditional methods, microwave based Ultra-wideband (UWB) imaging has been proposed by many researchers as a new and healthier-safe technology (Song, Li, Coates, & Men, 2017; Susila & Fathima, 2017; Rahman, Islam, Singh, Kibria & Akhtaruzzaman, 2016; Shahzad, O'Halloran, Jones & Glavin 2016; Baran, Kurrant, Zakaria, Fear & LoVetri, 2014; Ünal, Türetken, Sürmeli, & Canbay, 2011). UWB is a type of technology that uses radio energy to transmit the information with an advantages of low-power, high bandwidth and secure technology. It can transmit huge data within a very short time. The bandwidth of UWB for unlicensed use is from 3.1 GHz to 10.6 GHz according to the Federal Communications Commission (FCC) (Sarjoghian, Alfadhl & Chen, 2016; Kshetrimayum, 2009). Most of the UWB based early breast cancer detection techniques are still under research enhancement stage in the biomedical area. UWB uses dielectric properties (permittivity and conductivity) to distinguish healthy and unhealthy breast and stages of breast cancer (Martellosio, Pasian, Bozzi, Perregrini, Mazzanti, Svelto, Summers, Renne & Bellomi, 2015; Al-Fraihat, Al-Mufti, Hashim & Adam 2014). Basically, there are two approaches in microwave based UWB

imaging: radar based imaging or confocal imaging and microwave tomography. Radar based imaging or confocal imaging uses scattered signals (change due to differences in dielectric properties) to image the breast tissue, whereas in microwave tomography, the electrical properties are constructed by calculating the nonlinear and ill-posed inverse scattering signal (Song, et. al. 2017; Susila, et. al. 2017; Rahman, et. al. 2016). The scattered signals at the receiver contain the characteristic of tumor.

1.2 Research Motivation

Breast cancer cases are increasing year by year. Two types of breast cancer detection technologies are available i.e.: body image based technology and microwave imaging based technology. Body image based technology includes mammograms, magnetic resonance imaging (MRI) and ultrasound. These types of technologies are widely used. These technologies are not 100% accurate, especially when deal with dense breasts. Therefore, the incidence of false negative and false positive can occur during the screening process. A false negative occurs when the system cannot detect cancer, even though cancer is present and false positive occurs when the system detect cancer, even though cancer is not present (American Cancer Society). Moreover, only an expert operator (radiologist) can handle these technologies to interpret the images. Besides, the detected cancer cannot distinguish between benign and malignant with the available equipment except through biopsy.

A safe, easy to handle and non-invasive technology is needed which resulted in the development of microwave based technology (Hang, Sim & Zakaria, 2017; Kwon & Lee, 2016). This technology consists of two approaches: microwave tomography (MT)

and radar based technology (Susila, et.al., 2017; Rahman, et. al., 2016; Shahzad, et. al., 2016). MT has its own drawbacks where it needs to be calibrated separately for dissimilar materials. The result is not trustworthy because a small change in measurement causes a large impact on the result. This causes MT sensitive to noise in the measurement data (Subotic, Pjevalica & Palfi, 2017; Rahiman, Kiat, Jack & Rahim, 2015). Mostly MT uses one antenna to transmit and multiple antennas to receive the signals. Radar based technology is able to produce effective results even in noisy places and usually use less number of antennas. This shortcoming causes a radar based technology method is more low-cost, efficient and better potential to detect a breast cancer compared to microwave tomography. There are many ways to simulate or process the receive UWB signals. Most of the researchers use vector network analyser (VNA) or real time oscilloscope to view the changes of the UWB signal. The antenna is attached to the VNA or a real time oscilloscope and the UWB signals are obtained which contain the breast cancer information (Salleh, Othman, Ali, Sulaiman, Misran & Aziz, 2015; Rahman, et.al., 2016; Preece, Craddock, Shere, Jones & Winton, 2016). However, the simulation is done manually and causes more time (Mehdy, Ng, Shair, Saleh & Gomes, 2017).

Automated classifier is really needed in order to upgrade the simulation process in terms of time consumption and efficiency. Therefore, researchers propose intelligent classifiers using machine learning for data classification and pattern recognition (Mehdy, et.al., 2017; Arif, Alam & Hussain, 2015; Wang, Wang H, Chen & Liu, 2015). Machine learning is known as a smart data analysis that is more pervasive for its usage in modelling, simulating and optimizing. By using machine learning, it offers high

processing speed and better efficiency for a classification model. Therefore, it is important to explore machine learning to obtain an efficient machine learning model.

1.3 Problem Statement

The capability of the machine learning model depends on the features fed into the machine learning model for the training purpose. The features are selected based on the different feature selection methods proposed by various researchers in breast cancer detection application. Researchers normally obtain features either by extracting, selecting or normalizing the features (Song, Li, Coates & Men, 2017; Zhao, Wang & Cui, 2017; Shirazi, Chabok, & Mohammadi, 2018; Reza, Khatun, Jamlos, Fakir & Morshed, 2015). From the previous research, the conventional feature selection methods performed by the researchers in breast cancer detection application are basically a single stage feature selection method. However, such method contributes poor stability and higher misclassification rate due to the deficiency of deep analysis of data (Liang, Ma, Yang, Wang & Ma, 2018). Stability means to maintain the performance across the different scenarios. The exploration and exploitation of the data are insufficient during the feature selection as the features are reduced in the beginning stage which cause the selected features are might be redundant features or some useful features are lost. Thus, optimal number of features are unable to identify using the conventional feature selection method (Suji & Rajagopalan, 2016). Exploration is the identification of features through multi-stage approach while exploitation is done by adding some additional related information to the previous best solution. Thus, the exploration of feature selection methods to obtain an efficient machine learning model is the motivation of this study.

Traditionally, a set of strong features is selected after data analysis to avoid the development of overfitting machine learning model. Based on previous studies in breast cancer detection application, some of the researchers select the subset of feature depending on the machine learning score and some of them select during the machine learning model construction (Chen, Yang, Wang, Wang, Liu & Liu, 2012; Aalei, Shahraki, Rowhanimanesh & Eslami, 2016; Celaya, Ortiz, Martinez, Solis, Castaneda, Garza, Martinez & Ortiz, 2016). But, the selected features may not be applicable for different types of machine learning model since the feature selection is highly dependent on the machine learning and computational inefficiency (Hira & Gillies, 2015). It causes the high possibilities to build machine learning models with the high misclassification rate if the same feature used for different types of machine learning Thus, determining features by using the intrinsic properties of the data with maximum importance and minimum similarity is another motivation for this study. This can be done by ranking up the feature in order to measure the importance of the features independently. At the same time, different selection methods are added after the features ranking up process to investigate the features in different angle. From this, the complexity of the model can be reduced and optimization problem can be solved (Geng, Liu, Qin& Li, 2007).

In the breast cancer detection studies, developing a complete framework for detection is one of the important issues. Only some of the researchers only manage to develop a complete framework (from the data sample collection to visualization) to detect the breast cancer's parameters in their studies. The parameters of breast cancer that usually detected by researchers are existence, location, size and type. For examples, Shirazi (2018), Saini (2015), and Huang (2017) build framework to identify the

existence either presence or absence of the breast cancer. While Santorelli (2014) builds a framework to detect the existence and location of the breast cancer. Reza (2015) and Conceição (2010) build framework only to identify the size of the breast cancer and Yi (2017) and Chaurasia (2018) and Vandenberghe (2017) build framework only to research on investigating the type (benign and malignant) of the cancer. But, the researchers mostly be considered in the development of a framework that is able to detect breast cancer's parameter separately, one parameter each time. This causes the detection of breast cancer process to be repetitive in detecting different breast cancer's parameter, which is inconvenient and high computational time. Thus, a complete framework of breast cancer detection (combination of more than one detection parameter in the early stage) should be developed.

1.4 Research Objectives

1. To design a new statistical feature generator method for effective feature selection to improve classification accuracy for early breast cancer's size detection

2. To identify a new hybrid feature for early breast cancer detection through proposed feature generator method.

3. To develop a complete framework for early breast cancer detection to detect location and size using proposed hybrid feature and classifier.

1.5 Significance of the Study

The finding of this study redounds to the benefit of society, considering that early breast cancer detection is very crucial for better treatment today. The outcomes of the study to be considered are the development of a new feature selection method, investigation of a new set of hybrid feature and development of complete breast cancer framework. The greater demand for breast cancer detection using machine learning justifies the need for more effective and efficient feature selection method, as the features are the key for better classification in the machine learning. Thus, the researchers that apply the recommended feature selection approach and the selected features that are derived from the results of this study are able to produce better classification result with low false negative and false positive. The features are selected based on the own intrinsic properties which is able to produce a better version of the feature and reduce the possibilities to select redundancy and irrelevant features. Thus, this study is able to contribute the improvement for the available breast cancer detection system. For the researchers, this study helps them to uncover some of the critical areas that other researchers are unable to explore.

1.6 Research Scope

This research work is the beginning stage for breast cancer detection in the early stage using the UWB system. Heterogeneous breast phantom is only considered in this study because it is more practical compared to homogeneous breast phantom. One breast phantom size of 75mm, 60mm and 1.9mm in width, height and thickness respectively is considered. Five different sizes of tumor (2mm, 3mm, 4mm, 5mm and

6mm) are placed in different locations of x, y and z. The developed breast phantom is placed in between of two antennas. UWB transceiver is used to transmit scattered UWB signals through breast phantom. The forward scattered signal is only captured and analyzed in both time and frequency domain for further process.

Several data normalization methods are used to normalize the data sample. Ten different types of data normalization methods are used in this study. Only five data normalization methods are selected based on the available method for breast cancer application which are z-score, min-max, linear scaling, decimal scaling and mean and standard deviation normalization method. Whereas the remaining five data normalization methods are newly introduced to this application. Out of ten, only five methods are chosen based on statistical test for further analysis.

There is time, frequency and non-linear features that can be extracted from the data sample. In this study, only ten different features combination of time, frequency and non-linear features are extracted from the five datasets. Therefore, 50 features are extracted. But, 10 significant features are selected for further analysis.

Three datasets are evaluated using supervised classifiers. Among available classifiers, which are used for breast cancer detection, three supervised classifiers are used. The selected classifiers are Support Vector Machine (SVM), Probabilistic Neural Network (PNN) and Naïve Bayes (NB) because of its usage and performance in the breast cancer detection application. The performance of classifiers is evaluated and the best classifier is selected to be implemented in the proposed breast cancer detection algorithm.

This work is a preliminary work for breast cancer detection. Thus, the scope of this research is limited to select best features and classifiers to detect the tumor's size and location, regardless of the other parameters and other influencing factors to be considered.

1.7 Research Study Module

Figure 1.1 shows the study and investigation directions taken into account to achieve objectives of the research. It describes the overall techniques (both hardware and software) that are implemented for breast cancer studies. The scope of this study is limited only by exploring some of the techniques (as in blue boxes).

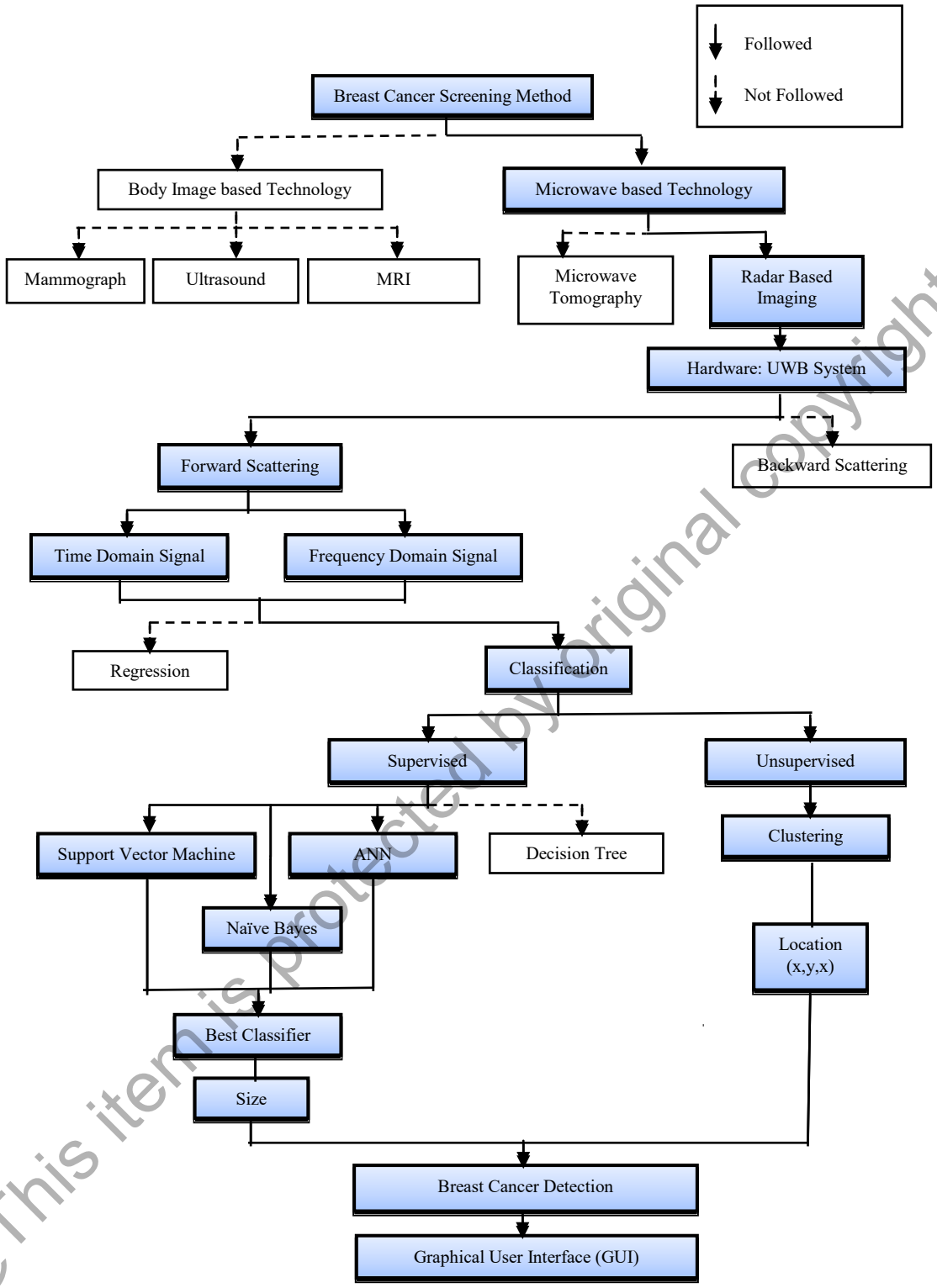


Figure 1.1 Research Study Module

1.8 Organization of the Thesis

This thesis contains six chapters. This current chapter gives an introduction of the studies. The problem statements are listed based on the limitation of previous studies and the objectives of the studies are stated based on the problem statement. The research scope and contribution are briefly described in this chapter.

Chapter 2 explains about the previous studies which related to the proposed research work. It presents the literature reviews on current breast cancer detection techniques, developed classifier and the achievement as well as the limitation of their studies.

Chapter 3 presents the data collection that is taken for analysis. This chapter also contains details about the breast phantom development, UWB antenna description, the experimental set-up and the pre-processing of collecting the data sample.

Chapter 4 describes of the processes to develop feature generator method for breast cancer detection and also the proposed a new set of hybrid significant features. The development and implementation of breast cancer detection algorithm are explained.

Chapter 5 elaborates the results of the work done based on chapter 4. The performance of the hybrid feature using supervised classifier for size detection is shown. The performance of location detection is discussed here.

Chapter 6 is the conclusion of the overall studies. The contribution of this study and the future work are explained

CHAPTER 2 : LITERATURE REVIEW

2.1 Overview

This chapter provides the review of breast cancer related works. Firstly, the detection technologies such as traditional (body image) based technology and microwave based technology which has two approaches i.e.: microwave tomography and radar based imaging technology are explained. Along with it, the signal processing methods such as focusing algorithms and machine learning for breast cancer detection are explained briefly and followed by the conclusion of this chapter.

2.2 Breast Cancer

Breast consists of fat, tissue, nerves and blood vessels (vein and arteries) as shown in Figure 2.1 and its physical structure is always changed throughout the lifetime. Here to mention, both man and woman have the breast tissue which the chance to develop breast cancer is high. However, breast cancer is more synonym to women as man does not involve with breast growth because of high testosterone and low estrogens levels. Thus, breast cancer can be stated as one of the leading diseases in the world for women compared to man (Susan G.Komen, 2015).

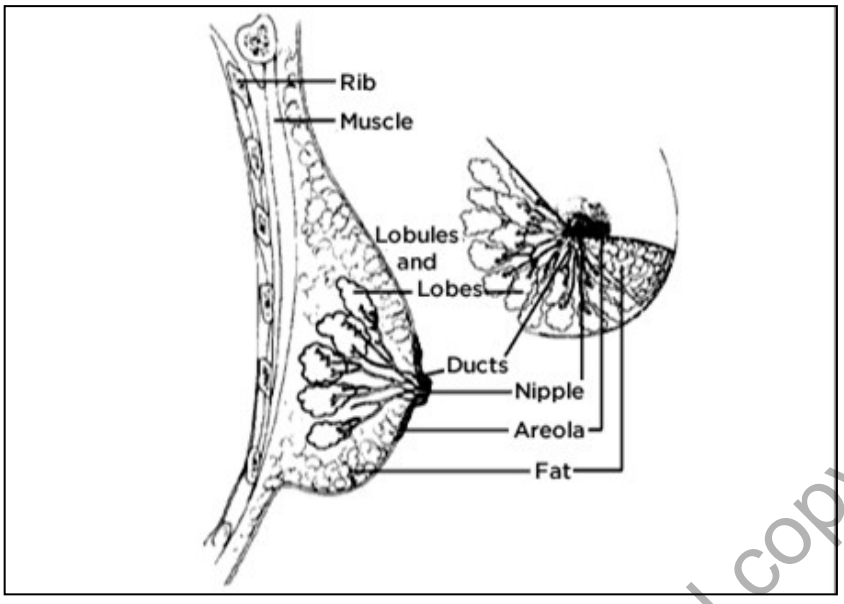


Figure 2.1 Breast Structure

Breast cancer can be grouped into two groups which are non-invasive and invasive breast cancer. A non-invasive cancer is usually found in duct and it is known as an early stage of cancer. Sometimes it will not grow and does not affect the life. In other hand, invasive breast cancer is cancerous where can be found in the cell throughout the breast duct or lobular and has high possibility of spreading to other part of the body. (Cruz-Roa, Gilmore, Basavanhally, Feldman, Ganesan, Shih, Tomaszewski, González, & Madabhushi, 2017; Hang, Sim & Zakaria, 2017; MedImmune, 2017; Sharma, Dave, Sanadya, Sharma & Sharma, 2010). Breast cancer cell develops because of some risk factors such as age, number of children, menopause and first childbirth, family background, birth control pills, and many more (Wellings, Vassiliades & Abdalla, 2016). However, the exact cause of breast cancer is still carrying a big question mark (American Cancer Society).

2.3 Commercial Breast Cancer Detection Tool and Its Limitation

The available breast cancer detection tool detects the breast cancer for further treatment to reduce the death risk (Hang, et.al., 2017; Fisher, Wilkinson & Valencia, 2016; Brown, 2016; Tabar, Yen, Vitak, Chen, Smith & Duffy, 2003) such as body imaging based technology and microwave imaging based technology (Kwon, et.al., 2016; Park, Kim, Jeong, Lee, Kim, Kim, Cha & Yoo, 2013).

2.3.1 Body Image based Technology

Body image based technology (mammography, magnetic resonance imaging (MRI) or ultrasound) obtains the image of the breast structure to examine and investigate the abnormality of the breast by the radiologist. Mammography uses low energy X-ray to scan the breast in order to get the breast picture (also refers as mammogram) as shown in Figure 2.2 (Kwon & Lee, 2016). Figure 2.3 shows the mammogram of the different level of breast density (American Cancer Society).

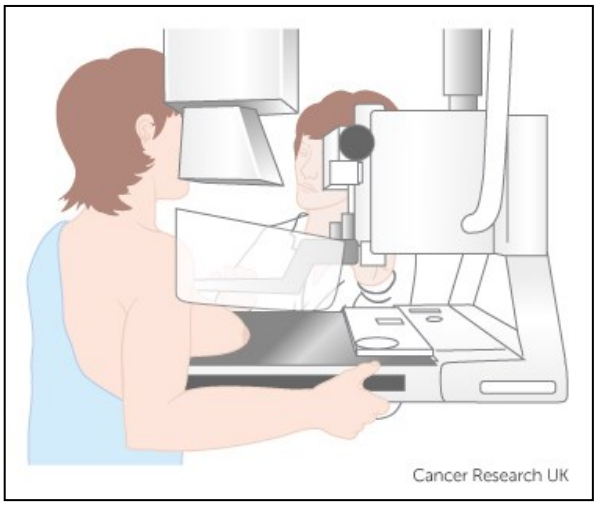
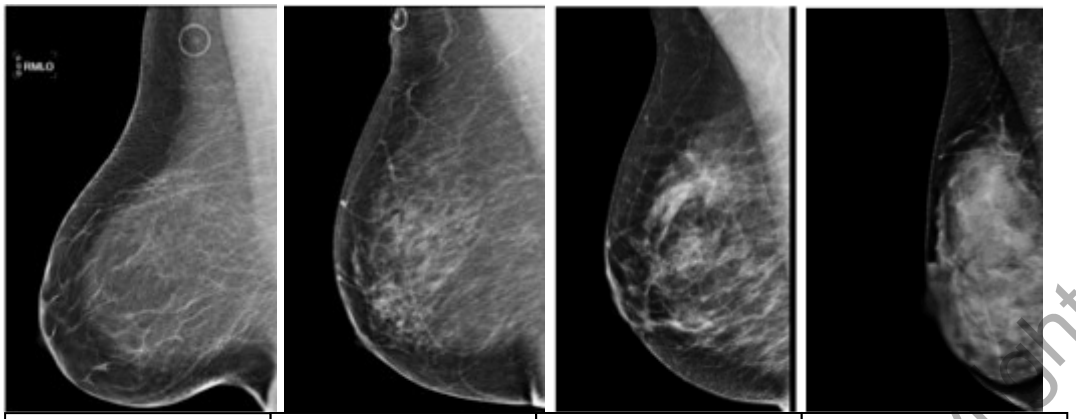


Figure 2.2 Mammography (Kwon, et.al., 2016)



Level 1: Fatty Tissue	Level 2: Average Density	Level 3: Heterogeneously Dense	Level 4: Extremely Dense

Figure 2.3 Mammogram of Breast based Dense (American Cancer Society)

Different from the mammography, MRI uses strong magnet fields and radio waves in order to get the breast image (Kwon, et.al., 2016). In addition, it can be considered as a better detection method compared to mammography and ultrasound because it can image breast in details especially for women with dense breast and can detect the breast cancer in the early stage (Chen, Huang, Shen, Liu & Xu, 2017). Figure 2.4 shows an MRI system available at the clinic / hospital.

Figure 2.4 Magnetic Resonance Imaging (MRI)

Same as MRI, ultrasound can be used after some lump is detected in the breast. Ultrasound uses sound waves with high-frequency to check whether the lump is in solid

(cancerous) or liquid (non-cancerous) form by imaging it as shown in Figure 2.5 (Kwon, et.al., 2016). It is a low-cost method and more suitable for younger women (lower than 35 years old) (Köşüş, Köşüş, Duran, Simavli & Turhan, 2010), and premenopausal women. It is because the younger woman breast structure is denser and more solid (Lee, Mariapun, Rajaram, Teo, & Yip 2017; Eugênio, Souza, Chojniak, Bitencourt, Graziano & Marques, 2017).

Figure 2.5 Ultrasound

Advantages and disadvantages of the body imaging based technology are briefly discussed in Table 2.1 (Kwon, et.al., 2016; Hang, et.al., 2017; Sylvia, et.al., 2011; Daud, Othman & Othman, 2014; Jalalian, Mashohor, Mahmud, Karasfi, Saripan & Ramli, 2017; Zhao, Zou, Geng & Zheng, 2015; Song, et.al., 2017; Kwong & Yucel, 2003)

Table 2.1 Advantages and Disadvantages for Mammography, MRI and Ultrasound

Method	Advantages	Disadvantages
Mammography	1. Early detection. 2. Regular check-up.	1. False negative and false positive. 2. Not 100% accurate. 3. Not suitable for young women and dense breast 4. Radiation from x-ray (side-effect). 5. Painful breast compression
MRI	1. Details picture of breast. 2. Show information which others could not show it. 3. Painless 4. Image in any direction and angle.	1. Magnetic field attracts metal objects. Need to remove it. 2. Side-effect (contrasting agent) 3. Expensive compared to mammography and ultrasound 4. Could not find all cancers. 5. High time consuming, 6. Need very experience radiologist to classify the breast cancer.
Ultrasound	1. Suitable for younger women. 2. Suitable for denser breast. 3. Give clear vision of the image than mammography. 4. Quick and painless.	1. Causes the allergic due to the gel applies on the skin. 2. Concentrate to detect benign tumor only. 3. Not accurate (malignant tumor mostly cannot be detected). 4. Need an experienced and trained doctor to operate.

2.3.2 Microwave Imaging based Technology

Microwave based technology is a potential technology which can replace the invasive and expensive screening traditional technology (mammography, MRI and ultrasound). Furthermore, this technology is safe, robust, ionizing radiation free and causes lesser physical harm to users (Kwon, et.al., 2016; Hang, et.al., 2017). There are three types of breast imaging methods in microwave imaging which are passive, hybrid and active as shown in Figure 2.6. The passive method classifies the detected tumor by measuring the differences of temperature between healthy and unhealthy breast and hybrid method uses more energy to identify the tumor and microwave to image the breast. Meanwhile, microwave signals are transmitted and received rapidly and illuminating the breast imaging using the microwave in active method (Daud, et.al., 2014).

Microwave based technology uses two approaches which are microwave tomography and radar based imaging (Song, et.al., 2017; Susila, et.al., 2017; Rahman, et.al., 2016; Shahzad, et.al., 2016; Baran, et.al., 2014; Unal, et.al., 2011). In both approaches, use the received UWB signals to classify breast cancer according to the dielectric properties (Martellosio, Pasian, Bozzi, Perregrini, Mazzanti, Svelto, Summers, Renne & Bellomi, 2015; Al-Fraihat, et.al., 2014; Noghanian, 2012). Basically, the received UWB signals is obtained either in time domain or frequency domain but frequency domain UWB signal is better because it has better signal-to-noise ratio compared to time domain UWB signal (Rahman, et.al., 2016).

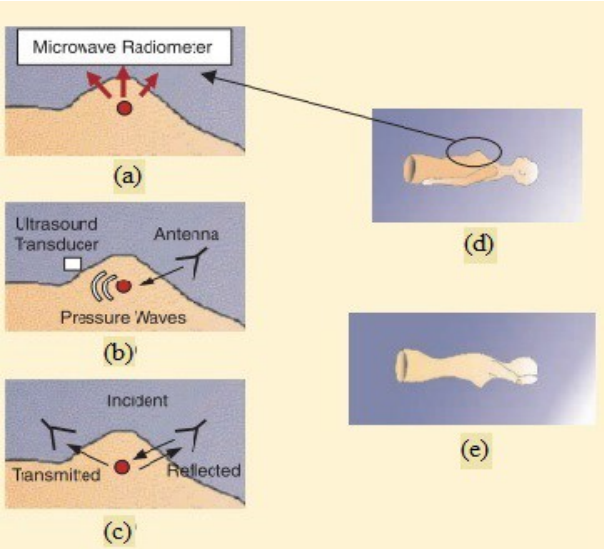


Figure 2.6 Microwave Breast Imaging Method (Daud, et.al., 2014)

UWB is a wireless technology and different from the narrowband radio frequency (NBRF) and spread spectrum technologies (SS) as shown in Figure 2.7 and Figure 2.8 (Navaz & Nawaz, 2014). US Federal Communication Commission (FCC) approves an unlicensed spectrum (3.1- 10.6GHz) for UWB (Kshetrimayum, 2009; Feng, Che & Xue, 2015).

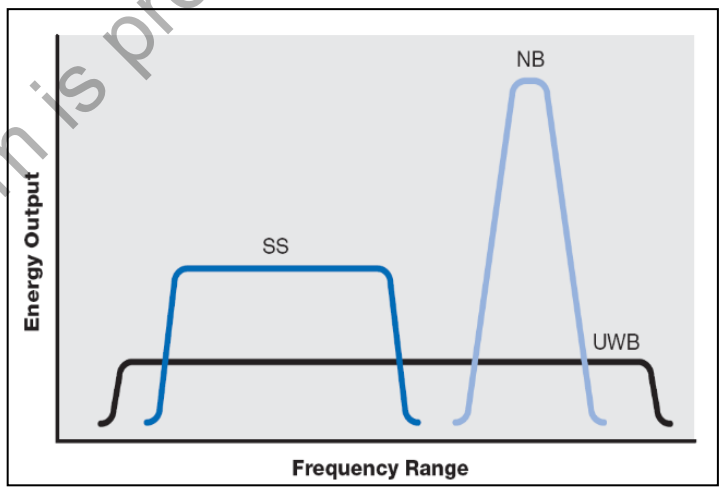


Figure 2.7 Frequency Range for SS, UWB and NB (Navaz, et.al., 2014)

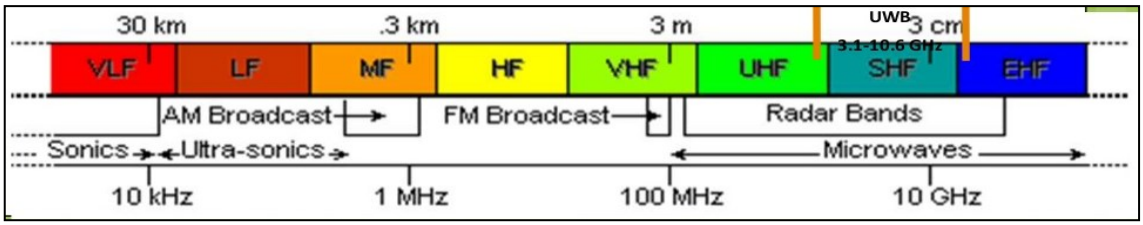


Figure 2.8 UWB Frequency Range

The advantages and disadvantages of the UWB are as shown in Table 2.2 (Dong, Li & Deng, 2017; Banke & Tiwari, 2016; Al-Samman, Rahman, Nasir, Jamaluddin, Khalily & Kamarudin, 2015; Kshetrimayum, 2009).

Table 2.2 Advantages and Disadvantages of UWB

Advantages	Disadvantages
1. Operate in wide frequency range (3.1-10.6GHz) as shown in Figure 2.8. Can share frequency spectrum. 2. High bandwidth (more than 500MHz). Sharing the bandwidth resources. 3. Can send large amount data at certain times. 4. Low cost. 5. Lower power consumption 6. High speed. 7. Using low power. 8. Convenience and flexibility. **9.** Secure. Could not interfere with others.	1. Short range communication. 2. UWB signal can interfere other license spectrum as it can be used in noisy place. 3. Complex circuitry for multipath energy 4. High speed ADCs for signal processing.

UWB technology is implemented and proposed for various types of applications such as data transfer in the home and office, video and voice application in short range,

military communication, replacing global positioning system (GPS) for identifying the location in short range, medical application, but some still under research (Sarjoghian, et.al., 2017; Dong, et.al., 2017: Kohno, Iinatti & Sameshima, 2016). The UWB is mostly used for indoor applications because of the UWB ability to immune to the multipath fading (Patel & Modi, 2015).

2.3.2.1 Microwave Tomography

Microwave tomography (MT) measures the scattered signal and applies the inverse non-linear algorithm to reconstruct the breast image (Susila, et.al., 2017; Shahzad, et.al., 2016; Rahman, et.al., 2016). MT uses one antenna to transmit UWB signal and the transmitted UWB signal penetrates the breast and reflects. Here, several antennas are used to receive the reflected UWB signals as shown in Figure 2.9 (Persson & Fhager, 2014; Rahiman, et.al., 2015).

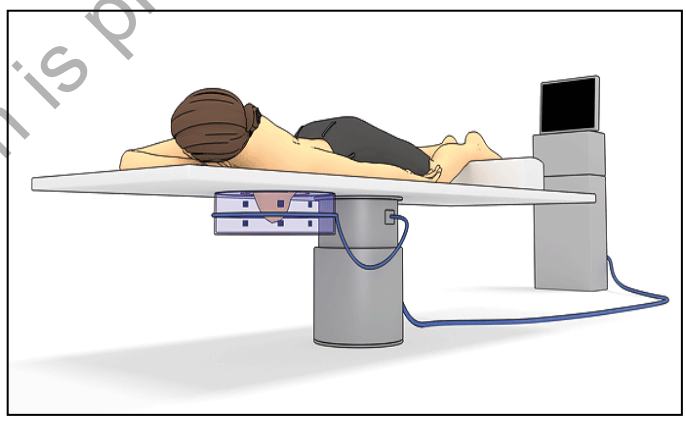


Figure 2.9 Microwave Tomography based Breast Cancer Detection (Persson, et.al., 2014)

2.3.2.2 Radar (UWB) based Imaging technology

Radar based imaging technology uses one or two antennas to transmit a UWB signal while all scattered signal (forward or backward) is received using another antenna (Henrikson, Klemm, Gibbins, Leendertz, Horseman, Preece, Benjamin & Craddock, 2011). It is considered more comfortable technology to detect breast cancer because of less number of antennas involved during the signal transmission and at the same time it can detect breast more accurately even in noisy places compared to microwave tomography (Baran, et.al., 2014).

2.4 Breast Phantom

Breast phantom is the best platform to test the breast cancer detection system before further investigation in microwave technology. Thus, researchers need accurate geometrical properties of the breast phantom, natural heterogeneity of the breast structure and the dispersive properties of the breast tissue because it can affect the signal transmission (Conceiçãao, O'Halloran, Glavin & Jones, 2011). Many breast phantoms are developed with different dielectric properties by using different types of material (Hahn & Noghanian, 2012; Lazebnik, Madsen, Frank & Hagness, 2005; Porter, Fakhoury, Oprisor, Coates & Popovic, 2010; Klemm, Leendertz, Gibbins, Craddock, Preece, & Benjamin, 2009; Alshehri, Khatun, Jantan, Raja Abdullah, Mahmood & Awang, 2011). A mixture of oil and water is the basic materials used in breast phantom development. Like Klemm (2009) uses different ratio of water, TX151 material, and polyethylene powder in developing inhomogeneous breast phantom as shown in Figure 2.10. Similarly, Keenan (2016) uses corn syrup solution and grapeseed oil while

Zarafshani (2017) uses two types of agar saline solution with different levels of conductivity to develop breast phantom as shown on Figure 2.11.

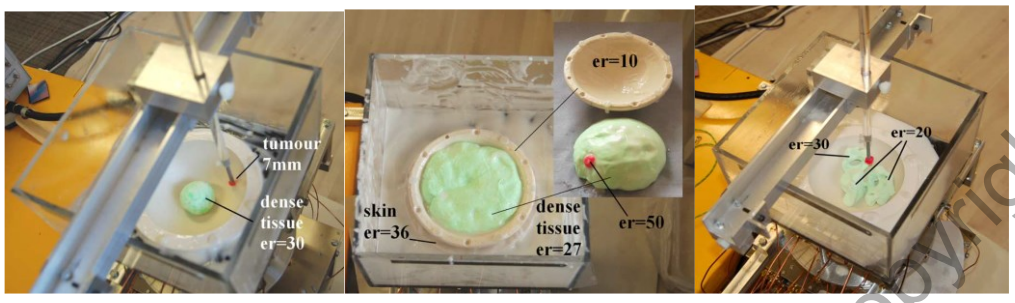


Figure 2.10 Inhomogeneous Phantom with Different Dense Tissue (Klemm, et.al., 2009)

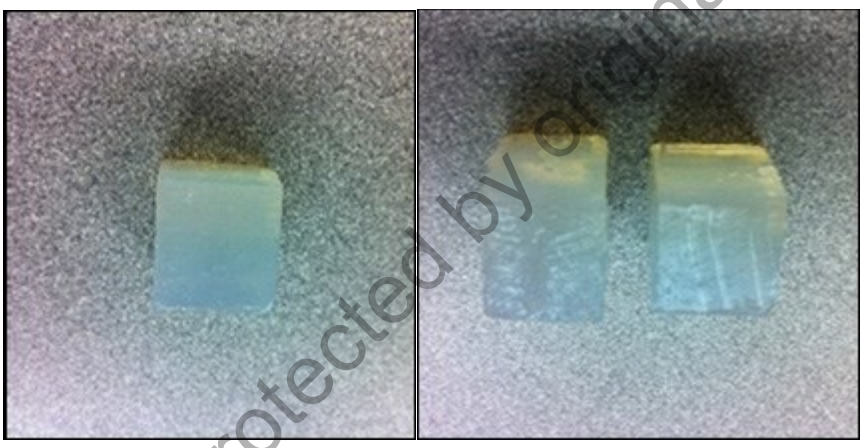


Figure 2.11 1cm^3 cube size of Fat and Stroma Tissue (Zarafshani, Dhurjaty, Wang, Tang, Xiang & Zheng, 2017)

Here, the researchers vary the amount of water in order to develop different dense of breast which resulted different value of dielectric properties. However, the real-like breast phantom is also still cannot be achieved since the researchers only considered fat, fibroglandular and skin of the breast and the conductivity is not close towards the real breast phantom (Hahn, 2012). Therefore, Hahn (2012) proposes a heterogeneous breast phantom by using different mixture ratio of distilled water, safflower oil, propylene glycol, gelatine, formalin and surfactant to develop breast phantom with four different layers i.e.: skin, fibro glandular, transitional and fatty tissue