Discrete thermal data analysis using various classifiers for early detection of breast cancer

Tan, Joanna Ming Yi

2009

Tan, J. M. Y. (2009). Discrete thermal data analysis using various classifiers for early detection of breast cancer. Master's thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/47134>

<https://doi.org/10.32657/10356/47134>

Nanyang Technological University

Downloaded on 26 Apr 2023 18:46:49 SGT

DISCRETE THERMAL DATA ANALYSIS USING VARIOUS CLASSIFIERS FOR EARLY DETECTION OF BREAST CANCER

TAN MING YI JOANNA

**SCHOOL OF MECHANICAL AND
AEROSPACE ENGINEERING**

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Masters in Engineering

2009

ACKNOWLEDGEMENT

The completion of this thesis was possible, thanks to the guidance and assistance of the people who have helped the author achieve more than the aspects of the project. Thus, the author would like to take this opportunity to express her sincere gratitude to the following people:

Firstly, the author would like to express her deepest appreciation and gratitude to her supervisor, A/P Ng Yin Kwee Eddie, for his invaluable guidance and also for his sharing of knowledge, allowing her to widen her perspectives for the understanding of the project. She would also like to thank him for taking care of her welfare in the course of the study.

She would also like to thank Dr Rajendra Udyavara Acharya for his kind assistance in the project. His experience and words of advice have encouraged and broaden the author's knowledge in the field of research.

Last but not least, she would like to thank all who have helped and guided the author in one way or another, making this project a fulfilling and enriching experience for her.

TABLE OF CONTENTS

ACKNOWLEDGEMENT I

SUMMARY V

LIST OF FIGURES.....VII

LIST OF TABLES IX

CHAPTER 1: INTRODUCTION..... 1

 1.1: BACKGROUND..... 1

 1.2: OBJECTIVE 4

 1.3: SCOPE 4

 1.4: OUTLINE OF THE THESIS..... 5

CHAPTER 2: LITERATURE REVIEW 6

 2.1: CURRENT MODALITIES FOR EARLY DETECTION OF BREAST CANCER 6

 2.1.1: Mammography..... 6

 2.1.2: Magnetic resonance imaging (MRI)..... 9

 2.1.3: Ultrasound..... 10

 2.1.4: Thermography 10

 2.2: CIRCADIAN RHYTHM CHAOS: A NEW BREAST CANCER MARKER..... 12

 2.3: DYNAMIC THERMAL ANALYSIS 16

 2.3.1: Evolving role for early detection of breast cancer 16

 2.3.2: Potential tool for early detection of breast cancer..... 18

 2.3.3 Sensitivity and Specificity improvement with ANN..... 20

 2.3.4 Exploration on different analytical software to enhance on the thermal analysis tool .. 22

CHAPTER 3: DATA ACQUISITION AND PREPROCESSING..... 23

 3.1: DATA COMPILATION 23

 3.2: DATA CLASSIFICATION 26

 3.3: DATA INSPECTION AND RECTIFICATION 27

CHAPTER 4: CLASSIFIERS AND RECEIVER OPERATING CHARACTERISTIC (ROC).....31

4.1: BACK-PROPAGATION ALGORITHM (BPA)	31
4.2: PROBABILISTIC NEURAL NETWORK (PNN)	34
4.3: FUZZY (SUGENO-TYPE)	36
4.4: GAUSSIAN MIXTURE MODEL (GMM)	39
4.5: SUPPORT VECTOR MACHINE (SVM)	41
4.6: FUZZY EQUIVALENCE	45
4.6.1 <i>Algorithm Prediction using Fuzzy Equivalence Relations</i>	45
4.7: COMPARISON ON THE ADVANTAGES AND DISADVANTAGES OF THE CLASSIFIERS	50
4.8: STATISTICAL ANALYSES USING RECEIVER OPERATING CHARACTERISTIC	53
(ROC)	53

CHAPTER 5: COMPARATIVE PERFORMANCE OF THE CLASSIFIERS55

5.1: COMPARISON STUDIES ON THE SIX CLASSIFIERS	55
5.1.1: <i>Results of the six classifiers</i>	56
5.2: PERFORMANCE COMPARISON FOR THE THREE CLASSIFIERS.	59
5.2.1: <i>Receiver Operating Characteristic (ROC) Analysis</i>	61

CHAPTER 6: IMPROVEMENT ON THE METHODS OF TEMPERATURE SCALING.64

6.1: TESTING ON THE CLASSIFIERS USING THE NEW NORMALIZATION	70
---	----

CHAPTER 7: PROCESSING DATA IN VARIOUS TIME SEQUENCES...72

7.1: CLASSIFICATION OF PATIENTS FOR VARIOUS TIME SEQUENCES.....	72
7.2: CLASSIFIERS' PERFORMANCE IN THE 10 MINUTES ANALYSIS	76
7.3: CLASSIFIERS' PERFORMANCE IN THE 15 MINUTES ANALYSIS	78
7.4: CLASSIFIERS' PERFORMANCE IN THE 20 MINUTES ANALYSIS	79
7.5: CLASSIFIERS' PERFORMANCE IN THE 25 MINUTES ANALYSIS	80
7.6: DISCUSSION ON THE CLASSIFIERS' PERFORMANCE IN THE VARIOUS TIME SEQUENCES.....	81

**CHAPTER 8: COMPARISON OF STATISTICS TECHNIQUES WITH
ARTIFICIAL NEURAL NETWORK TECHNIQUES 83**

8.1: T-TEST ON THE TWO DIFFERENT DIAGNOSES 86

8.2: ANOVA ANALYSIS ON THE 3 DIFFERENT CLASSES (NORMAL, BENIGN AND CANCER) 91

8.3: ANOVA ANALYSIS ON THE 4 DIFFERENT CLASSES (NORMAL, BENIGN, CANCER, SUSPECTED CANCER)..... 96

**CHAPTER 9: DISCUSSION ON THE PERFORMANCE OF THE
CLASSIFIERS 98**

CHAPTER 10: CONCLUSION AND FUTURE WORKS 100

10.1: CONCLUSION 100

10.2: LIMITATIONS AND FUTURE WORK 103

REFERENCES 105

**APPENDIX 1: ANOVA ANALYSIS ON THE STATISTICAL
PARAMETERS FOR THE 3 DIFFERENT DIAGNOSESI**

**APPENDIX 2: ANOVA ANALYSIS ON THE STATISTICAL
PARAMETERS FOR THE 4 DIFFERENT DIAGNOSESVIII**

SUMMARY

Breast cancer is the second leading cause of death in women. It occurs when cells in the breast begin to grow out of control and invade nearby tissues or spread throughout the body. The limitations of mammography as a screening modality, especially in young women with denser breasts, necessitated the development of novel and more effective screening strategies with acceptable sensitivity and specificity. In this work, the interpretive software system was further enhanced and developed for the detection and classification of the breast cancer patients. This system incorporated dynamic thermal analysis and analytical software to produce a potential tool for detection of breast cancer.

The performance of the six classifiers; namely, Fuzzy Equivalence, Back-Propagation Algorithm (BPA), Probabilistic Neural Network (PNN), Fuzzy (Sugeno-type), Gaussian Mixture Model (GMM) and Support Vector Machine (SVM) was compared and evaluated. The accuracy of these 6 classifiers generally depends on the size and quality of the training data, the rigor of the training imparted and also the parameters used to represent the input (breast surface temperature). GMM and Fuzzy Equivalence had performed better than the other classifiers and they were able to obtain at least 85% of accuracy in most of the analysis.

A study on the different methods of temperature scaling was demonstrated and this was to determine a more suitable normalization method which was independent on

the temperature scaling. $\left[\frac{(T_{\max} - T)}{(T_{\max} - T_{\min})} \right]$ was introduced as the new temperature

scale and the performance of the classifiers was not affected by the new method as they were able to attain an average accuracy of 83.1%.

In the various time sequences analysis, shorter time duration of temperature data was collected to test on the accuracy of the six classifiers. The results showed that the four classifiers did not perform relatively well. They had only managed to achieve approximately 50% to 80% of accuracy. The decrease in percentage could be due to the small number of temperature data available for classification. Hence, the analysis showed that longer time duration of temperature data in classifying the different diagnoses would yield higher accuracy.

Lastly, using the statistics techniques, the temperature data was processed into various statistical parameters and tested on the classifiers. The aim is to compare and evaluate on the two methods: statistical and artificial neural network. The result had shown that artificial neural network being non-linear, is a more powerful method in classifying the various diagnoses correctly. This had further assured us in using this method for classification and detection of the breast cancer.

In conclusion, fuzzy equivalence and GMM have shown greater capability in performing better than the rest of the classifiers. Regardless of any analysis, both classifiers had obtained relatively good results. The current work had further enhanced on the application of analytical software techniques together with the real-time functional thermal analysis so as to develop a potential tool for the detection and classification of breast cancer.

LIST OF FIGURES

Figure 2.1: Dynamic thermal analysis for patient diagnosed with fibroadenoma	17
Figure 2.2: Dynamic thermal analysis for patient diagnosed with TI breast cancer	17
Figure 3.1: Sensors and the relevant equipment	24
Figure 3.2: Sensors attached to the surface of the breasts	24
Figure 3.3: Specific location of the 16 sensors	24
Figure 3.4: Drastic fluctuations of the temperature in several sensors for patient #1001	28
Figure 3.5: Drastic fluctuations of the temperature in several sensors of a typical subject	28
Figure 4.1: Feed-forward classifier used for the 4 classifications	33
Figure 4.2: Probabilistic neural network structure	34
Figure 4.3: Block diagram of Sugeno rule diagram	37
Figure 4.4: Working of SVM algorithm	42

Figure 4.5: Flowchart for prediction using fuzzy equivalence relations	47
Figure 5.1: Comparison of ROC curves for the 6 different classifiers	58
Figure 5.2: Comparison of ROC curves for the three different classifiers	63
Figure 6.1: Temperature data in sensor 1 for normal and cancer patient	65
Figure 6.2: Temperature scaling of method 1: $abs\left[\frac{(T - T_{ave})}{(T_{max} - T_{ave})}\right]$	66
Figure 6.3: Temperature scaling of method 2: $abs\left[\frac{(T - T_{ave})}{S_{tdev}}\right]$	67
Figure 6.4: Temperature scaling of method 3: $\left[\frac{(T - T_{min})}{(T_{max} - T_{min})}\right]$	67
Figure 6.5: Temperature scaling of method 4: $\left[\frac{(T_{max} - T)}{(T_{max} - T_{min})}\right]$	68
Figure 7.1: Average accuracy of the classifiers in different time duration	81
Figure 7.2: Percentage of classification in different time duration	82
Figure 8.1: Example of temperature measurements and tabulation of statistical parameters.	86
Figure 8.2: Diagram on the 3 parameters being fed into the classifier	93

LIST OF TABLES

Table 2.1: Result on mammography findings	18
Table 2.2: Results on dynamic thermal analysis	18
Table 3.1: Patients classification	26
Table 4.1: Detailed algorithm for Prediction using Fuzzy Equivalence Relations	48
Table 4.2: Comparison of the 6 classifiers	50
Table 4.3: Summary of definitions for true positive, false positive, false negative and true negative	53
Table 4.4: Summary of formula for sensitivity, specificity, positive/negative likelihood ratio and predictive value	54
Table 5.1: Number of training and testing data used for classification	56
Table 5.2: Comparison on the percentage of correct classification for the classifiers	57
Table 5.3: Results of Performance Indices for the six classifiers used	58
Table 5.4: Number of training and testing data used	59

Table 5.5: Comparison on the percentage of correct classification for the three classifiers	59
Table 5.6: Results of ROC for the three classifiers	61
Table 6.1: Tabulated results for methods 3 and 4	69
Table 6.2: Comparison on the percentage of correct classification for the 5 classifiers	71
Table 7.1: Number of patients used for the various time sequence study	74
Table 7.2: Number of training data and testing data used for various time sequence study	75
Table 7.3: Result of the 4 classifiers in the 10 minutes analysis	77
Table 7.4: Results of the 4 classifiers in the 15 minutes analysis	78
Table 7.5: Results for the 4 classifiers in the 20 minutes analysis	79
Table 7.6: Results for the 4 classifiers in the 25 minutes analysis	80
Table 8.1: Definition and formula of the statistical parameters	83
Table 8.2: Number of temperature data	86

Table 8.3: Result of t-test for the two different diagnoses (normal and cancer)	89
Table 8.4: Result of ANOVA for 3 the different diagnoses	92
Table 8.5: Result of correct classification using statistical parameters	94
Table 8.6: Result of correct classification for 3 different diagnoses	94
Table 8.7: Result of ANOVA between the 4 different diagnoses	96
Table 8.8: Result of classifier for 4 different diagnoses	97

Chapter 1: Introduction

1.1: Background

Breast Cancer is a heterogeneous disease which varies in different women of different age group and different cell populations within the tumor itself. It occurs when malignant cells in a part of the body grow out of control and start to invade nearby tissues or spread throughout the body. Even till today, breast cancer still ranked second only to lung cancer as a cause of cancer death in women. The number of women developing breast cancer has increased tremendously from 1:20 in 1960 to 1:7 today [1]

According to the American Cancer Society [2], in 2008, an estimated 182,460 new cases and 40,480 deaths from breast cancer (in women) are expected to occur in the United States. However, the death rate from breast cancer had gradually decreased since 1990 and this could probably due to the availability of early detection and improved treatment in breast cancer. Given such circumstances, early detection of breast cancer is considered a vital prognostic factor, as statistic have shown in the five year relative survival rate for localized breast cancer, there is a rise from 80% in the 1950s to 98% today[2].

Mammography is currently the most commonly used screening modality for the early detection of breast cancer. However, it is a gold standard that does not always shine brightly since published reports have shown wide variations in its sensitivity and specificity [3]. Furthermore, its limitation on young and premenopausal women had hindered its clinical value and effectiveness in mammogram, as denser breast tissue produces mammographic images which are difficult if not impossible

to interpret. In addition, the cost and radiation exposure related to mammograms do not justify its use in young women with low risk of developing breast cancer [4].

New modalities, such as magnetic resonance imaging (MRI), have been shown to be more sensitive in the early detection of occult breast cancers, particularly in premenopausal women in whom the sensitivity of mammography is compromised [5]. In the recent reports, MRI is able to detect cancer in the contralateral breast that is missed by mammography and/or in clinical examination at the time of the initial breast examination [6]. In addition, MRI has proved to be a better screening tool for women with BRCA1 or BRCA2 mutations [7], and in those women with strong family history of breast cancer [8, 9]. Although the sensitivity of MRI is better than that of mammography, the technique is flawed by a lower specificity and a far greater cost. However, recently, The American Cancer Society announced a change in its breast cancer screening guidelines, recommending that women with high genetic risk (such as those who have mutation in the BRCA1 or BRCA2 genes or those with a strong family history of breast cancer) be screened with magnetic resonance imaging [10].

The breast has been recognized to exhibit a circadian rhythm which reflects its physiology. There now exists increasing evidence that those rhythms associated with malignant cell proliferation are largely non-circadian [11, 12]. Gros et al, have examined the relationship between breast skin temperature and breast cancer as well as the differences between the characteristics in circadian rhythm changes in skin temperature of clinically healthy and cancerous breasts and found that these changes were both definitely real and measurable [13, 14]. Furthermore, comparison of rhythm parameters associated with the tumour area and corresponding site on the contralateral breast showed that abnormal thermal characteristics such as elevated mesor values, decreased amplitude and changes in

the timing of the acrophase may be used in the exploration for the early detection of breast cancer [15].

Thermographic technology was originally introduced to complement mammography, because it was felt that thermography was able to detect early breast cancer development up to 10 years earlier than with most conventional methods[16]. However, the accuracy of thermography has remained uncertain. Several factors, such as the symmetry and stability of the breasts' temperature during the menstrual cycle and the use of oral contraception, have been confusing. Presently, there is no ideal technique used to evaluate the results of all methods of early breast cancer detection. Studies on the use of a

feasible and non invasive dynamic thermal analysis[12, 17] have been carried out and reported to be sensitive in detection of breast cancer. These were further enhanced with the use of the artificial neural network systems to implement a smart thermal analysis tool [18, 19].

In addition, it was reported that various analytical software such as Gaussian mixture model and support vector machine were also used in the thermal analysis tool. Their performance was compared with the artificial neural network systems and it seems that Gaussian mixture model had produced the highest percentage of correct classification for unknown data and sensitivity [20]. In order to further enhanced on the development of the interpretive software system, more studies and improvements were required to reassess on this thermal analysis tool, as it may provide promising and significant contributions to the medical and research areas.

1.2: Objective

The aim of this study is:

1. To determine the best classifiers (able to achieve within 90% of specificity and sensitivity) used for the development of the interpretive software system.
2. To find out if the classifiers were able to obtain same result (sensitivity and specificity) when shorter duration of temperature data is being processed.
3. To find out the feasibility of using statistical parameters of temperature data for classifying the various diagnoses.

1.3: Scope

In this study, the performance of the six classifiers (fuzzy equivalence, back-propagation algorithm, probabilistic neural network, fuzzy (Sugeno-type), Gaussian mixture model and support vector machine) were being evaluated to determine which classifier had greater ability in classifying the unknown data for the four different diagnoses. Statistical analysis of receiver operating characteristic (ROC) was used to determine the accuracy of the classifiers based on their specificities, sensitivities and areas under the curve. Secondly, the temperature data was processed into various time sequences and the aim is to find out whether the accuracy of the classifiers will be affected when shorter duration of temperature data are being used. Lastly, the temperature data will be processed into various statistical parameters to find out the feasibility of using these parameters for classification and also determine which analytical method is more powerful.

1.4: Outline of the thesis

Chapter 1 gives the introduction to the topic and the project. The literature review is compiled and critically assessed in Chapter 2. Chapter 3 will provide a detailed outline on the acquisition and preprocessing of the data collected from Lifeline Biotechnologies, Inc, Florida, USA [21]. A theoretical background on the six classifiers and application of receiver operating characteristic (ROC) will be briefly explained in Chapter 4. The performance of the six classifiers will be evaluated in Chapter 5. The improved method for temperature scaling will be elaborated in Chapter 6. Processing of the temperature data in various time sequences will be evaluated in Chapter 7. The feasibility of using statistical parameters in temperature data for classification will be analysed in Chapter 8. Chapter 9 will provide a discussion on the performance of the six classifiers and conclusion will finally be addressed in Chapter 10.

Chapter 2: Literature Review

As a prerequisite to start the work, an initial analysis was carried out on the current available modalities for early detection of breast cancer. This literature study also provides information on new findings of circadian rhythm chaos related to the breast cancer. Emergence of a non-invasive and painless thermal analysis device was also discussed in this Chapter and this will further enhanced the work in this thesis.

2.1: Current modalities for early detection of breast cancer

Early detection of breast cancer is crucial as it allows 98% of five year survival rate. Currently, mammography is the gold-standard modality used by hospitals for the detection of breast cancer. Moreover in recent years, the technologies on imaging had changed and improved significantly. Detailed descriptions on the modalities will be provided.

2.1.1: Mammography

Mammography plays a central part in early detection of breast cancers because it can show changes in the breast up to two years before a patient or physician can feel them. There are screening mammography and diagnostic mammography. A diagnostic mammogram is used to diagnose breast disease in women who have abnormal clinic findings such as breast lump. It can also be done after an abnormal screening mammogram is being observed and this is to determine the cause of the abnormalities on the area of concern.

Screening mammogram is used to look for breast disease in women who are who appear to have no breast problems. It utilizes a small dose of ionizing radiation to

image breast tissue. The examination is performed by compressing the breast firmly between a plastic plate and an x-ray cassette that contains special x-ray film. Examination films are taken in mediolateral oblique and craniocaudal projections in which both views should include breast tissue from the nipple to the pectoral muscle. Two-view examinations decrease the recall rate compared with single-view examinations by eliminating concern about abnormalities due to superimposition of normal breast structures [22].

Most screening mammography uses screen-film technology. However, there are currently two recent enhancements to traditional mammography and these include digital mammography and computer-aided detection (CAD). Digital mammography, also known as full-field digital mammography (FFDM), is a system that replaced x-ray film with solid-state detectors which convert x-rays into electrical signals. These signals are used to develop images of the breast which can be seen on a computer screen or printed on special film similar to conventional mammograms. From the patient's point of view, digital mammography study is exactly like film mammography. However, it works faster than film mammography as its images are acquired in less than a minute, while the latter required two to five minutes for the film to be developed by the technologist [23].

Computer-aided detection (CAD) is a sophisticated computer program which use a digitized mammographic image obtained from either a film mammogram or a digitally acquired mammogram. CAD will search for areas of density, mass, or calcification with potentially suspicious areas that may indicate the presence of cancer. The system highlights these areas on the images and alerts the radiologist to review the area again. Studies had shown that using CAD had increased the accuracy of mammography by up to 20% [24].

2.1.1.1: Issue related to mammography's deficiencies

Mammography is currently the gold standard screening modality for the detection of early breast cancer and it is considered as a major achievement in the annals of modern medical imaging by the radiologists. However in the review of the mammography literature published in the last decade, it reads more like an apology for its deficiencies rather than a testimonial to its sensitivity in detecting breast cancer [25].

The sensitivity of mammography is the proportion of breast cancer detected when breast cancer is present. The overall sensitivity is approximately 75% [26] but it ranges from 54% to 58% in women younger than 40 years to 81% to 94% in those older than 65 years [27, 28]. It is a screening tool facilitated for older women whose breast tissues are primarily fatty and provide better visualization during mammography scanning. Women under 40 are not advised to undergo mammography unless their families have a clear history of breast cancer in one or more first-degree relatives. This is due to the denser breast tissue of the younger women which hinder the mammography effectiveness. In addition, the exposure to the potentially hazardous effects of radiation does not justify for the low risk of developing breast cancer in younger women.

An important factor in determining the mammographic sensitivity is the radiologist's interpretation. Studies have shown substantial variability in interpretation and reading accuracy among radiologists [29-37]. This does not provide patients or physicians the confidence on the accuracy of mammography's sensitivity and specificity. Identifying a radiologist who is more accurate than

another is difficult. Hence, using mammography as the sole screening tool does not justify for its result in the early detection of breast cancer.

Furthermore, in the recent enthusiasm to promote mammography screening may lead women to unrealistic anticipations, influencing them to falsely believe that a negative examination is an assurance for the absence of cancer in its earliest detectable stage. Thus, it is necessary to complement the physical examination and mammography with a third screening modality based on the concept of the presence of circadian rhythm breast temperature chaos. This is a non-invasive and non-radiogenic tool which allows the clinicians to provide patients with every opportunity for early diagnosis.

2.1.2: Magnetic resonance imaging (MRI)

Magnetic Resonance Imaging (MRI) uses a strong magnetic field and radio waves to create images of the body. Radiofrequency (RF) pulses are generated to interrogate the hydrogen atoms that make up the human body. Powerful computers process the returning RF signals (echoes) to form medical images.

MRI has a higher sensitivity in the diagnosis of breast cancer ranging from 86-100% than mammography which only had a sensitivity of 20-50%. Many studies have confirmed the ability of MRI to detect cancer in the contralateral breast that is missed by mammography and/or in clinical examination at the time of the initial breast examination [6]. However, MRI has poor specificity and the cost for MRI scans is exorbitant which makes it unaffordable for many people. Recently, The American Cancer Society announced a change in its breast cancer screening guidelines, recommending that women with elevated risk (such as those who have mutation in the BRCA1 or BRCA2 genes or those with a strong family history of breast cancer) be screened with MRI [10].

2.1.3: Ultrasound

Ultrasound uses high-frequency sound waves to outline a part of the body. These sound waves are transmitted into the area of the body concerned and echoed back. The echoes are absorbed by the ultrasound probe, and it will be further altered into an image which will be displayed on a computer screen. There is no exposure to radiation during the ultrasound scanning.

Ultrasound cannot replace mammography in breast cancer screening as it cannot reliably identify microcalcifications which lead to its high false positive rate. However, it may be used to complement mammography in women with dense breasts as results have showed that screening with ultrasound allowed the detection rate to increase by 17% [38]. Furthermore, it was reported by researchers in May 14 issue of The Journal of the American Medical Association that 28% more cancers were found when ultrasound was added to mammography as compared to when mammography was the only screening method used [39]. Therefore, ultrasound can be used as a complement tool together with mammography in the detection of breast cancer.

2.1.4: Thermography

In thermography scanning, there is neither x-ray radiation nor pain encountered. It is based on the digital infrared thermal imaging (DITI), which uses the principle of higher metabolic activity and vascular circulation found in both pre-cancerous tissue and area surrounding a developing breast cancer. In these areas, there will be an increase in the regional surface temperatures of the breast. DITI uses ultra-sensitive medical infrared cameras and sophisticated computers to detect and analyse. This will produce high-resolution diagnostic images which show the

variation of the temperature. In this project, the modality used is based on circadian rhythm which is a similar concept to thermography.

Thermography has the ability to image the thermovascular aspects of the breast, extremely early warning signals (from 8-10 years before any other detection method) have been observed in long-term studies. Furthermore, thermography is the earliest known indicator for the future development of breast cancer. It is for this reason that an abnormal infrared image is the single most important marker of high risk for developing breast cancer. Thus, thermography has an important role in becoming one of the major front-line methods for breast cancer detection.

2.2: Circadian rhythm chaos: a new breast cancer marker

In the lack of evidence for the sensitivity and specificity of mammography and also the physical evidence of breast cancer, the presence of circadian rhythm breast temperature chaos may be a vital addendum for early detection of breast cancer. Circadian rhythm chaos is a term used to indicate the disruption of the normal circadian rhythm fluctuations that occur in the healthy breast [25]. It can be defined as statistically validated physiologic changes recurring with a reproducible waveform. The biological times series of rhythms can be characterized as having a time period of twenty hours.

Variations in measurements of breast temperature and their related circadian rhythms chaos may be the secondary effects of tumour angiogenesis. When these circadian variations are compared to normal circadian rhythms, it is possible to identify a group of women at high risk for breast cancer development. It has been suggested that a circadian rhythm shifts may correlate well with tumour formation. An increase in the local temperature on the skin surface of the breast is due to the growth of angiogenesis (new blood vessels) found near the tumours. The lack of receptivity in angiogenesis had caused a rise in the amount of constant blood flow, and hence increases the local temperature.

The circadian rhythm fluctuation of most physiological variables is not apparent. It becomes overt only when properly measured at frequent intervals along a twenty-four hour time scale. Because of the “invisible” nature of these rhythms, many investigators have ignored them, thinking that the circadian variation represents only a minor labor required to explore it extensively. However, research over the past forty years indicates the potential value of circadian rhythm analysis as a high-risk marker in early cancer detection.

As early in 1938, Fay and Henny [40] were the first modern investigators to realize the differences in the temperature of the skin overlying cancers of the breast but delayed the publication of their findings for many years. In 1953, Voutilainen reported extensive work on the biologic rhythms of malignant tumours. Twenty-one cases of human cancers were studied and a clear biphasic rhythm was demonstrated with peak mitotic activity, occurring in the early morning between 2 and 4 am and during the day at 10 to 2 pm [69]. In 1956, Lawson [41] documented surface temperature increases in untreated breast cancer patients; one year later he described thermographic imaging and its capacity to detect a 1 to 3 degree Centigrade warmer temperature overlying malignant tumors compared to other areas of the breast [42]. From 1963 to 1974, Mansfield, Dodd and Somolensky have studied similar series of patients and showed circadian rhythm advances of more than three hours in cancerous breast as compared to the healthy breast [70].

The special relationship between circadian alteration and the degree of malignancy or growth rate was also reported by Lianos et al, in 1970 [71], Barcia-Sainz in 1966 [72] and Tahti in 1956 [73]. The analysis of cell mitoses in human tissues has shown that rhythms associated with malignant cell proliferation are largely non-circadian modulations. This shows that there are some defects in the time integration mechanism of cells which created a form of temporal anarchy and histological abnormality. In 1970, Luca also found that cancer cell activity falls outside of the overall temporal harmony of the body [74].

Gautherie and Gros in 1977 analysed a series of twenty-six cancer patients. In fifteen cases, the cancerous breast exhibited a shortened circadian rhythm. These patients were characterized as having rapidly growing histological undifferentiated tumours with increased thermo-circadian output while the remaining tumours were designated as slow growing with a wave form which appeared to have the same

shape as normal tissue but contained wave form characteristics with dynamics significantly different from those of normal tissue [43].

The significant connection between the disturbances of circadian rhythms and breast cancer is substantially reviewed in [11]. There are three observations concluded on the phenomenon of thermal and circadian rhythm chaos found in the breast.

1. The temperature and thermal conductivity of blood flow are higher in the cancerous breast than in the contralateral healthy breast. In the comparison between the temperature behaviour of a normal breast and a cancerous breast, the temperature change in the latter will produce a bell-shaped profile, while the temperature in the former will rise continuously from the skin into the deep tissue. Thermal conductivity usually remains constant in the healthy breast.
2. Changes in thermal and circulatory blood flow are directly proportional to the growth rate of the tumour. For instance, fast-growing tumours will exhibit higher temperatures and rising blood flow which will extend beyond the limits of the tumour. Such observation is contra to the slow-growing tumours in which changes occur might be undetectable.
3. The absence of mammographic or physical evidence of cancer does not preclude the presence of cancer in its early stages due to the existence of circadian rhythm breast temperature chaos. Patients facing such phenomenon can be considered 'high risk' for breast cancer, as evident had shown that majority of the breast cancer are diagnosed within 4-5 years of the abnormal thermal examination. In a study conducted by Gantherine et

al, 21.3% of patients who had abnormal thermograms but no abnormality on physical examination and mammography developed breast cancer within the next 3 years [44]

In the view of early diagnosis, tumours characterized by radical changes in the pattern of low frequency rhythms appear to have rapid growth rates, and exhibit undifferentiated cell structure. If this is true, the relationship between rhythm disturbances and angiogenic activity could lead to a breakthrough in the screening and management of breast cancer patients. The development of a therapeutic index could lead to improved treatment timing as a function of a probable susceptibility-resistance cycle of the patient.

The circadian rhythm fluctuations associated with thermo-circadian variations in the microvascular bed of tumors may reflect mitotic cell activity and represent a clinical method of obtaining or verifying the presence or formation of a cancerous lesion. It was reported by investigators that the circadian signals have preceded the diagnosis of visible breast cancers by several months to several years. There is also a direct relationship between the circadian signal of a cancer, the tumour-volume doubling-time and lymphatic permeation. If these factors are true, the feasibility of using this modality for detection, quantification and evaluation of the biologic phenomenon will definitely make a great contribution to the early diagnosis and treatment of breast cancer. The establishment of cyclic changes characteristic of breast cancer might well serve to identify the disease in the in situ phase, its most curable phase.

2.3: Dynamic thermal analysis

2.3.1: Evolving role for early detection of breast cancer

Dynamic thermal analysis is the evolving role for early detection of breast cancer. It is safe, non invasive and seems to be sensitive in the early detection of breast cancer especially among the young women. With the advancement in the technology, it is possible to record the variations of circadian rhythm in the breast and also analyze these data using highly sophisticated computer statistical software. Thermal information is collected from the eight sensors of each breast. This information is recorded and stored by a miniaturized microprocessor. The eight sensors are placed in anatomically critical positions elicited by data obtained from tumour registries as to where cancers are likely to develop.

The temperature data provided and conducted by Lifeline Biotechnologies, Inc, Florida, USA are measured using the First Warning System. The thermal data are collected every five minutes for a period of 48 hours. A total of 9000 pieces of data are recorded by microprocessors during the test period, and after which these data are analyzed using specially developed statistical software. The temperature points from each contralateral sensor are plotted against each other to form a thermal motion picture of a lesion's physiological activity. Figures 2.1 and 2.2 show the graphs of dynamic thermal analysis in two patients diagnosed with fibroadenoma and TI breast cancer respectively [12].

Figure 2.1: Dynamic thermal analysis for patient diagnosed with Fibroadenoma [12]

Figure 2.2: Dynamic thermal analysis for patient diagnosed with TI breast Cancer [12]

One of the main challenges to this new technology is the number of false positive cases found. This can cause confusion for women who are identified either as being positive or high risk by dynamic thermal analysis in the absence of mammographic or physical evidence of cancer. This group of women may or may not have cancer in its earliest stages. Further analysis of the thermal data using artificial neural network (ANN) may increase the sensitivity and reduce the number of false positives cases. As promising result was obtained during the initial stage, dynamic thermal analysis is foreseen as a potential tool in the detection of breast cancer.

2.3.2: Potential tool for early detection of breast cancer

In Chapter 2.3.1, dynamic thermal analysis was introduced as the new evolvement tool in the early detection of breast cancer. Evaluation on the feasibility of this tool will be discussed in this chapter to determine its potential as a breast cancer screening tool [17].

The method for measurement and collection of temperature data were mentioned in the previous chapter. There were a total of 173 women underwent mammography, and 118 (68%) of them were found to have breast cancer in the final histology. Tabulated result on mammography screening was found in Table 2.1. In dynamic thermal analysis, thermal data obtained from breast surface were analyzed in 160 participants. Out of these 160 participants, 113 of them were classified as positive while the remaining 47 were found to be negative. Thermal abnormalities were found in 76 (72%) out of 105 breast cancer cases and 37 (67%) out of the 55 benign cases. The result for dynamic thermal analysis was shown in Table 2.2.

Table 2.1: Result on mammography findings [17]

Test	Disease			
	Cancer Present	118	Cancer Absent	55
Mammogram	True Positive	55	True Positive	32
	47%		58%	
	False Negative	9	False Negative	7
	8%		13%	
	Indeterminate	54	Indeterminate	16
	46%		29%	

Table 2.2: Results on dynamic thermal analysis [17]

Test	Cancer Patients		Cancer absent	
DTA	True Positive	76	True Positive	18
	72%		33%	
	False Negative	29	False Negative	37
	28%		67%	

A new algorithm using multiple-layer perception and SoftMax output artificial neural network (ANN) was developed. This algorithm was used on a subgroup of participants (n=38). MLP had improved the sensitivity to 76% by detecting cancer correctly in 16 out of 21 participants. The number of false positive cases was also decreased to 26% which further improved the specificity to 74%.

From the result obtained, it showed that mammography had low positive predictive value as compared to dynamic thermal analysis. And with the use of ANN in dynamic thermal analysis, it was able to analyse the thermal data recorded and hence improve the sensitivity and specificity of the test. However, ANN required 540 input data for a reliable analysis, thus there is a need to validate this new algorithm further.

The use of dynamic thermal analysis of the breast can be concluded as a feasible and non invasive method which seems to be sensitive for the detection of breast cancer. The limitation on the specificity of this test can be further enhanced using ANN. Prospective multi-centre trials are required to validate this promising modality as an adjunct to screening mammography especially in young women with dense breasts.

2.3.3 Sensitivity and Specificity improvement with ANN

Dynamic thermal analysis is a potential tool for the early detection of breast cancer. This tool can be further enhanced with the use of ANN, as the latter has the capability of improving the specificity and sensitivity of the tool. However as mentioned previously in Chapter 2.3.2, more input data are required for the ANN to have a reliable analysis. Therefore in this chapter, two recent papers are presented to evaluate the combination of both the dynamic thermal analysis and the ANN as the intelligent tool for early detection of breast cancer.

There are various types of the ANN available, thus as a start off in using the ANN, exploration on the different combinations of networks is necessary as this allows the best networks to be used in the detection and classification of breast cancer [18]. The ANN is able to detect and learn from new data when more databases are collected during routine screening. Moreover, the data provided for ANN must be 'clean' and have 'no abnormalities'. This is to prevent any errors from occurring, thus data preparation is a vital phase in project management.

In this paper [18], two possible types of networks; namely MLP and radial basis function (RBF) are used. Parameter settings for the two networks are carried out by using the small data file (13 patients), as this limit the time and resources needed to perform the test if the large data file were used instead. From the result obtained in the small data file, the two networks managed to achieve nearly 90% of correct classification. However when the large data file (38 patients) was used, only 51% and 48% of correct classification were obtained for MLP and RBF respectively. A leave-one-out approach was also used to test the robustness of the ANN, and after which receiver operating characteristic (ROC) was used to compute the uncertainty and applied two scoring systems to calculate the classification. A comparison

between RBF and MLP shows that, RBF is able to classify data that is similar to the trained environment, whereas MLP with SoftMax output performs better when new or unknown data are presented for mass screening. Hence, these two networks are able to work hand-in hand to form a perfect team on testing various kinds of data.

In order to further evaluate the feasibility of applying the ANN in the detection and classification of breast cancer, a study was carried out [19]. It uses 1170 sixteen sensor data collected from 54 patients. Three different kinds of breast conditions: namely, normal, benign and cancerous breast are found in these patients. Comparative work was done on the two different kinds of ANN classifiers: the back-propagation algorithm classifier (BPA) and the radial basis function classifier (RBF). Temperature data from the 16 temperature sensors on the surface of the two breasts (eight sensors on each side) are fed as input to the classifiers. In the result shown from the study, the BPA and RBF classifiers are able to achieve a sensitivity of 84% and 91% respectively, and also a specificity of 100%. Therefore, the application of the ANN for the early detection and differentiation of abnormal patient states in health screening was achievable with the use of thermographic technology.

2.3.4 Exploration on different analytical software to enhance on the thermal analysis tool

In the paper [19], two different ANN classifiers such as back-propagation algorithm (BPA) and radial basis function (RBF) were used to identify three different stages of breast cancer, namely normal, benign and cancerous breast. These classifiers were able to achieve at least 80% of sensitivity and specificity. However, to further enhance on the development of the dynamic thermal analysis tool, other analytical software techniques were used to test on this potential tool [20].

Temperature data was collected from 90 individuals exhibiting four different breast conditions, namely: normal, benign, cancer and suspected-cancer. 6000 pieces of thermal data were collected from sixteen-sensors with eight placed on the surface of each breast. The temperature data collected from these 16 sensors placed on the surface of the each breast were fed as inputs to the classifiers. Comparison works were made on the five different kinds of classifiers: back-propagation algorithm, probabilistic neural network, fuzzy (Sugeno-type), Gaussian mixture model and support vector machine.

These classifiers were able to attain approximately 80% accuracy in classifying the four different diagnoses. Gaussian mixture model was the most sensitive classifier, achieving the highest sensitivity of 94.8%. Support vector machine was considered the best classifier as it was able to produce the most specific and accurate results. These evaluations show the feasibility of applying analytical software techniques together with the real-time functional thermal analysis to develop a potential tool for the detection and classification of breast cancer. More analytical software will be used in this study to reassess and improve on this thermal analysis tool.

Chapter 3: Data Acquisition and Preprocessing

3.1: Data Compilation

The measurement of the patients' temperature was provided and conducted by Lifeline Biotechnologies, Inc, Florida, USA [21]. Thermal data was measured and recorded by the Lifeline's equipment (First WarningTM) as shown in Figure 3.1. There were a total of sixteen sensors, with eight sensors being placed on each breast. The sensors were a biological series of thermistors manufactured for Lifeline by Yellow Springs Instrument Company (YSI). YSI is reputable for its most accurate thermal test equipment in the world and they supply the US Bureau of Standards with all their equipment. They have guaranteed the thermistors to be interchangeable with a repeatable accuracy of one; one hundredth of a degree Centigrade. Placement of sensors was accomplished with the use of a templet design from information provided by the national tumor registry to insure that the information was collected in areas of the breast where most breast cancers develop.

The sixteen sensors are marked number 8L and 8R for easy identification of each thermistor. Each thermistor and its signals are consequently identified with a specific position on the breast. This consistency simplifies subsequent processing and improves accuracy of the signals in terms of individual signal correlation with calibration data and selection of specific signal sources for manipulation in developing the generalization of physiological condition. The sensors are taped in place with rosette-shaped surgical tape which enabled intimate contact with the breast skin.

Examination was done concurrently on contralateral areas of both breasts under close monitoring and controlled ambient temperature for a specific period of time. Figures 3.2 and 3.3 show the sensors being attached to the surface of the breasts and the specific location of the sixteen sensors placed on each breast respectively. The system would record the temperature of all the sixteen sensors at a constant interval of approximately 5 minutes for duration of 48 hours.

In addition to temperature readings, each patient’s demographic details (age and status of menopause); mammogram and biopsy results were also provided by Lifeline Biotechnologies, Inc, Florida, USA [21], and this information was compiled in either Argentina or GMH Org Spreadsheets, depending on where the patients took their tests from. In the spreadsheet under the mammogram result, various stages of cancer such as benign and suspicious cancer were mentioned. The biopsy result showed the specific types of cancer such as ductual carcinoma and carcinoma NOS, being diagnosed by each patient. The information on the location of the tumour was also included. It was categorized as LUOQ/RUOQ, which refers

to the location of tumour at the left/right upper outer quadrant of the breast. Lastly due to confidentiality, each patient was identified with a patient number throughout the whole analysis process.

3.2: Data Classification

The temperature readings were classified according to the results of the patients' pathological diagnoses. Data were arranged into four different files, namely normal, benign, cancer and suspected cancer. The patients categorized under these classifications were based on the mammography data provided in the Argentina and GMH Org Spreadsheet. A total of 105 patients had undergone the First WarningTM system test. However, some of these patients were excluded from this study due to the abnormalities found in their temperature data. These abnormalities will be further elaborated in the following chapter. After compilation, there were a total of 67 patients involved in this data analysis. Table 3.1 shows the classification of these 67 patients being categorized under the four different diagnoses groups.

Table 3.1: Patients classification

Diagnosis	No of patients categorized in each group	Patient number (#)
Normal	11	#54, #88, #120, #144, #145, #150, #151, #152, #169, #178, #185
Benign	20	#40, #43, #47, #56, #68, #79, #81, #84, #92, #94, #95, #117, #125, #1002, #1003, #1010, #1011, #1013, #A103, #A107
Cancer	22	#16, #22, #42, #51, #60, #63, #69, #96, #97, #130, #145, #1004, #1005, #1018, #1022, #A104, #A105, #A114, #A117, #A121, #A127, #1016
Suspected Cancer	14	#18, #20, #23, #26, #27, #31, #33, #34, #50, #53, #86, #106, #1015, #1024

3.3: Data Inspection and Rectification

Data inspection was required to ensure that temperature readings were ‘clean’ and had no ‘abnormalities’. Inspection was done through graphical analyses by converting all the sets of temperature readings from each patient into graphs. Sixteen different colours representing the sixteen sensors used for temperature measurement were shown in the graph. If any abnormalities were found, such as great fluctuations of the temperature on a particular sensor, the patient’s data would be excluded from analysis. The potential causes for such phenomena are as follows:

- Poor contact between the sensors and the breast surface
- Continuous data recording after the screening system has shut off
- Sensors dropped off from the breast surface in the midst of temperature recording

Figure 3.4 shows the plot of temperature readings for patient #1001 at a particular time interval (minutes). There were sixteen different colour lines shown on the graph and they represent the sixteen sensors used for temperature measurement. From the graph, drastic fluctuations of the temperature were seen for some sensors such as sensors 8 and 10. This could be due to the poor contact between the sensors and the breast surface at that point of time. Similar temperature fluctuation behaviour was also seen in Figure 3.5 for patient #A116. Hence with such irregularities detected, the sets of temperature readings for patients #1001 and #A116 were removed from the data collection and cannot be used in the data analysis.

Figure 3.4: Drastic fluctuations of the temperature in several sensors for patient #1001

Figure 3.5: Drastic fluctuations of the temperature in several sensors of a typical subject.

Several solutions could be considered for the above two abnormalities [18]:

- i. Based on the graphical analysis, manually remove the temperature reading data that has the lowest temperature during data preparation
- ii. Enhance the sensors or equipment features during the first activity. Lifeline Biotechnologies, Inc, Florida, USA had placed an alarm on each sensor to detect abnormality during monitoring and data collection.
- iii. Develop a computer program that could detect temperature which was out of normal range and delete the corresponding data automatically. Ideally, use a high-end, real-time infrared imaging system that could provide temperature reading for each pixel which allows a better representation of the thermal profile on the breast surface.
- iv. Apply a regression approach to select the best appropriate input data for the classifiers.

Normalization of the temperature readings was required since each set of temperature readings had its own temperature range, depending on individual's health and body conditions. For instance, some patients had temperature ranges from 30°C to 35 °C, while others ranged from 32°C to 36°C. Therefore, to adjust for the different temperature ranges and to constrain the correlation into a manageable range, the normalized temperature readings were used instead of the actual readings. The temperature readings were normalized based on the following formula:

X / Y where X is the temperature variable and Y is the maximum temperature obtained from all the temperature readings

A new temperature scaling method $\left[\frac{(T_{\max} - T)}{(T_{\max} - T_{\min})} \right]$ will be adopted in Chapter 6

and it will be used in the analysis for Chapter 7 and 8. The new method was introduced as it is independent of the temperature scale used.

Chapter 4: Classifiers and Receiver Operating Characteristic (ROC)

In this study, six classifiers namely, back-propagation algorithm (BPA), probabilistic neural network (PNN), fuzzy (Sugeno-type), Gaussian mixture model (GMM), support vector machine (SVM) and fuzzy equivalence were used. The software for the classification of the different classes was written in MATLAB 7.0.4 for BPA, PNN, fuzzy, GMM and SVM. Fuzzy equivalence was coded in C and run on Microsoft Visual Studio 2008.

4.1: Back-Propagation Algorithm (BPA)

Back propagation was created by generalizing the Widrow-Hoff learning rule to a multiple layer network and non-linear differentiable transfer function. Input vectors and corresponding target vectors are used to train a network until it can approximate a function, associate input vectors with specific output vectors, or classify input vectors in an appropriate way as defined in this study. Networks with biases, a sigmoid layer and a linear output layer are capable of approximating any function with a finite number of discontinuities. The back propagation algorithm consists of two paths: the forward path and the backward path. The forward path includes creating a feed forward network, initializing weight, simulation and training the network. The network weights and biases are updated in the backward path.

The feed forward networks often have one or more hidden layers of sigmoid neurons followed by an output layer of linear neurons. Multiple layers of neurons with non-linear transfer functions allow the network to learn non-linear and linear

relationships between input and output vectors. The linear output layer allows the network to produce values outside the range -1 to +1.

Before training a feed forward network, the weight and biases must be initialized. We used random numbers around zero to initialize weights and biases in the network. The training process requires a set of proper inputs and targets as outputs. During training, the weights and biases of the network are iteratively adjusted to minimize the network performance function. The default performance functions for feed forward networks are the mean square errors, the average squared errors between the network outputs and the target output.

The weight update aims at maximizing the rate of error reduction, and, hence, it is termed as a 'gradient descent' algorithm [45]. The weight increment is done in 'small' steps; the step size is chosen heuristically, as there is no definite rule for its selection. In the present case, a learning constant $\eta = 0.9$ (which controls the step size) was chosen by trial and error.

The ANN structure used for the classification is shown in Figure 4.1. It consists of 16 nodes to accept the data, and the two hidden layers with 17 neurons process the data using the activation function. The output layer will give rise to an output of four possible classes, whereby the network is initially trained to identify these four classes given by decoded binary output [0001 0010 0100 1000].

Figure 4.1: Feed-forward classifier used for the 4 classifications.

4.2: Probabilistic Neural Network (PNN)

It works on the similar principal as that of K-nearest neighbor (K-NN) model. The basic idea is that, the predicated value of an item is likely to be about the same as other items that have close values of the predictor variables. The Figure 4.2 shows the diagram of the PNN.

Figure 4.2: Probabilistic neural network structure.

PNN has four layers [46]. They are as follows:

- (a) Input layer: The input data is fed to this layer. In our work, we have 16 input layers.
- (b) Hidden layer: this layer has one neuron for each training data set. In this layer, the neuron stores the values of the predictor variables along with the target value. This hidden neuron computes the Euclidean distance of the test case from the neuron's center point and then applies the RBF kernel function using the sigma value(s). The resulting value is passed to the neurons in the pattern layer.
- (c) Pattern (summation) layer: there is one pattern neuron for each class of the target variable. The actual target class of each training case is stored with each hidden neuron; the weighted value coming out of a hidden neuron is fed only to the pattern neuron that corresponds to the hidden neuron's category. The pattern neurons add the values for the class they represent (hence, it is a weighted vote for that category).
- (d) Decision layer: the decision layer compares the weighted votes for each target classes accumulated in the pattern layer and uses the largest vote to predict the target category.

4.3: Fuzzy (Sugeno-type)

In a fuzzy classification system, pattern space is divided into multiple subspaces. And for each subspace, the relationships between the target patterns and their classes are described by if-then type fuzzy rules [47]. The advantage of this system is that a nonlinear classification boundary can be easily implemented. Unknown patterns are classified by fuzzy inference, and patterns that belong to an unknown class which is not considered by learning, can be easily rejected. Ishibuchi et al [48] proposed methods using a simple learning procedure and a genetic algorithm (GA) to acquire a fuzzy classification system automatically. With these methods, however, a pattern space is a divided lattice-like structure. Therefore, many fuzzy rules corresponding to fine subspaces are required to implement a complicated classification boundary.

A Fuzzy classifier [49] using subtractive clustering and the Sugeno fuzzy inference system is implemented as shown in Figure 4.3. The algorithm for implementation is as followed:

- Step 1 (Fuzzify inputs)
 - The input is fuzzified using symmetric Gaussian membership function given by:

$$f(x; \sigma, \mu) = \frac{e^{-(x-\mu)^2}}{2\sigma^2}, \text{ where } \sigma \text{ and } \mu \text{ are variance and mean, respectively}$$
- Step 2 (Fuzzy inference)
 - Fuzzy inference is the process of formulating the mapping from a given input to an output using fuzzy logic for making decisions

- From the fuzzified inputs, the cluster centers are determined using subtractive clustering method
- In this method, the data point with the highest potential will be selected as the first cluster center
- All data points in the vicinity of the first cluster center (as determined by radii) are removed in order to determine the next data cluster and its center location
- This process is iterated until all of the data is within the radii of a cluster center
- Step 3 (Obtaining the membership function (MF) computation)

Figure 4.3: Block diagram of Sugeno rule operation.

Assign the inputs as: Input 1 = 'x' and Input 2 = 'y' and the Output will be $z = ax + by + c$. The output level 'z' is a constant ($a=b=0$) for a zero-order Sugeno model. The output level 'zi' of each rule is weighted by the firing strength 'wi' of the rule. Input 1 = 'x' and Input 2 = 'y', the firing strength is $w_i = \text{AndMethod}(F1(x), F2(y))$. The final output of the system is the weighted average of all rule outputs as computed below:

$$\text{Final Output} = \frac{\sum_{i=1}^N w_i z_i}{\sum_{i=1}^N w_i}, \quad \text{where } N = \text{the total number of fuzzy rules}$$

(1)

4.4: Gaussian Mixture Model (GMM)

A Gaussian Mixture Model (GMM) is a parametric model used to estimate a continuous probability density function from a set of multi-dimensional feature observations. It is widely used in data mining, pattern recognition, machine learning and statistical analysis. This Gaussian mixture distribution can be described as a linear superposition of K multidimensional Gaussian components given by:

$$p(x) = \sum_{k=1}^K \pi_k N(x | \mu_k, \Sigma_k)$$

(2)

Where π_k, μ_k, Σ_k are mixing coefficients, mean and covariance respectively.

The solution for determining the parameters of GMM is estimated by using the maximum likelihood (ML) criterion. A powerful method for maximizing the likelihood solution models is by the general form of Expectation-Maximization (EM) algorithm. The steps to carry out the EM algorithm are as followed:

- i. Initialize the means μ_k , covariances Σ_k and mixing coefficients π_k , and evaluate the initial value of the log likelihood.
- ii. E step: Evaluate the responsibilities using the current parameter values

$$\gamma(z_{nk}) = \frac{\pi_k N(x_n | \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j N(x_n | \mu_j, \Sigma_j)} \quad (3)$$

- iii. M step: Re-estimate the parameters using the current responsibilities

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) x_n \quad (4)$$

$$\sum_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk})(x_n - \mu_k^{new})(x_n - \mu_k^{new})^T$$

$$\pi_k^{new} = \frac{N_k}{N}$$

$$\text{Where } N_k = \sum_{n=1}^N \gamma(z_{nk})$$

- iv. Evaluate the log likelihood and check for convergence of either the parameters or the log likelihood. If the convergence criterion is not satisfied return to Step (ii).

$$\ln p(X | \mu, \Sigma, \pi) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K \pi_k N(x_n | \mu_k, \Sigma_k) \right\} \quad (5)$$

However in EM algorithm, it takes more iteration to reach convergence compared with the K-means algorithm and each cycle needs more computations. Hence, it is common to use the K-means algorithm to find the initial estimates of the parameters obtained from a sample of the training data. The K-means algorithm uses the squared Euclidean distance as the measure of dissimilarity between a data point and a prototype vector. This not only limits the type of data variables to be considered but also makes the determination of the cluster means non-robust to the outliers [50-52]. This algorithm starts off by choosing randomly the initial means and assumed unit variances for the diagonal covariance matrix which is being adopted in the current work.

4.5: Support Vector Machine (SVM)

In the recent years, Support Vector Machine (SVM) classifiers have demonstrated excellent performance in a variety of pattern recognition problems [53-55]. The methodology is described in details by Vapnik [54] and in the tutorial by Burges [53].

The SVM is a supervised learning method that generates input-output mapping functions from a set of labeled training data. The mapping function can be either a classification function or a regression function. For classification, non-linear kernel functions are often used to transform input data to a high-dimensional feature space in which the input data become more separable compared to the original input space. Maximum-margin hyper-planes are then created; hence, the model produced depends on only a subset of the training data near the class boundaries [56]. This classification method is currently adopted in the current work.

The aim of SVM modeling is to find a separating hyperplane which separates positive and negative examples from each other with optimal margin; in other words, the distance of the decision surface and the closest example is maximal as shown in Figure 4.4. Essentially, this involves orienting the separating hyperplane to be perpendicular to the shortest line separating the convex hulls of the training data for each class, and locating it midway along this line. The vectors that constrain the width of the margin are the support vectors.

Figure 4.4: Working of SVM algorithm.

Let the separating hyperplane be defined by $x \cdot w + b = 0$, where w is its normal. For linearly separable data labeled $\{x_i, y_i\}$, $x_i \in \mathbb{R}^{N_d}$, $y_i = \{-1, 1\}$, $i = 1, \dots, N$, the optimum boundary chosen with maximal margin criterion is found by minimizing the objective function:

$$E = \|w\|^2$$

Subject to $(x_i \cdot w + b)y_i \geq 1$, for all i .

(6)

The solution for the optimum boundary w_0 is a linear combination of a subset of the training data, $s \in \{1 \dots N\}$: the *support vectors*. These support vectors define the margin edges and satisfy the equality $(x_s \cdot w_0 + b)y_s = 1$. Data may be classified by computing the sign of $x \cdot w_0 + b$.

Generally, the data are not separable, and the inequality in the equation (6) cannot be satisfied. In this case, a “slack” variable ξ_i that represents the amount by which each point is misclassified is introduced. The new objective function is now reformulated as

$$E = \frac{1}{2} \|w\|^2 + C \sum_i L(\xi_i)$$

Subject to $(x_i \cdot w + b)y_i \geq 1 - \xi_i$ for all i . (7)

The second term on the right-hand side of equation (7) is the empirical risk associated with those points that are misclassified or lie within the margin. L is a cost function and C is a hyper-parameter that trades-off the effects of minimizing the empirical risk against maximizing the margin. The first term can be thought as a regularization term, deriving from maximizing the margin which gives the SVM its ability to generalize well on sparse training data.

Kernel functions can be used to extend to solution of non-linear boundaries problems [57]. Kernel functions define a non-linear mapping from the input space (observed data) to a manifold in higher dimensional feature space, which is defined implicitly by the kernel functions. The hyperplane is constructed in the feature space and intersects with the manifold, creating a non-linear boundary in the input space. In practice, the mapping is achieved by replacing the value of the dot products between two vectors in the input space with the value that results when the same dot product is carried out in the feature space. The dot product in the

feature space is expressed by functions (i.e., the kernels) of two vectors in input space. The polynomial and radial basis function (RBF) kernels are commonly used, and they are

$$K(x_i, x_j) = (x_i \cdot x_j + 1)^n \quad (8)$$

and

$$K(x_i, x_j) = \exp \left[-\frac{1}{2} \left(\frac{\|x_i - x_j\|}{\sigma} \right)^2 \right] \quad (9)$$

Respectively, where n is the order of the polynomial and σ is the width of the RBF, the dual for the nonlinear case is given by:

$$\alpha^* = \max_{\alpha} \left(\sum_i \alpha_i + \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j) \right) \quad (10)$$

$$\text{Subject to } \begin{cases} 0 \leq \alpha_i \leq C \\ \sum_i \alpha_i y_i = 0 \end{cases}$$

With the above formulation on the use of kernels, an explicit transformation of the data to the feature space is not required. Several algorithms extend the basic binary SVM classifier to be a multi-class classifier. Examples consist of one-against-one SVM, one-against-all SVM [58, 59], half against half SVM [60] and Directed Acyclic Graph SVM (DAGSVM) [61].

4.6: Fuzzy Equivalence

In fuzzy based prediction, it is necessary to make database queries to extract data required from the databases [62]. Also it might be required to classify unknown data into the predefined classes. In such cases, the fuzzy relations can be used effectively to establish a relationship between the unknown and known data. The classification of unknown data can then be done depending on its closeness to the known data. The fuzzy equivalence relation gives the relationship among the data, although the formation of this relationship cannot usually be done directly, a fuzzy compatibility relation (reflexive and symmetric) can be readily determined in terms of an appropriate distance function applied to given data. Then, a meaningful fuzzy equivalence relation is defined as the transitive closure of this fuzzy compatibility relation.

4.6.1 Algorithm Prediction using Fuzzy Equivalence Relations

The unknown data which is to be predicted is first appended onto the data whose classes are already known. The entire data set is then normalized to maintain the uniformity in the data values. The normalization is done attribute wise, the maximum in each attribute (column) is chosen and the column values are divided by this maximum value.

From the second step, the compatibility relation is formed. Given that the set of data X have p -tuples of R^p . A fuzzy compatibility relation, R , on x be defined in terms of an appropriate distance function of the Minkowski class given the formula as:

$$R(x_i, x_k) = 1 - \delta \left(\sum_{j=1}^p |x_{ij} - x_{kj}|^q \right)^{1/q} \quad (11)$$

for all pairs $(x_i, x_k) \in X$, where $q \in R^+$, and δ is a constant which ensures that $R(x_i, x_k) \in [0, 1]$. δ is usually taken as inverse value of the largest distance in X .

In the algorithm being considered, the value of q is taken to be equal to two. Thus, the Euclidean distance is used. If $q=1$ is used, then the distance is hamming distance. For the value of d , since the data values are normalized, it is adjusted such that the value resulting from the equation always lies between zero and one. In this algorithm the value is taken to be equal to $1/p^{1/2}$. The result of the above step is a matrix of the size of the data set where the values give a degree of relation between the data.

The fuzzy compatibility relation defined above is not necessarily an equivalence relation. Hence, the transitive closure of R is determined to get the corresponding fuzzy equivalence relation. The transitive closure is determined using the algorithm given as follows:

Step 1: $R' = R \cup (R \circ R)$.

Step 2: If $R' \neq R$, make $R = R'$ and go to Step 1.

Step 3: Stop $R' = R_T$.

The above steps result in a fuzzy equivalence relation being formed between the data; this gives a diagonal matrix where the values denote a degree of relationship between the data. The unknown data as said earlier are appended onto the known data; hence the classes of the unknown data can be obtained by looking at the membership degree of each of these data with the classified data. The class to which they belong is given by the data with which the degree of membership is the highest. Figure 4.5 shows the flowchart for prediction using fuzzy equivalence

relations and Table 4.1 provides a detailed algorithm for prediction using fuzzy equivalence relations.

Figure 4.5: Flowchart for prediction using fuzzy equivalence relations

Table 4.1: Detailed Algorithm for Prediction using Fuzzy Equivalence Relations

Step	Phase	Description
1	Initialization	<p>1.1 Read classified data from an input file</p> $[input_data]_{i,j} [classes]_k \leftarrow Buffer$ <p>i: 1 to m (number of data) j: 1 to n (number of attributes) k: 1 to i; where there are 'o' classes</p> <p>1.2 Read unclassified data from test file</p> $[unclassified]_{i,j} \leftarrow Buffer$ <p>i: 1 to p (number of data) j: 1 to n (number of attributes)</p>
2	Pre-processing	<p>2.1 Append unclassified_data onto input_data</p> $[input_data]_{l,j} \leftarrow [input_data]_{i,j} \Downarrow [unclassified_data]_{k,j}$ <p>l: 1 to $(m+p)$ i: $m+1$ to $m+p$ k: 1 to p j: 1 to n</p> <p>2.2 Normalize the data matrix attribute wise selecting the maximum in each attribute</p> $[input_data]_{l,j} \leftarrow \frac{[input_data]_{l,j}}{\max \{ [input_data] \}_j}$ <p>l: 1 to $(m+p)$ j: 1 to n</p>

3	Compute the Fuzzy Equivalence Relation	<p>3.1 Find the Compatibility relation between the data using distance function of Minkowski class</p> $R_c(x_i, x_j) = 1 - \delta \left(\sum_{l=1}^n (x_{il} - x_{jl})^2 \right)^{1/2}$ <p>$i, j: 1 \text{ to } m+p$ Where $\delta = 1/n^{1/2}$</p> <p>3.2 Find the transitive closure of R_c using algorithm given below</p> <p>3.2.1 $R' = R_c \cup (R_c \bullet R_c)$</p> <p>$\cup$ = Max operator $\bullet$ = Max-Min composition Where $R \bullet R = r_{ij} = \max_k \min(r_{ik}, r_{kj})$ $i, j, k: 1 \text{ to } m+p$</p> <p>3.2.2 If $R' = R_c$ then Stop. Else make $R_c = R'$ goto step 3.2.1.</p> <p>3.2.3 $R_t = R'$; R_t is the Transitive closure matrix $(m+p) \times (m+p)$</p>
4	Prediction	<p>Beginning from row $m+1$ search for the maximum membership degree until column m and store the corresponding index in matrix Max_membership</p> $[Max_membership]_i = \max [R_t]_{i,j};$ <p>$i: m+1 \text{ to } m+p$ $j: 1 \text{ to } m$</p> <p>corresponding to each unclassified data the class to which it belongs is given by $[classes]_{[Max_membership]_i}; i: 1 \text{ to } p$</p>

4.7: Comparison on the advantages and disadvantages of the classifiers

In this section, the advantages and disadvantages of the six classifiers will be discussed in Table 4.2

Table 4.2: Comparison of the 6 classifiers

Classifiers	Advantages	Disadvantages
Back-Propagation Algorithm (BPA)	<ul style="list-style-type: none">• Good at prediction and classification• Easy to implement• Applicable for a wide range of data	<ul style="list-style-type: none">• Uncertain in mapping of the hidden layers• Time consuming due to slow learning time• Prone to over-fitting
Probabilistic Neural Network (PNN)	<ul style="list-style-type: none">• Higher accuracy• Fast training process• A parallel structure guaranteed to converge to an optimal classifier as the size of the training set increases<ul style="list-style-type: none">○ Addition or removal of training samples do not require extensive retraining [63]	<ul style="list-style-type: none">• Large memory requirements [63]• Slow execution of network• Requires a representative training set

Fuzzy (Sugeno-type)	<ul style="list-style-type: none">• Provide the fastest learning and interpretation of the parameters	<ul style="list-style-type: none">• Harder hardware implementation
Gaussian Mixture Model (GMM)	<ul style="list-style-type: none">• Able to form smooth approximations for any arbitrarily-shaped densities• Time and complexity can be reduced significantly	<ul style="list-style-type: none">• Many parameters required to fit, thus making learning difficult
Support Vector Machine (SVM)	<ul style="list-style-type: none">• Control model complexity through a regularization term [64]<ul style="list-style-type: none">○ Less prone to over-fitting○ Robust to noise• Simple geometric interpretation• Induce nonlinear features through a nonlinear kernel function<ul style="list-style-type: none">○ Suitable for complicated classification tasks	<ul style="list-style-type: none">• High algorithmic complexity<ul style="list-style-type: none">○ Sensitive to the choice of variable settings• Extensive memory requirements of the required quadratic programming in large-scale tasks• Difficulty in the choice of kernel functions

Fuzzy Equivalence	<ul style="list-style-type: none">• Need not derive rules for segmenting the data• Unknown data is predicted into known classes based on its similarity	<ul style="list-style-type: none">• Time consuming due to slow learning and executing time
------------------------------	--	--

4.8: Statistical analyses using Receiver Operating Characteristic (ROC)

The ROC curve was a plot of sensitivity against (1- specificity). Sensitivity, also known as true positive fraction (TPF) refers to the probability that a test result was positive when the disease was present [65]. As for specificity (true negative fraction, TNF), it was the probability that a test result was negative when the disease was not present. The formula and definitions of these parameters were summarized in Tables 4.3 and 4.4.

Table 4.3: Summary of definitions for true positive, false positive, false negative and true negative

	Disease				
	Present	n	Absent	n	Total
Positive	True positive, TP	a	False positive, FP	c	a + c
Negative	False negative, FN	b	True negative, TN	d	b + d
Total		a + b		c + d	

The parameters mentioned in Table 4.3 were used for calculating sensitivity, specificity, positive/negative likelihood ratio and predictive value. The definitions for these parameters were [65]:

- True positive (TP): The number of cases with disease present were correctly classified as positive

- False positive (FP): The number of cases without disease present were classified as positive
- True negative (TN): The number of cases without disease present were correctly classified as negative
- False negative (FN): The number of cases with disease present were classified as negative

Table 4.4: Summary of formula for sensitivity, specificity, positive /negative likelihood ratio and predictive value

Sensitivity	$a / (a + b)$	Specificity	$d / (c + d)$
Positive Likelihood Ratio (+LR)	$\text{Sensitivity} / (1 - \text{Specificity})$	Negative Likelihood Ratio (-LR)	$(1 - \text{Sensitivity}) / \text{Specificity}$
Positive Predictive Value (+PV)	$a / (a + c)$	Negative Predictive Value (-PV)	$d / (b + d)$

The area under the ROC curve was an important parameter as it measured the performance of the classifier across the entire range of cut-off points. In short, it determined the overall test accuracy. Conventionally, the area under the ROC curve must range between 0.5 and 1 [66]. If the area was closer to 1, this showed that the classifier had better accuracy in the testing. Currently, the area under the ROC curve was the best indicator for the classifier’s performance, as compared to the misclassification rate and the measure of risk based on confusion and loss matrices [67]. This was because it was able to provide the most complete way of quantifying the diagnostic accuracy.

Chapter 5: Comparative performance of the classifiers

5.1: Comparison studies on the six classifiers

Six classifiers namely; back-propagation algorithm (BPA), probabilistic neural network (PNN), fuzzy (Sugeno-type), Gaussian mixture model (GMM), support vector machine (SVM) and fuzzy equivalence were used to classify the four different diagnoses (normal, benign, cancer and suspected-cancer) of breast cancer. As mentioned by Joanna [20], five of the classifiers (BPA, PNN, Fuzzy (Sugeno-type), GMM and SVM) were able to achieve more than 80% of accuracy. GMM was able to produce the highest percentage of correct classification for unknown data and sensitivity. SVM was considered as an excellent classifier as it had the greatest specificity, positive predictive value and largest area under the curve which implies its accuracy.

In this section, Fuzzy Equivalence will be introduced as the sixth classifier to further enhance on the development of the interpretive software system. Fuzzy Equivalence was tested with the same sets of temperature readings used in [20] and its performance will be compared with the rest of the classifiers which had earlier conducted their tests as mentioned in [20].

There were a total of 6000 sets of temperature readings, and they were further sorted into two groups; 5000 sets of training data and 1000 sets of testing data. The 6000 sets of temperature readings were collected from four different diagnoses and in each diagnosis, 1500 sets of temperature data were randomly collected from different patients belonging to the same group. Each set of the temperature data consists of sixteen temperature measurements collected concurrently by the sixteen sensors over the test period. It was necessary to have more training data than testing data to allow for better training of the classifiers. Table 5.1 shows the

number of training and testing data used in comparing the performance of the six classifiers.

Table 5.1: Number of training and testing data used for classification

Diagnosis	Number of training data	Number of testing data
Normal	1250	250
Benign	1250	250
Cancer	1250	250
Suspected-Cancer	1250	250
Total	5000	1000

5.1.1: Results of the six classifiers

The performance of the six classifiers is shown in Table 5.2. Similarly to the previous test in [20], GMM had once again obtained the highest percentage of correct classification of 90.6% even after Fuzzy Equivalence was introduced. This was followed by Fuzzy Equivalence which had attained an accuracy of 87.1%. As shown in Table 5.3, both GMM and Fuzzy Equivalence had also achieved at least 90% of sensitivity as compared to the rest of the classifiers. They had sensitivity of 94.8% and 92.2% respectively.

Table 5.3 shows that SVM had the highest specificity of 90.4%. GMM and Fuzzy Equivalence had achieved low specificity. The overall specificities for all the classifiers were also relatively low and this means that fewer true negative cases were correctly classified. This could be due to noise data present in the normal diagnosis as specificity determines the probability that a test result was negative when the disease was not present.

In Figure 5.1, it shows that SVM (pink dotted line) had again the largest area under the ROC curve and this was accountable from Table 5.3 as it had obtained the largest area of 0.872. Fuzzy Equivalence may not seem to perform as well as GMM and SVM but it still has the ability of performing well in classifying the unknown data. Moreover, more tests are required to verify this result. Hence, in the next section, the three classifiers namely, GMM, SVM and Fuzzy Equivalence will undergo one more test to determine which classifier can perform the best. In this test, there will be an increased in the number of temperature data used and statistical analysis using the receiver operating characteristic (ROC) will be applied.

Table 5.2: Comparison on the percentage of correct classification for the classifiers

Type of classifiers	No of training data used	No of testing data used	Percentage (%) of correct classification
BPA	5000	1000	83.1
PNN	5000	1000	86.1
Fuzzy	5000	1000	77.4
GMM	5000	1000	90.6
SVM	5000	1000	85.6
Fuzzy Equivalence	5000	1000	87.1
Average			85.0

Table 5.3: Results of Performance Indices for the six classifiers used

Classifier	TN	TP	FP	FN	Sensitivity (%)	Specificity (%)	Area under the curve	PPV
BPA	209	622	41	128	82.9	83.6	0.833	93.8
PNN	195	666	55	84	88.8	78	0.834	92.4
Fuzzy	189	585	61	165	78	75.6	0.768	90.6
GMM	195	711	55	39	94.8	78	0.864	92.8
SVM	226	630	24	120	84	90.4	0.872	96.3
Fuzzy Equivalence	193	678	57	72	90.4	77.2	0.838	92.2

Figure 5.1: Comparison of ROC curves for the 6 different classifiers

5.2: Performance comparison for the three classifiers.

Previously, it was mentioned that GMM, SVM and fuzzy equivalence have the potential of becoming the best classifier. In this section, the number of training and testing data used were increased to 7500 and 1500 respectively. These increments are to further determine which classifier (GMM, SVM or fuzzy equivalence) can perform better. Table 5.4 shows the number of training data and testing data used in this test.

Table 5.4: Number of training and testing data used

Diagnosis	Number of training data	Number of testing data
Normal	1875	375
Benign	1875	375
Cancer	1875	375
Suspected-cancer	1875	375
Total	7500	1500

Table 5.5: Comparison on the percentage of correct classification for the three classifiers

Type of classifiers	No of training data used	No of testing data used	Percentage (%) of correct classification
GMM	7500	1500	99.4
SVM	7500	1500	76.8
Fuzzy Equivalence	7500	1500	100
Average			92.1

The average percentage of correct classification for the three classifiers was 92.1% as tabulated in Table 5.5. The result showed that Fuzzy Equivalence had the best performance as it had achieved 100% of accuracy in classifying the unknown data. This was followed by GMM which had obtained 99.4% of correct classification. SVM did not perform as well as the other two classifiers; it only managed to attain 76.8% of accuracy.

However, this cannot simply determine that Fuzzy Equivalence was the best and most accurate classifier. Receiver operating characteristic (ROC) analysis will be applied in the next section to conclude on the performance of the three classifiers.

5.2.1: Receiver Operating Characteristic (ROC) Analysis

The MedCalc statistical software [65] was used for the analyses. The ROC results based on the sensitivity, specificity, positive predictive value and area under curve for the three classifiers were tabulated in Table 5.6.

Table 5.6: Results of ROC for the three classifiers

Classifier	TN	TP	FN	FP	Sensitivity (%)	Specificity (%)	Area under the curve	+PV
GMM	374	1117	8	1	99.3	99.7	0.995	99.9
SVM	318	834	291	57	74.1	84.4	0.795	93.6
Fuzzy Equivalence	375	1125	0	0	100	100	1.00	100

The results obtained from the 1500 testing data were classified under true negative (TN), true positive (TP), false positive (FP) and false negative (FN), depending on each classifier’s situation. Based on Table 5.6 using ROC analysis, Fuzzy Equivalence had the highest sensitivity of 100%, and this shows that it had correctly classified all the true positive cases. GMM and SVM had sensitivity of 99.3% and 74.1% respectively. Hence, it is concluded that Fuzzy Equivalence is the most sensitive classifier.

Fuzzy Equivalence showed the highest specificity of 100% and this was justified by the number of true negative cases. It had the greatest number data classified under true negative cases as compared to the rest of the classifiers, thus explained its highest percentage in specificity. This result was followed by GMM with specificity of 99.7% and lastly, SVM with specificity of 84.4%. In tabulating the positive predictive value, Fuzzy Equivalence classifier showed the highest value of 100, followed by GMM with value of 99.9 The PPV value for SVM was 93.6.

The area under the curve is also an important parameter as it determines the overall classification accuracy for the three classifiers. Figure 5.2 illustrated the comparison of the ROC curve for the three classifiers. The Fuzzy Equivalence (red line) had the largest area under the curve. It had obtained the maximum area for the ROC curve as Table 5.6 showed that it had an area of 1.00. SVM (maroon dotted line) had the smallest area among the three classifiers. This result was reinforced based on Table 5.6 as it showed that SVM had an area of 0.795 while GMM had an area of 0.995. Thus, using the ROC concepts, we can conclude that Fuzzy Equivalence works better than the rest of the classifiers.

Figure 5.2: Comparison of ROC curves for the three different classifiers

With the increment of the temperature data, Fuzzy Equivalence seems to perform much better than the other two classifiers. It had obtained the highest percentage of correct classification in Chapter 5.2. In this statistic analysis of ROC curves, Fuzzy Equivalence was considered the outstanding classifier. This result was based on the four performance indices, namely sensitivity, specificity, positive predictive value and area under curve, in which Fuzzy Equivalence had attained the best result in all the indices. It had the greatest sensitivity, specificity and positive predictive value and had also attained the largest area under the curve which implies its accuracy. Therefore, Fuzzy Equivalence was considered to be an excellent classifier.

Chapter 6: Improvement on the methods of temperature scaling.

Previously in Chapter 3, the temperature readings were normalized based on the following formula (X / Y), where X is the temperature variable and Y is the maximum temperature obtained from all the temperature readings. This will cause the results to be dependent on the temperature scale used. Hence, several methods of temperature normalization are explored in this Chapter to determine which method is the most suitable for our application. A total of four different methods on normalization will be compared to determine the most suitable method for temperature scaling.

The temperature scaling is carried out by obtaining the temperature data from one normal patient and one cancer patient. In each sets of temperature data, there are sixteen columns of temperature measurements which represent the sixteen sensors used to measure the temperature. In this analysis, we will only use the temperature data from the first sensor to determine the best appropriate method. Graphs will also be plotted to find out how the temperature data behave in different methods of normalization.

The four methods and their parameters are evaluated as followed:

- Method 1: $abs\left[\frac{(T - T_{ave})}{(T_{max} - T_{ave})}\right]$
- Method 2: Z-score : $abs\left[\frac{(T - T_{ave})}{S_{dev}}\right]$
- Method 3: $\left[\frac{(T - T_{min})}{(T_{max} - T_{min})}\right]$

- Method 4: $\left[\frac{(T_{\max} - T)}{(T_{\max} - T_{\min})} \right]$

T is the temperature variable, T_{ave} is the average temperature of the range of temperature data, $T_{\max}$ is the maximum temperature of the temperature range, $T_{\min}$ is the minimum temperature and $Stdev$ is the standard deviation of the temperature data.

Figure 6.1 shows how the temperature data of a normal patient and a cancer patient behave in the first sensor. This is done before proceeding with normalization. The next Figure 6.2 shows the normalized data from the conversion of the temperature data using method 1: $abs \left[\frac{(T - T_{ave})}{(T_{\max} - T_{ave})} \right]$. Similarly, Figures 6.3, 6.4 and 6.5 show the conversions of the temperature data for method 2, 3 and 4 respectively.

Figure 6.1: Temperature data in sensor 1 for the normal and cancer patient

Based on Figures 6.2 and 6.3, we can see that the normalization ranged from 0 to 3.5 and 0 to 5.5 respectively. These two methods will be rejected as ANN can only perform within the range of 0 to 1. Thus, it is impossible to carry out ANN classification with such wide ranges of temperature data. In Figures 6.4 and 6.5, the temperature ranges from 0 to 1, and this shows that these two methods are suitable for the ANN testing.

Figure 6.2 shows the temperature scaling of method 1: $abs \left[\frac{(T - T_{ave})}{(T_{max} - T_{ave})} \right]$

Figure 6.3 shows the temperature scaling of method 2: $abs\left[\frac{(T - T_{ave})}{S_{dev}}\right]$

Figure 6.4 shows the temperature scaling of method 3: $\left[\frac{(T - T_{min})}{(T_{max} - T_{min})}\right]$

Figure 6.5 shows the temperature scaling of method 4: $\left[\frac{(T_{\max} - T)}{(T_{\max} - T_{\min})} \right]$

The temperature data normalized using methods 3 and 4 will be used as inputs fed into the classifiers for classifying the unknown data into the four different diagnoses (normal, benign, cancer and suspected-cancer). The classifiers used are PNN, SVM and Fuzzy Equivalence. 60 sets of temperature data were randomly collected from the patients of the four different diagnoses (normal, benign, cancer and suspected-cancer). These data were further compiled into 200 training data and 40 testing data used to carry out the classification test.

Table 6.1: Tabulated results for methods 3 and 4

Classifier	PNN	SVM	Fuzzy Equivalence
Accuracy of classifier using Method 3: $\left[\frac{(T - T_{\min})}{(T_{\max} - T_{\min})} \right]$	50%	75%	27%
Accuracy of classifier using Method 4: $\left[\frac{(T_{\max} - T)}{(T_{\max} - T_{\min})} \right]$	50%	75%	27%

From Table 6.1, all the classifiers had obtained the same accuracy regardless of whether method 3 or 4 is being used and this is because the parameters of these two methods are complement of one another. However, method 4: $\left[\frac{(T_{\max} - T)}{(T_{\max} - T_{\min})} \right]$ will be chosen to normalize the temperature data [68] as we are locating for higher temperature data which likely indicates the presence of breast tumours.

6.1: Testing on the classifiers using the new normalization

In this analysis, 8000 training data and 1600 testing data were used to compare on the performance of the five classifiers. Fuzzy is not included in this analysis as it had performed badly using the new method of normalization. Furthermore, the performance of Fuzzy in Chapter 5 was also poor; hence, Fuzzy was omitted from this current study. The new normalization method $\left[\frac{(T_{\max} - T)}{(T_{\max} - T_{\min})} \right]$ will be adopted

in this analysis. Table 6.2 shows the performance of the five classifiers. All the five classifiers had performed relatively well in the new normalization method $\left(\left[\frac{(T_{\max} - T)}{(T_{\max} - T_{\min})} \right] \right)$, obtaining an average accuracy of 83.1%.

GMM had the best performance and had obtained the highest percentage of correct classification of 90.3%. PNN, Fuzzy Equivalence and SVM obtained 85.3%, 82% and 77.4% respectively. BPA classifier was trained under 2 hidden layers of 25 neurons and the 8000 training data were successfully trained after going through 72 learning cycles to meet a goal of 0.000860206. However, the BPA classifier was only able to classify the unknown data correctly with an accuracy of 80.4%.

The analysis showed that the performance of the classifiers was not greatly affected by the new normalization method. It was more appropriate for normalization as the results are independent of the temperature scale used. Hence, the formula: $\left[\frac{(T_{\max} - T)}{(T_{\max} - T_{\min})} \right]$ will be adopted as the new temperature scaling for the remaining of the analysis as shown in the next two chapters.

Table 6.2: Comparison on the percentage of correct classification for the 5 classifiers

Type of classifiers	No of training data used	No of testing data used	Percentage (%) of correct classification
			Using $\left[\frac{(T_{\max} - T)}{(T_{\max} - T_{\min})} \right]$ for temperature scaling
BPA	8000	1600	80.4
PNN	8000	1600	85.3
Fuzzy Equivalence	8000	1600	82.0
GMM	8000	1600	90.3
SVM	8000	1600	77.4
AVERAGE			83.1

Chapter 7: Processing data in various time sequences

7.1: Classification of patients for various time sequences

Lifeline's equipment (First Warning TM) is a non-invasive, painless and easy to wear system which consists of sixteen sensors that are connected to a monitor (of a size of a pager) attached to the bra. It is designed to be worn under normal clothing and most importantly, it allows the woman to maintain her normal daily life. However, as mentioned in Chapter 3, the patients had undergone 48 hours of temperature recording using the Lifeline's equipment (First Warning TM) [21].

The long hours of temperature recording may deter women from wanting to use the system for early detection of breast cancer as it can be tiring and provide unnecessary inconvenience for them even though it was specified that they were still able to maintain normal daily life. Furthermore, the accuracy of recording may be affected due to problems such as loosen of the sensors. Thus, the aim of this chapter is to find out whether the accuracy in classifying the diagnoses will be affected when the duration of temperature recording is shortened. The four classifiers (back-propagation algorithm (BPA), probabilistic neural network (PNN), Gaussian mixture model (GMM) and fuzzy equivalence) which had performed well in the previous chapter will be used in this study to determine how their accuracies, sensitivities and specificities will be affected when unknown data obtained from shorter temperature recording duration is being used for classification.

As a start to this study, we will use the minimal time duration to find out how can the accuracy be affected with minimum temperature data set used. Thus, the 48

hours temperature data provided for us previously will be processed into various time sequences ranging from 10 minutes to 25 minutes.

Before commencing the study, there are several factors to be considered:

1. Ensure that all the patients chosen for this study have undergone constant temperature recording of 48 hours.
2. Ensure that the temperature is recorded with a time interval of approximately 5 minutes.
3. Instead of randomly selecting the temperature data used for classification; a suitable time frame is selected for the various time sequences. Since circadian thermal activity usually takes place from 6pm to 12am. Thus, the time frame for the various time sequences are:
 - a. 10 minutes (6pm-6.10pm)
 - b. 15 minutes (6pm-6.15pm)
 - c. 20 minutes (6pm-6.20pm)
 - d. 25 minutes (6pm-6.25pm)

In Chapter 3, we stated that 67 patients were being selected for the software analytical test after eliminating those with abnormalities. However in this section, only 49 patients are chosen for this study and this is because those eliminated have more than 48 hours of temperature recording. This information could be found in the spreadsheet provided by Lifeline Biotechnologies, Inc, Florida, USA [21] and it shows that the start and end dates for these patients are more than two days. Therefore, in order to be consistent with our data collection, such patients are being omitted from this study.

The number of patients used for the four classes of diagnoses is shown in Table 7.1. The patients are randomly categorized into training and testing which would be used for compiling of training and testing data. The number of patients used will remain the same regardless of which time sequence is being studied. The selected time frame for the various time sequences is from 6pm to 12am and this is due to the circadian thermal activity which usually takes place during this period of time.

Table 7.1: Number of patients used for the various time sequence study

Diagnosis	Number of Patients		Patients' Identification
	Training	Testing	
Normal	9	3	#178, #163, #88, #120, #144, #145, #169, #149, #150, #151, #152, #155,
Benign	9	3	#125, #1011, #56, #92, #95, #117, #1002, #1003, #1013, #47, #A103, #A107
Cancer	9	3	#A108, #22, #42, #51, #96, #130, #A105, #A114, #A121, #A127, #A115, #A117
Suspected-cancer	9	3	#15, #20, #23, #26, #27, #34, #39, #65, #59, #86, #106, #1015
Total number of patients	48		

Table 7.2 Number of training data and testing data used for various time sequence study

	Number of training data	Number of testing data
Diagnosis / Time duration (minutes)	10, 15, 20, 25	10,15,20,25
Normal	36	12
Benign	36	12
Cancer	36	12
Suspected-cancer	36	12
Total	144	48

The number of training and testing data used for the various time sequences were tabulated in Table 7.2. A total of 144 training data and 48 testing data were compiled for the analysis. Due to consistency, same number of data was used for the all the different time durations. Also, there are more training data than the testing data as this will allowed the classifiers to train better and obtained higher accuracy in classification. Once compilation was done, these data will be fed into the four classifiers namely BPA, PNN, GMM and Fuzzy Equivalence to classify the unknown data into the four different diagnoses. Results obtained by the four classifiers will be further elaborated in the following sections.

7.2: Classifiers' performance in the 10 minutes analysis

In this analysis, an average of 49.9% of accuracy was obtained by the four classifiers. Table 7.3 shows the result obtained by the 4 classifiers. PNN had the highest percentage of correct classification of 56.2% while GMM had obtained the lowest accuracy of 43.7%. The percentage of sensitivity and specificity of the classifiers were also tabulated in the table. The results obtained from the 48 testing data were classified under true negative (TN), true positive (TP), false positive (FP) and false negative (FN), depending on each classifier's situation.

Both PNN and BPA showed the highest sensitivity of 52.8%. This was followed by Fuzzy Equivalence which had attained 36.1% of sensitivity and GMM had the lowest sensitivity of 30.6%. This observation showed that the higher sensitivity of the classifier would result in a greater detection rate by causing the false negative rate to be lower.

Both GMM and Fuzzy Equivalence had the highest specificity of 83.3% and this was justified by the number of true negative cases. There are a total of 12 normal test data and the classifiers had correctly classified 10 of these data into true negative cases. Thus, enable them to achieve 83.3% of specificity. This result was followed by PNN and BPA with specificity of 66.7% and 50% respectively.

Table 7.3: Result of the 4 classifiers in the 10 minutes analysis

Classifiers	Percentage of correct classification (%)	TN	TP	FN	FP	Sensitivity (%)	Specificity (%)
BPA	52.1	6	19	17	6	52.8	50
PNN	56.2	8	19	17	4	52.8	66.7
GMM	43.7	10	11	27	2	30.6	83.3
Fuzzy Equivalence	47.9	10	13	23	2	36.1	83.3
Average	49.9					43.1	70.8

7.3: Classifiers’ performance in the 15 minutes analysis

In Table 7.4, it showed that all the four classifiers had obtained an average accuracy of 52%. GMM and Fuzzy Equivalence had the worst performance among the four classifiers. They had only classified 43.7% of the test data correctly, and conversely, PNN had achieved the highest accuracy of 64.5%.

The sensitivity and specificity analyses showed that all the classifiers had an average of 56.9% and 39.6% respectively. PNN had also attained the highest sensitivity of 72.2% and this was further reinforced by the number of true positive data which it had correctly classified. Although Fuzzy Equivalence was neither the most accurate nor the most sensitive classifier, it had also obtained the highest specificity of 75%.

Table 7.4: Results of the 4 classifiers in the 15 minutes analysis

Classifiers	Percentage of correct classification (%)	TN	TP	FN	FP	Sensitivity (%)	Specificity (%)
BPA	56.25	2	25	11	10	69.4	16.7
PNN	64.5	6	26	10	6	72.2	50
GMM	43.7	2	19	17	10	52.8	16.7
Fuzzy Equivalence	43.7	9	12	24	3	33.3	75
Average	52.0					56.9	39.6

7.4: Classifiers’ performance in the 20 minutes analysis

The 20 minutes analysis showed that all the classifiers except GMM had obtained the highest accuracy of 64.6%. GMM had only managed to obtain 33.3% of correct classification. Based on the table, we can see that the most accurate classifier may not necessary become the most sensitive and specific classifier. Although GMM had the lowest accuracy, it had the highest specificity of 100% similar to Fuzzy Equivalence and PNN. This result is self explanatory since all the normal unknown test data were correctly classified into its own class. BPA is the only classifier which had obtained the highest sensitivity of 80.6%.

Table 7.5: Results for the 4 classifiers in the 20 minutes analysis

Classifiers	Percentage of correct classification (%)	TN	TP	FN	FP	Sensitivity (%)	Specificity (%)
BPA	64.6	2	29	7	10	80.6	16.7
PNN	64.6	12	19	17	0	52.8	100
GMM	33.3	12	4	32	0	11.1	100
Fuzzy Equivalence	64.6	12	19	17	0	52.8	100
Average	56.8					49.3	79.2

7.5: Classifiers’ performance in the 25 minutes analysis

In this final analysis, the average accuracy for the classifiers is the highest as compared to the previous three analyses. The average accuracy had increased from approximately 50% to 77.6%. According to the table, PNN had the highest percentage of correct classification of 83.3%. This was followed by BPA, Fuzzy Equivalence and GMM which have classified the unknown data correctly with an accuracy of 79.2% and 68.7% respectively.

PNN may be the most accurate classifier but it did not manage to classify the greatest number of true negative and true positive cases. Conversely, among the four classifiers, GMM had the highest specificity of 100% and this means that it had correctly classified all the true negative cases. BPA had managed to classify most of the true positive cases correctly, thus allowing it to obtain the highest sensitivity of 91.7%.

Table 7.6: Results for the 4 classifiers in the 25 minutes analysis

Classifiers	Percentage of correct classification (%)	TN	TP	FN	FP	Sensitivity (%)	Specificity (%)
BPA	79.2	5	33	3	7	91.7	41.7
PNN	83.3	9	30	6	3	83.3	75
GMM	68.7	12	21	15	0	58.3	100
Fuzzy Equivalence	79.2	6	32	4	6	88.9	50
Average	77.6					80.5	66.7

7.6: Discussion on the classifiers’ performance in the various time sequences

From this analysis, we can see that the four classifiers did not perform as good as what they had achieved in Chapter 5 and 6. They had only managed to obtain approximately 50% to 80% of accuracy. The drop in percentage could be due to the small number of temperature data used for classification. This is proven in the previous two chapters that these classifiers can perform relatively well when more temperature data were being used.

Based on the observations made, PNN was the most accurate classifier in the four different time sequences; this was followed by BPA and Fuzzy Equivalence. This result was unexpected, as previously, GMM was considered to be one of the most accurate classifier but in this analysis it had the worst performance. Thus, we can only conclude that GMM had greater ability in classifying the unknown data when greater number of temperature data was used.

Figure 7.1: Average accuracy of the classifiers in different time duration

Using the various time sequences analysis, we can see that the percentage of correct classification had gradually improved from 10 minutes to 25 minutes and this is illustrated in Figure 7.1. The average accuracy of the classifiers increased from 49.9% to 77.6% and this shows that longer time duration gives better accuracy. This is further evaluated in Figure 7.2 which shows how the accuracy of each classifier behaves in the four different time duration. The accuracy of all the classifiers increased from 10 minutes to 25 minutes. The red line which represents BPA shows the increment in the percentage of correct classification, similarly to PNN. As for GMM and Fuzzy Equivalence, there is a slight drop in accuracy at 20 minutes and 15 minutes, but the accuracy of these classifiers ultimately increase. Hence, it is inevitable to conclude that longer time duration used in classifying the different diagnoses will yield higher accuracy. This simply means that the accuracy in classifying the diagnoses will be affected when the duration of temperature recording is shortened.

Figure 7.2: Percentage of classification in different time duration

Chapter 8: Comparison of Statistics techniques with Artificial Neural Network Techniques

In the previous 5 chapters, we had looked into the performance of the classifiers using temperature measurements as inputs and it had showed the feasibility of using the non-linear method to detect breast cancer. In this section, we are analysing the statistical features of the temperature measurements and to find out whether they can be used as inputs fed into the classifiers for testing. The purpose of this chapter is to compare and evaluate on the two different methods used.

The statistical parameters used are mean, median, standard deviation, variance, skewness, kurtosis and moment. These parameters will be obtained based on each set of the temperature measurement in each subject. The definition and formula of the statistical parameters will be explained in Table 8.1.

Table 8.1: Definition and formula of the statistical parameters

Parameter	Definition	Formula
Mean	Arithmetic average of a set of values, or distribution	$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$ $\bar{x}$ - mean, n - no of variables, x_i - variable
Median	The number separating the higher half of a sample or a probability distribution, from the lower half.	Median of a finite list of numbers <ul style="list-style-type: none">• Arrange all observations from lowest value to highest value and pick the middle one Even number observations <ul style="list-style-type: none">• Take the mean of the two middle values

Standard deviation	<p>Measure the variability or dispersion of a data set.</p> <p>Low standard deviation</p> <ul style="list-style-type: none"> Data points tend to be very close to the same value (the mean) <p>High standard deviation</p> <ul style="list-style-type: none"> Data are “spread out” over a large range of values 	$s = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2}$ <p>s - standard deviation, $\bar{x}$ - mean, n - no of variables, x_i - variable</p>
Variance	<p>Measure of statistical dispersion, averaging the squared distance of its possible values from the expected value (mean)</p>	$\frac{\sum (x_i - \bar{x})^2}{n-1}$ <p>$\bar{x}$ - mean, n - no of variables, x_i - variable</p>
Skewness	<p>The degree of asymmetry of a distribution around its mean.</p> <p>Positive skewness</p> <ul style="list-style-type: none"> Asymmetric tail extending toward more positive values <p>Negative skewness</p> <ul style="list-style-type: none"> Asymmetric tail extending toward more negative values. 	$\frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^3$ <p>s - standard deviation, $\bar{x}$ - mean, n - no of variables, x_i - variable</p>
Kurtosis	<p>Relative “peakedness” of a distribution compared with the normal distribution.</p> <p>Positive kurtosis</p> <ul style="list-style-type: none"> Relatively peaked distribution. 	$\left[\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^4 \right] - \frac{3(n-1)^2}{(n-2)(n-3)}$ <p>s - standard deviation, $\bar{x}$ - mean, n - no of</p>

	Negative kurtosis • Relatively flat distribution.	variables, x_i - variable
Moment	The moments about its mean μ are called <i>central</i> moments; these describe the shape of the function, independently of translation.	$m_n = E(x - \mu)^k$ m_n - moment, $E(x)$ - expected value of x , μ - mean, k - order

The procedure of the analysis is as followed:

1. Analyse on two diagnoses (normal and cancer) to find out its feasibility using t-test
2. Include one more diagnosis (benign) and use Anova to test for their differences
3. Use the temperature measurements of the three classes as mentioned in Step 2 for classification testing. Also, feed the statistical parameters of the temperature measurements as inputs to the classifiers and compare the result.
4. Lastly, include the last diagnosis (suspected-cancer) and repeat Steps 2 and 3. Conclude on the findings.

8.1: T-test on the two different diagnoses

As a head start to this analysis, two diagnoses (normal and cancer) are used to test on the feasibility of using statistical features of the temperature data for classification. In each diagnosis, 100 sets of temperature data are collected from 4 different patients as shown in Table 8.2. In each set of temperature data, there are 16 temperature measurements which represent the 16 sensors attached to the patient. The statistical parameters are calculated from each set of the temperature data. An example of the tabulated parameters is shown in Figure 8.1 and the calculations on one typical set of this temperature data will be briefly explained.

Table 8.2: Number of temperature data

Classification	No of patients	Number of temperature data
Normal	4	100
Cancer	4	100

Sensor	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Temp	32.64	32.45	33.3	32.68	32.5	32.35	32.07	33.7	32.58	32.98	32.01	32.54	31.48	33.81	33.06	32.33
	32.72	32.53	33.3	32.64	32.46	32.2	31.88	33.66	32.54	33.02	31.9	32.5	31.52	33.48	33.14	32.37
	32.92	32.81	33.63	32.87	32.69	32.47	32.42	34.08	32.98	33.18	32.2	32.78	31.74	33.98	33.34	32.56
	33.04	32.92	33.71	32.99	32.81	32.59	32.34	34.21	33.14	33.31	32.2	32.89	31.97	34.11	33.46	32.68
	33.2	33	33.71	33.07	32.89	32.63	32.3	34.25	33.38	33.35	32.32	32.93	32.08	34.23	33.54	32.88
Mean	32.655	32.56	0.607453702				0.369	0.300246344			0.244008978			-0.004050986		
	32.61625	32.535	0.594058078				0.352905	0.04879755			-0.428977487			0.107500105		
	32.915625	32.84	0.621857098				0.38670625	0.311258552			0.095962824			0.199354613		
	33.023125	32.955	0.63152428				0.398822917	0.36578735			-0.198358806			-0.134066482		
	33.11	33.035	0.63015342				0.397093333	0.291807747			-0.271555879			1.526924616		

Figure 8.1: Example of temperature measurements and tabulation of statistical parameters.

The top table of Figure 8.1 shows the temperature measurements of the 16 sensors. The bottom table shows the statistical parameters calculated from each set of the temperature measurements. The calculations of the parameters for the second row as highlighted in red are shown below.

$$\begin{aligned}
 \text{Mean, } \bar{x} &= \frac{1}{n} \cdot \sum_{i=1}^n x_i \\
 &= \\
 &= \frac{32.72 + 32.53 + 33.3 + 32.64 + 32.46 + 32.2 + 31.88 + 33.65 + 32.54 + 33.02 + 31.9 + 32.5 + 31.52 + 33.48 + 33.14 + 32.37}{16} \\
 &= \frac{521.86}{16} = 32.61625
 \end{aligned}$$

Median: Rearrange the 16 temperature from smallest to largest and find the mean of the middle two values in order to get the median

$$\begin{aligned}
 &[31.52, 31.88, 31.9, 32.2, 32.37, 32.5, 32.5, 32.53, 32.54, 32.64, 32.7, 33.02, 33.14, 33.3, 33.48, 33.66] \\
 \text{Median} &= \frac{32.53 + 32.54}{2} = \frac{65.07}{2} = 32.535
 \end{aligned}$$

$$\begin{aligned}
 \text{Standard deviation, } s &= \sqrt{\frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2} \\
 &= \\
 &= \sqrt{\frac{1}{15} [(32.72 - 32.61625)^2 + (32.53 - 32.61625)^2 + \dots + (32.37 - 32.61625)^2]} \\
 &= 0.594058078
 \end{aligned}$$

$$\begin{aligned}\text{Variance, } & \frac{\sum (x_i - \bar{x})^2}{n-1} \\ &= \frac{(32.72 - 32.61625)^2 + (32.53 - 32.61625)^2 + \dots + (32.37 - 32.61625)^2}{15} \\ &= 0.352905\end{aligned}$$

$$\begin{aligned}\text{Skewness, } & \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^3 \\ &= \frac{16}{15 \cdot 14} \left(\frac{(32.72 - 32.61625) + (32.53 - 32.61625) + \dots + (32.37 - 32.61625)}{0.594058078} \right)^3 \\ &= 0.04879755\end{aligned}$$

$$\begin{aligned}\text{Kurtosis, } & \left[\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^4 \right] - \frac{3(n-1)^2}{(n-2)(n-3)} \\ &= \left(\frac{16 \cdot 17}{15 \cdot 14 \cdot 13} \right) \left(\frac{(32.72 - 32.61625) + (32.53 - 32.61625) + \dots + (32.37 - 32.61625)}{0.594058078} \right)^4 - \left(\frac{3(15)^2}{14 \cdot 13} \right) \\ &= -0.428977487\end{aligned}$$

$$\text{Moment, } m_n = E(x - \mu)^k = 0.107500105$$

$$(k=3 \text{ and } \mu = \bar{x} = 32.61625)$$

Once the statistical parameters are tabulated, *t*-test is carried out. A *t*-test is any statistical hypothesis test in which the test statistic has a Student's *t* distribution if the null hypothesis is true. Independent two sample *t*-test is used in this section. P value will be calculated to determine if the two samples are significantly independent. If the P value is less than 0.01, this shows that the two samples are statistically different. The results of the *t*-test for the two different diagnoses (normal and cancer) are tabulated in Table 8.3.

Table 8.3: Result of *t*-test for the two different diagnoses (normal and cancer)

	Diff	95% of CI	T	DF	P value
Mean	0.08708	-0.1125931 to 0.28675	0.86	198	0.3908
Median	0.1038	-0.0876173 to 0.29521	1.069	198	0.2862
Standard deviation	0.1708681	0.069114979 to 0.2726	3.311	198	0.0011
Variance	0.47712767	0.302294428 to 0.651960912	5.382	198	< 0.0001
Skew	0.08245402	-0.058151341 to 0.223059381	1.156	198	0.2489
Kurtosis	-0.650621	-0.921400081 to -0.379841919	-4.738	198	< 0.0001
Moment	-0.0195916	-0.204633897 to 0.165450637	-0.209	198	0.8348

From Table 8.3, several parameters obtained from the t-test analysis are tabulated. However, only the P value makes a significant impact on this analysis. The table shows that the P value for standard deviation, variance and kurtosis are less than 0.01, and this means that these parameters are significantly independent and different. With this result obtained, we can see the feasibility of using statistical features of the temperature data for classifying the different diagnoses (normal and cancer). In the next section, we will include one more diagnosis (benign) to further elaborate on this study.

8.2: ANOVA analysis on the 3 different classes (normal, benign and cancer)

In this section, Benign is the 3rd diagnosis added to the analysis and the number of temperature data will also be increased. The total sets of temperature data collected are 3000, with 1000 data from each diagnosis. These temperature measurements will be converted into the seven statistical parameters as mentioned in the previous section. T-test can no longer be used to analyse on the relationship between the diagnoses as it is only suitable for two different classes with small limited sample sizes. With the addition of one more diagnosis and the increment of the temperature data, we need to use another tool for analysis.

Anova known as the analysis of variance will be used to analyse and compare the statistical parameters for the 3 different diagnoses instead. ANOVA is a general technique used to test the hypothesis that the means among two or more groups are equal, under the assumption that the sampled populations are normally distributed. One way analysis of variance is used to compare the parameters. The statistical analysis of Anova are compiled in a graph and tabulated in a table. This analysis is included in Appendix 1. However, the results of F statistic and P value are shown in Table 8.4 as these two values play an important role in determining whether the parameters are significantly different. If the P value is less than 0.01, this suggests that the parameters are significantly different, and the value of F statistic is conversely proportional to the P value. If the value of F statistic is large, the value of P will be smaller as shown in Table 8.4.

Table 8.4: Result of ANOVA for 3 the different diagnoses (normal, benign and cancer)

Parameters	F Statistic	P value
Mean	200.43	0
Median	195.72	0
Standard deviation	227.56	0
Variance	308.01	0
Skew	56.04	0
Kurtosis	106.66	0
Moment	2.73	0.0657

The value of P for parameters mean, median, standard deviation, variance, skew, and kurtosis is 0 and this indicates that these parameters from the 3 different diagnoses are significantly different. However for moment, its P value is 0.0657 and the F statistic value is conversely low. Thus, there is little evidence to imply that the parameter of moment for the 3 different diagnoses had slight effect on one another.

Next, we will apply the six parameters which are significantly different and use them to test on the classifiers. Moreover, as a head start to this analysis, only 3 parameters are selected to test on the classifiers and they are mean, median and standard deviation. They will be fed as inputs to the classifiers to classify the 3 various diagnoses namely, normal, benign and cancer. Figure 8.2 shows the diagram on how the 3 parameters are being fed into the classifier. The classifiers used for this test is Probabilistic Neural Network (PNN), Gaussian Mixture Model (GMM) and Fuzzy Equivalence. These parameters will be normalized based on the

formula: $\left[\frac{(T_{\max} - T)}{(T_{\max} - T_{\min})} \right]$ as evaluated in Chapter 6, after which it will be fed into the classifiers for testing.

Figure 8.2: Diagram on the 3 parameters being fed into the classifier

Table 8.5 indicates the result of correct classification obtained by the 3 parameters (mean, median and standard deviation). The average percentage of correct classification obtained is 72.7%, and this shows that it is feasible to use statistical parameters as inputs for classifying the unknown data. To further testify for this analysis, the remaining three parameters (variance, skew and kurtosis) will be included and there will be a total of 6 statistical parameters fed as inputs to the classifiers for testing.

The method mentioned in Chapter 5 will also be applied in this section. Instead of using statistical parameters, the temperature data used to convert the statistical parameters will be fed as inputs to the classifiers for identifying the various diagnoses. The aim is to compare and evaluate on the results of using two different

types of inputs (temperature data and statistical parameters of the temperature data) to correctly classify the 3 different diagnoses. This will enable us to determine which is the more powerful and appropriate method for classification. The results of these two methods are tabulated in Table 8.6.

Table 8.5: Result of correct classification using 3 statistical parameters

Classifier	No of training data	No of testing data	Percentage of correct classification
PNN	2700	300	75.6
GMM	2700	300	76.3
Fuzzy Equivalence	2700	300	66.3
Average			72.7

Table 8.6: Result of correct classification for 3 different diagnoses

Classifier	No of training data	No of testing data	Percentage of correct classification (%)	
			6 Statistical Parameters	Temperature data
PNN	2700	300	70.3	93.6
GMM	2700	300	65.0	99.6
Fuzzy Equivalence	2700	300	63.6	98.6
Average			66.3	97.3

Table 8.6 shows that all the classifiers had better performance when using the temperature data as inputs than using the statistical parameters of the temperature data. Six statistical parameters are being fed as inputs and they are only able to obtain an average of 66.3% of correct classification, while the latter had average of 97.3% accuracy. Hence, this proved that using the temperature data as inputs had greater performance than the statistical method. Lastly in the next section, the last diagnosis (suspected cancer) will be included and the procedure will be repeated to conclude on the performance of the two methods (temperature data and statistical parameters of the temperature data).

8.3: ANOVA analysis on the 4 different classes (normal, benign, cancer, suspected cancer)

The last diagnosis, suspected-cancer was added in this analysis and a total of 7600 temperature data is collected. These data will be converted into the 7 statistical parameters. Anova analysis is also carried out to analyze on the relationship between the parameters of the 4 different diagnoses. While the details of the Anova analysis are shown in Appendix 2, the value of F statistic and P value are tabulated in Table 8.7.

Table 8.7: Result of ANOVA between the 4 different diagnoses

Parameters	F Statistic	P value
Mean	343.63	0
Median	297.42	0
Standard deviation	147.21	0
Variance	202.74	0
Skew	177.10	0
Kurtosis	106.11	0
Moment	0.41	0.7461

All the parameters except moment had obtained relatively high F statistic value and zero P value. Thus, this shows that these parameters are significantly different when compared with the 4 various diagnoses. The six parameters will be fed into the classifiers for accuracy test and the results are tabulated in the Table 8.8. In the

7600 temperature data, 6600 of them were used as training data while the remaining 1000 data were used as testing data for the classification.

Similarly to Chapter 8.2, the classifiers using temperature data as inputs had better performance as compared to the statistical parameters of the temperature data. Their average accuracies are 92.8% and 56.7% respectively. The method using temperature data as inputs can also be referred as Artificial Neural Network (ANN) method. GMM had achieved the highest percentage of correct classification of 96.7% in this method. Therefore, it was concluded that the ANN being non-linear had better performance and is a much more powerful method. This had also assured us that we had used the right method in classification as mentioned in Chapters 5 and 6.

Table 8.8: Result of classifier for 4 different diagnoses

Classifier	No of training data	No of testing data	Percentage of correct classification (%)	
			6 Statistical Parameters	Temperature data (ANN method)
PNN	6600	1000	53.2	94.5
GMM	6600	1000	64.6	96.7
Fuzzy Equivalence	6600	1000	52.4	87.1
Average			56.7	92.8

Chapter 9: Discussion on the performance of the classifiers

In this thesis, six classifiers are used for the detection and classification of the various stages of the breast cancer. The performance of the classifiers is being evaluated throughout the whole thesis. Out of the six classifiers, Fuzzy (Sugeno-type) and Support Vector Machine (SVM) did not perform relatively well as compared to the rest of the classifiers. Fuzzy (Sugeno-type) is suited for modeling non-linear systems, has a fuzzy inference system which is able to interpolate the linear gains applied across the input space, and it also required interpolation between multiple linear models so that the classifier can perform ideally well. Moreover, the non-linearity behavior of the temperature measurements may have deterred the good performance of the classifier.

Similarly, Support Vector Machine (SVM) belongs to a family of generalized linear classifiers. Therefore, it did not perform well in classifying the non-linear temperature data even though it uses the non-linear kernel functions to transform the input data to a high-dimensional feature space in which input data become more separable compared to the original input space. Furthermore, it has a greater ability in classifying two classes as compared to four different classes of diagnoses.

The performance of PNN, BPA and Fuzzy Equivalence as evaluated in Chapter 6 was relatively close, with each of them performing at least 80% of accuracy. In BPA, the temperature data were divided into training and testing data. The training data will be fed into the feed forward networks, consisting of two hidden layers of sigmoid neurons which have non-linear transfer functions, allowing the network to learn non-linear and linear relationships between the input and output vectors. The training will be repeated until the error is minimized and the set of testing data will

be fed into the neurons to determine the desired output. In this way, this increased the accuracy level of BPA.

Moreover, PNN has a higher accuracy than BPA and this is proven in all the analysis. PNN has a fourth layer which includes a loss matrix. This is multiplied by the probability estimates in the third layer, and the class with lowest estimated cost is selected. This allows the PNN to generate accurate predicted target probability scores. Furthermore, the neural networks (PNN and BPA) are more resistant to insufficient data amount. They are able to perform relatively well even though smaller temperature data were being used in Chapter 7 as compared to GMM and Fuzzy Equivalence which require more temperature data to perform well.

Fuzzy Equivalence has a better performance than Fuzzy (Sugeno-type) because the latter do not need to derive rules for segmenting the data, and the unknown data is predicted into known classes based on its similarity. Lastly, Gaussian mixture model is an unsupervised classifier which had obtained the highest accuracy in most of the analyses carried out during this work. Being an unsupervised classifier, the training samples of a classifier are not labeled to show their category membership, it is during training when GMM will try to approximate the underlying probability density functions (pdf's) of the observations. In addition, GMM has the ability to reduce data complexity, greater efficiency in data representation and it has also proved to be useful if the data features are weak or missing. Thus, these could probably be the unique characteristics of GMM which makes it the best classifier as compared to the rest of the classifiers.

Chapter 10: Conclusion and Future Works

10.1: Conclusion

In this work, the interpretive software system was further enhanced and developed for the detection and classification of the breast cancer patients. This system incorporated dynamic thermal analysis and analytical software to produce a potential tool for detection of breast cancer. The performance of the six classifiers; namely, fuzzy equivalence, back-propagation algorithm (BPA), probabilistic neural network (PNN), fuzzy (Sugeno-type), Gaussian mixture model (GMM) and support vector machine (SVM) was compared and evaluated. The accuracy of these 6 classifiers generally depends on the size and quality of the training data, the rigor of the training imparted and also the parameters used to represent the input (breast surface temperature).

In Chapter 5, we reported that all the classifiers were able to obtain an average accuracy of 85%. The studies had also showed that Fuzzy equivalence, GMM and SVM had performed better than the rest of the classifiers. More tests are carried out to further analyse on this deduction. Hence, with more temperature data being analysed, the three classifiers were able to achieve an average accuracy of 92% in classifying the four different diagnoses (normal, benign, cancer and suspected-cancer). Fuzzy Equivalence was considered to be an excellent classifier as it had obtained 100% of accuracy in classifying the unknown data. Furthermore, it had also achieved the best result in the four performance indices, namely, sensitivity, specificity, positive predictive value and area under the curve which simply implies its accuracy. Therefore among the three classifiers, Fuzzy Equivalence can be considered as the most outstanding classifier.

An analysis on the different methods of temperature scaling was demonstrated in Chapter 6. The purpose is to determine a more suitable normalization method as the previous method was dependent on the temperature scaling. Through the analysis, we had concluded on using $\left[\frac{(T_{\max} - T)}{(T_{\max} - T_{\min})} \right]$ for the new temperature scaling. The performance of the classifiers was not affected as they are able to attain an average accuracy of 83.1%. In this case, GMM had achieved the highest accuracy of 90.3%.

In Chapter 7, another analysis was implemented to find out if temperature data with shorter time duration can affect the accuracy of the classifiers. The temperature data were arranged into various period of time ranging from 10 minutes to 25 minutes. Moreover, this duration had to follow a suitable time frame from 6pm to 12am due to the occurrence of circadian thermal activity. The result of this analysis showed that the four classifiers did not perform relatively well as compared to the results obtained in Chapters 5 and 6. They had only managed to achieve approximately 50% to 80% of accuracy. The decrease in percentage could be due to the small number of temperature data available for classification and this is proven in Chapters 5 and 6 where the classifiers have better performance when more temperature data were being used.

The analysis had also showed that BPA and PNN could perform better with lesser temperature data. This was an unexpected phenomenon for GMM and Fuzzy Equivalence as they had better results in Chapter 5. Thus, it is concluded that GMM and Fuzzy Equivalence have greater ability in classification when more temperature data were used for testing. Furthermore, we can see that the percentage of correct classification had gradually improved from 49.9% to 77.6% in the time duration of 10 minutes to 25 minutes. In addition, the accuracy of each classifier

had also increased from 10 minutes to 25 minutes. Hence, these showed that using longer time duration of temperature data in classifying the different diagnoses would yield higher accuracy.

Lastly, the statistical features of the temperature measurements were analysed to determine whether they can be used as inputs fed into the classifiers for testing. The aim is to compare with the method used in Chapter 5 and evaluate on the performance of the two different methods. Based on the result, we could conclude that the classifiers performed better when the artificial neural network method was being used since it was a non-linear and more powerful method. This further assured us in using this method for classification and detection of the breast cancer. Based on all the analyses done throughout this study, the two classifiers, Fuzzy equivalence and GMM, have showed greater capability in performing better than the rest of the classifiers. Regardless of whether the new method of temperature scaling is used or the statistics techniques compared with the ANN method, both classifiers had relatively good results. Although the ANN classifier had obtained good results too, but it required longer training time to achieve such results. Hence, it is inevitable to conclude that both Gaussian mixture model and Fuzzy Equivalence had the potential to be further developed as the best analytical software which will be incorporated with dynamic thermal analysis for the detection of breast cancer.

10.2: Limitations and Future Work

In this study, we have successfully developed the analytical software which will be incorporated with the dynamic thermal analysis for detection of breast cancer. However, there are some limitations which we faced doing the analysis. Firstly, the measurement of the patients' temperature was provided and conducted by Lifeline Biotechnologies, Inc, Florida, USA. This may not hinder our analysis as we are only required to obtain the patients' temperature and fed into the classifiers for identifying the different diagnoses. But the quality of the temperature measurements may serve as a problem to us. Information was provided on how the temperature measurements were taken but details on where and what is the ambient temperature of the environment when the measurements were taken was not provided. This may affect on the quality of the temperature measurements which will in turn affect the accuracy of the classifiers' testing.

Secondly, the unusual behavior of the temperature measurements can be a problem to our analysis. It is not easy for an engineer to study on the physiology of the temperature behavior. We had mentioned in Chapter 2 that higher temperature is proportionally related to the growth and location of the breast tumours. But this may not be true as it could be due to the menstruation which leads to such high temperature. Hence, choosing a good quality of temperature measurements for classifiers' testing can be a challenge at times.

Lastly, we had concluded that most of the classifiers are only able to work better when more temperature data were being used. This may become a problem in future when the device was implemented. We need to obtain sufficient temperature data in order for the classifiers to behave well and this simply means longer time

recording is required by the patients. This may hinder the patients from wanting to wear a device on their bodies for such long hours as compared to mammography which saves a lot of hassle.

In future, it may be possible to use the sets of temperature measurements where the breast tumors were found to test on the classifiers. We can based on the information provided by Lifeline Biotechnologies, Inc, Florida, USA and identify where the tumours were being found and locate the temperature measurements around that area. This could help to improve on the quality of classifiers' testing.

The results of the classifiers may further be improved by using more diverse (data from different gene pools) training data for longer durations and using the temperature data taken under identical conditions such as after food and before/after menstruation. We can also use non-linear features of the temperature data to improve on the accuracy of the classifiers. The future work will definitely served as a further enhancement for the development of this system.

References

1. National Breast Cancer Foundation, INC.
2. American Cancer Society, Cancer Facts and Figures 2008: Atlanta: American Cancer Society; 2008.
3. Sobti A, Sobti, P.S., Keith L.G., Screening and diagnostic mammography: Why the Gold Standard Does Not Shine More Brightly? *Int J Fert Women's Med*, 2005. **50**: p. 199-206.
4. Breast Cancer in Young Women.
5. Wright H, Rim J.L.A., Chellman-Jeffers M., Patrick R., Rybicki L., Kim J., Crowe J., Magnetic resonance imaging as a diagnostic tool for breast cancer in premenopausal women. *Am J Surg*, 2005. **190**: p. 572-575.
6. Constance D. Lehman, C.G., Christiane K. Kuhl, R. Edward Hendrick, Etta D. Pisano, Lucy Hanna, Sue Peacock, Stanley F. Smazal, Daniel D. Maki, Thomas B. Julian, Elizabeth R. DePeri, David A. Bluemke, Mitchell D. Schnall, for the ACRIN Trial 6667 Investigators Group MRI Evaluation of the Contralateral Breast in Women with Recently Diagnosed Breast Cancer 2007. **356**: p. 1295-1303 (13).
7. Warner E, P.D., Hill KA, Causer PA, Zubovits JT, Jong RA, Cutrara MR, DeBoer G, Yaffe MJ, Messner SJ, Meschino WS, Piron CA, Narod SA, Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. . *JAMA*, 2004. **292**(11): p. 1317-1325.
8. Kriege M, B.C., Boetes C, Besnard PE, Zonderland HM, Obdeijn IM, Manoliu RA, Kok T, Peterse H, Tilanus-Linthorst MM, Muller SH, Meijer S, Oosterwijk JC, Beex LV, Tollenaar RA, de Koning HJ, Rutgers EJ, Klijn JG, Efficacy of MRI and Mammography for Breast Cancer Screening in

- Women with a Familial or Genetic Predisposition. *New England J. of Medicine*, 2004. **351(5)**: p. 427-437.
9. Lehman CD, B.J., Weatherall P, Thickman D, Hylton N, Warner E, Pisano E, Schnitt SJ, Gatsonis C, Schnall M, DeAngelis GA, Stomper P, Rosen EL, O'Loughlin M, Harms S, Bluemke DA, International Breast MRI Consortium Working Group, Screening women at high risk for breast cancer with mammography and magnetic resonance imaging, *Cancer*. 2005. **103(9)**: p. 1898-1905.
 10. Local cancer specialists agree with breast MRIs for high-risk patients. 6th April 2007.
 11. Keith LG, O.J., Laguens M, Circadian rhythm chaos: A new breast cancer marker. *Int J Fertil Women Med*, 2001. **46**: p. 238-247.
 12. Salhab M, Sarakbi W AI and Mokbel K, The evolving role of the dynamic thermal analysis in the early detection of breast cancer. *International Seminars in Surgical Oncology*, 2005. **2:8** p. doi: 10.1186/1477-7800-2-8
 13. Gros C., Bourjat M.G.P., Prognosis and post therapeutic follow-up of breast cancers by thermography. Aarts NJM, Gautherine M, Ring EFJ., editor. *Thermography*. Karger, Basel, 1975: p. 77-90.
 14. Gautherine M.C.G., Contribution of infrared thermography to early diagnosis, pretherapeutic prognosis and post-irradiation follow-up of breast carcinomas. *Med Mundi*, 1976. **21**: p. 135-149.
 15. Philips MJ, W.D., Simpson HW, Fahmy DR, Groom GV, Philips ME, Pierrepont CG, Blamey RW, Halberg F, Griffiths K, Characterisation of breast skin temperature rhythms of women in relation to menstrual status. *Acta Endocrinologica*, 1981. **96(3)**: p. 350-360.
 16. Ng E.Y.K., Fok S.C., Ng F.C., Sim L.S, Computerized detection of breast cancer with artificial intelligence and thermograms. *Int J Med Engg. & Technol*, 2002. **26(4)**: p. 152-157.

17. Salhab M; Keith LG; Laguens M; Reeves W; Mokbel K, The potential role of dynamic thermal analysis in breast cancer detection. *International Seminars in Surgical Oncology* 2006. **3:8** p. doi: 10.1186/1477-7800-3-8. (ISSN: 14777800)
18. Ng, E.Y.-K., Tan, M.S., Lockwood, S., and Keith, L.G, ANN based Classification of Breast Cancer with Discrete Temperature Screening: Facts and Myths, Part-IV, *Emerging Technologies in Breast Imaging and Mammography*, American Scientific Publishers: USA/Canada, Editors: Profs. Jasjit S. Suri, R. Rangayyan and S. Laxminarayan, (Chap.21, ISBN: 1-58883-090-X): p. 1-20.
19. Ng, E.Y.-K., R Acharya U, Keith, L.G., and Lockwood, S, Detection and Classification of Breast Cancer using Neural Classifiers with First Warning Thermal Sensors. . *Information Sciences*, 2007. **177**(20): p. 4526-4538.
20. Joanna MY Tan, Ng E.Y.K., Rajendra Acharya U, Louis G. Keith, Jim Holmes Comparative study on the use of analytical software to identify the different stages of breast cancer using discrete temperature data. *Journal of Medical Systems*, 2009. **33**(2): p. 141-153.
21. Lifeline Biotechnologies, Inc, Florida, USA.
22. Sickles. EA, Findings at mammographic screening on only one standard projection: outcomes analysis. *Radiology* 1998. **208**(2): p. 471-5.
23. Full-field digital mammography: A potential alternative to the traditional film-screen technique?
24. Discover the benefits of a digital mammogram and learn the importance of early screening for breast health.
25. Louis G. Keith, Jaroslaw J. Oleszczuk and Martin Laguens, Are Mammography and Palpation Sufficient for Breast Cancer Screening? A Dissenting option. *Journal of Women's Health & Gender-Based Medicine*, 2002. **11**(1).

26. Carney PA, Miglioretti DL, Yankaskas BC, et al., Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Ann Intern Med* 2003. **138**(3): p. 168-75.
27. Rosenberg RD, Hunt WC, Williamson MR, Effects of age, breast density, ethnicity, and estrogen replacement therapy on screening mammographic sensitivity and cancer stage at diagnosis: review of 183,134 screening mammograms in Albuquerque, New Mexico. *Radiology* 1998. **209**(2): p. 511-8.
28. Kerlikowske K, Grady D, Barclay J, Likelihood ratios for modern screening mammography. Risk of breast cancer based on age and mammographic interpretation. *JAMA*, 1996. **276**(1): p. 39-43.
29. Kerlikowske K, Grady D, Barclay J, Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting and Data System. *J Natl Cancer Inst* 1998. **90**(23): p. 1801-9.
30. Elmore JG, Wells CK, Lee CH, Variability in radiologists' interpretations of mammograms. *N Engl J Med* 1994. **331**(22): p. 1493-9.
31. Elmore JG, Wells CK, Howard DH, The impact of clinical history on mammographic interpretations. *JAMA*, 1997. **277**(1): p. 49-52.
32. Barlow WE, Chi C., Carney PA, Accuracy of screening mammography interpretation by characteristics of radiologists. *J Natl Cancer Inst* 2004. **96**(24): p. 1840-50.
33. Smith-Bindman R, Chu P., Miglioretti DL, Physician predictors of mammographic accuracy. *J Natl Cancer Inst* 2005. **97**(5): p. 358-67.
34. Esserman L, Cowley H., Eberle C, Improving the accuracy of mammography: volume and outcome relationships. *J Natl Cancer Inst* 2002. **94**(5): p. 369-75.

35. Kan L, Olivotto IA, Warren Burhenne LJ, Standardized abnormal interpretation and cancer detection ratios to assess reading volume and reader performance in a breast screening program. . Radiology. 2000. **215**(2): p. 563-7.
36. Beam CA, Conant EF, Sickles EA, Association of volume and volume-independent factors with accuracy in screening mammogram interpretation. . J Natl Cancer Inst 2003. **95**(4): p. 282-90.
37. Coldman AJ, Major D., Doyle GP, Organized breast screening programs in Canada: effect of radiologist reading volumes on outcomes. . Radiology, 2006. **238**(3): p. 809-15.
38. Screening with other imaging modalities: early investigations.
39. Salynn Boyles, Ultrasound Helps Breast Cancer Detection in WebMD Health News. 2008.
40. Fay T, Henry GC, Correlation of body segmental temperature and its relation to the location of carcinomatous metastasis. Surg Gynecol Obstet 1966. **38**: p. 512-24.
41. Lawson RN, Implications of surface temperatures in the diagnosis of breast cancer. . Can Med Assoc J 1956. **75**: p. 309-10.
42. Lawson RN, Thermography: a new tool in the investigation of breast lesions. . Can Serv Med J, 1957. **23**: p. 517-24.
43. Gautherie M, Gros Ch, Circadian rhythm alteration of skin temperature in breast cancer. Chronobiologia, 1977. **4**: p. 1-17.
44. Gautherie M, H.P., Walter JP, Keith LG, Thermovascular changes associated within situ and minimal breast cancers; results of an ongoing prospective study after four years. 1987. **11**: p. 833-842.
45. Wasserman, P.D., Advanced Methods in Neural Computing. 1993, New York, USA: Van Nostrand Reinhold.
46. DTREG: Software for predictive modeling and forecasting.

47. Nozaki, I., T, Selecting fuzzy if-then rules with forgetting in fuzzy classification systems. *Journal of Japan Society for Fuzzy Theory and Systems*, 1994. 6(3): p. 585-602.
48. Ishibuchi, M., Tanaka, Construction of fuzzy classification systems using genetic algorithms. *Journal of Japan Society for Fuzzy Theory and Systems*, 1995. 7(5): p. 1022-1040.
49. George, K., Bo, "Fuzzy Sets and Fuzzy Logic: Theory and Applications". India: Prentice Hall India, 1995.
50. Christopher M. Bishop, *Pattern Recognition and Machine Learning*, Springer 2006.
51. Reynolds, D.A., Rose, R.C., "Robust text-independent speaker identification using Gaussian mixture speaker models." *IEEE Transactions on Speech and Audio Processing*, Vol 3, 1995. p. 72-83.
52. Seo, C., Lee, K.Y., Lee, J., "GMM based on local PCA for speaker identification," *Electronics Letters*, 2001. 37: p. 1486-1488.
53. Burgess, C.J.C., "A tutorial on support vector machines for pattern recognition". in *Data mining and Knowledge discovery*, 1998. 2(2): p. 1-47.
54. Vapnik, V., *Statistical Learning Theory*. 1998, New York: Willey.
55. David, V., and Sanchez, A., *Advanced support vector machines and kernel methods*. in *Neurocomputing*: Elsevier, 2003.
56. Lipo Wang, *Support Vector Machines: Theory and Applications*. 2005: Springer.
57. Muller, K.R., Mika, S., Ratsch, G., Tsuda, K., and Scholkopf, B., An introduction to Kernel Based Learning Algorithms. *IEEE Transactions on Neural Networks*, vol 12, 2001.
58. Hsu C.W and Lin C.J, A comparison of methods for multi-class support vector machines. *IEEE Trans. on Neural Networks*, 2002. Vol 13: p. 415-425.

59. Weston J. and Watkins C., Multi-class support vector machines. Technical Report CSD-TR-98-04. 1998, Department of Computer Science, Royal Holloway, University of London, Egham, TW20 0EX, UK.
60. Lei H., G.V., "Half-Against-Half Multi-class Support Vector Machines", in Proceeding Sixth International Workshop on Multiple Classifier Systems (MCS'05), Berlin, Germany: Springer-Verlag, 2001. p. 156-164.
61. Platt J.C, Chrisianni N., and Shawe-Taylor J, "Large Margin DAGs for multiclass classification", Advances in Neural Information Processing Systems, 2000. 12: p. 547-553. MIT Press.
62. Patrick Bosc, Henri P., An introduction to the fuzzy set and possibility theory-based treatment of soft queries and uncertain or imprecise databases. 1994.
63. GORUNESCU Florin, Benchmarking Probabilistic Neural Network Algorithms. 2006. [Last accessed 2009 4th January] Available from: <http://www.inf.ucv.ro/~aidc/proceedings/2006/1%20fgorunescu.pdf>
64. Rong Jin, Huan Liu, Robust Feature Induction for Support Vector Machines. 2004.
65. Schoonjans, F. MedCalc Online Help, Help-Topics: Statistics, ROC. MedClac Software and Receiver Operating Characteristics (ROC).
66. DeLeo, J.M. Receiver Operating Characteristic Laboratory (ROCLAB): Software for developing decision strategies that account for uncertainty management in artificial neural network decision-making. In Proceedings of Second International Symposium on Uncertainty Modeling and Analysis, 25-28 Apr 1993. 1993.
67. Downey, T.J., Jr; Meyer, D.J.; Price, R.K.; Spitznagel, E.L, Using the Receiver Operating Characteristic to Assess the Performance of Neural Classifiers. Neural Networks, 1999. 5: p. 3642-3646.

68. Ng, E.Y.-K., "Is Thermal Scanner Losing its Bite in Mass Screening of Fever due to SARS?" *Medical Physics*, American Association of Physicists in Medicine 2005, 32 (1): Pp. 93--97.
69. Saltus, R. Scientists grapple with mammogram controversy. New York Times Syndicate, Your Daily Health (quoted from The Boston Globe on internet), pp1-3).
70. Solmolensky, M., G.D.T. Dodd, A. Zermeno and L.M. Marsh. "Application of Rhythmometry for the Evaluation of Temporal J. Chronobiology. 1974; 360-6.
71. Echave Lianos, J.M., R.E. Nash. "Mitotic Circadian Rhythm in a Fast Growing and a Slow Growing Hepatoma: Mitotic Rhythms in Repatonia." *J.Nat. Cancer Inst.* 1970; 44:581-5.
72. Garcia-Sainz, M., F. Halberg. "Mitotic Rhythm in Human Cancer, Re-evaluated by Electronic Computer Programs: Evidence for Temporal Pathology." *J.Nat. Cancer Inst.* 1966; 37:279-292.
73. Tahti. et al "Studies of the Effects of X-Radiation on 24-hour Variations in the Mitotic Activity in Human Malignant Tumours." *Acta Path Microbiol. Scand.* 1956; 117:1-61.
74. Luce, C.G. "Biological Rhythms in Psychiatry and Medicine." Public Health Service Publ. Chevy Chase, MD. 1970; 67-70.

Appendix 1: Anova analysis on the statistical parameters for the 3 different diagnoses

Anova analysis for mean

Anova analysis for median

Anova analysis for standard deviation

Anova analysis for variance

Anova analysis for skewness

Anova analysis for kurtosis

Anova analysis for moments

Appendix 2: Anova analysis on the statistical parameters for the 4 different diagnoses

Anova analysis for mean

Anova analysis for median

Anova analysis for standard deviation

Anova analysis for variance

Anova analysis for skewness

Anova analysis for kurtosis

Anova analysis for moments