Development of Features and Feature Reduction Techniques for Mammogram Classification

Ph. D. Thesis

by

Shradhananda Beura

Department of Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela - 769008, India
June 2016

Development of Features and Feature Reduction Techniques for Mammogram Classification

A dissertation submitted to the department of

Computer Science and Engineering

of

National Institute of Technology Rourkela

in partial fulfilment of the requirements
for the degree of

Doctor of Philosophy

by

Shradhananda Beura
(Roll No- 512CS1008)

under the supervision of

Prof. Banshidhar Majhi

and

Prof. Ratnakar Dash

Department of Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela - 769008, India
June 2016

Department of Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela - 769 008, Odisha, India.

June 11, 2016

Certificate of Examination

Roll Number: 512CS1008
Name: Shradhananda Beura
Title of Dissertation: Development of Features and Feature Reduction Techniques for Mammogram Classification

We the below signed, after checking the dissertation mentioned above and the official record book (s) of the student, hereby state our approval of the dissertation submitted in partial fulfillment of the requirements of the degree of **Doctor of Philosophy in Computer Science and Engineering** at **National Institute of Technology Rourkela**. We are satisfied with the volume, quality, correctness, and originality of the work.

Ratnakar Dash
Co-supervisor

Banshidhar Majhi
Principal Supervisor

Pankaj Kumar Sa
Member, DSC

Bidyut Kumar Patra
Member, DSC

Supratim Gupta
Member, DSC

Sumantra Dutta Roy
External Examiner

Santanu Kumar Rath
Chairperson, DSC & HOD

Department of Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela - 769 008, Odisha, India.

June 11, 2016

Supervisors' Certificate

This is to certify that the work in the thesis entitled **Development of Features and Feature Reduction Techniques for Mammogram Classification** by **Shradhananda Beura**, bearing roll number 512CS1008, is a record of an original research work carried out by him under our supervision and guidance in partial fulfillment of the requirements for the award of the degree of **Doctor of Philosophy** in **Computer Science and Engineering**. Neither this thesis nor any part of it has been submitted for any degree or academic award elsewhere.

Ratnakar Dash
Co-supervisor

Banshidhar Majhi
Principal Supervisor

Dedicated To My Family

Acknowledgment

No work goes unfinished
I take this opportunity to thank all those who have contributed in this journey.

Foremost, I would like to express sincere gratitude to my advisor, Prof. Banshidhar Majhi for providing motivation, enthusiasm, and critical atmosphere at the workplace. His profound insights and attention to details have been true inspirations to my research. Prof. Majhi has taught me to handle difficult situations with confidence and courage.

I would like to thank Prof. Ratnakar Dash for his constructive criticism during the entire span of research. His insightful discussions has helped me a lot in improving this work.

My sincere thanks to Prof. P. K. Sa, Prof. S. K. Rath, Prof. B. K. Patra, and Prof. S. Gupta for their continuous encouragement and valuable advice.

I would like to thank my friends and colleagues at NIT Rourkela for the help they have offered during the entire period of my stay.

Finally, I owe the heartfelt thanks to my in-laws and parents for their unconditional love, support, and patience. Special thanks go to my father-in-law who has supported me a lot to finish this piece of work. Words fall short to express gratitude to my wife, Sujata Mallick, who has been the constant source of inspiration to me. I am indeed grateful to you for your support and understanding.

Shradhananda Beura

Abstract

Breast cancer is one of the most widely recognized reasons for increased death rate among women. For reduction of the death rate due to breast cancer, early detection and treatment are of utmost necessity. Recent developments in digital mammography imaging systems have aimed to better diagnosis of abnormalities present in the breast. In the current scenario, mammography is an effectual and reliable method for an accurate detection of breast cancer. Digital mammograms are computerized X-ray images of breasts. Reading of mammograms is a crucial task for radiologists as they suggest patients for biopsy. It has been studied that radiologists report several interpretations for the same mammographic image. Thus, mammogram interpretation is a repetitive task that requires maximum attention for the avoidance of misinterpretation. Therefore, at present, Computer-Aided Diagnosis (CAD) system is exceptionally popular which analyzes the mammograms with the usage of image processing and pattern recognition techniques and classify them into several classes namely, malignant, benign, and normal. The CAD system recognizes the type of tissues automatically by collecting and analyzing significant features from mammographic images.

In this thesis, the contributions aim at developing the new and useful features from mammograms for classification of the pattern of tissues. Additionally, some feature reduction techniques have been proposed to select the reduced set of significant features prior to classification. In this context, five different schemes have been proposed for extraction and selection of relevant features for subsequent classification. Using the relevant features, several classifiers are employed for classification of mammograms to derive an overall inference. Each scheme has been validated using two standard databases, namely MIAS and DDSM in isolation. The achieved results are very promising with respect to classification accuracy in comparison to the existing schemes and have been elaborated in each chapter.

In Chapter 2, hybrid features are developed using Two-Dimensional Discrete Wavelet Transform (2D-DWT) and Gray-Level Co-occurrence Matrix (GLCM) in succession. Subsequently relevant features are selected using *t-test*. The resultant feature set is of substantially lower dimension. On application of various classifiers it is observed that Back-Propagation Neural Network (BPNN) gives better classification accuracy as compared to others. In Chapter 3, a Segmentation-based Fractal Texture Analysis (SFTA) is used to extract the texture features from the mammograms. A Fast Correlation-Based Filter (FCBF) method has been used to generate a significant feature subset. Among all classifiers, Support Vector Machine (SVM) results superior classification accuracy. In Chapter 4, Two-Dimensional Discrete Orthonormal S-Transform (2D-DOST) is used to extract the features from mammograms. A feature selection methodology based on null-hypothesis with statistical *two-sample t-test* method has been suggested to select most significant features. This feature with AdaBoost and Random Forest (AdaBoost-RF) classifier outperforms other classifiers

with respect to accuracy. In Chapter 5, features are derived using Two-Dimensional Slantlet Transform (2D-SLT) from mammographic images. The most significant features are selected by utilizing the Bayesian Logistic Regression (BLogR) method. Utilizing these features, LogitBoost and Random Forest (LogitBoost-RF) classifier gives the better classification accuracy among all the classifiers. In Chapter 6, Fast Radial Symmetry Transform (FRST) is applied to mammographic images for derivation of radially symmetric features. A t-distributed Stochastic Neighbor Embedding (t-SNE) method has been utilized to select most relevant features. Using these features, classification experiments have been carried out through all the classifiers. A Logistic Model Tree (LMT) classifier achieves optimal results among all classifiers. An overall comparative analysis has also been made among all our suggested features and feature reduction techniques along with the corresponding classifier where they show superior results.

Keywords: Computer-Aided Diagnosis, DWT, GLCM, DOST, Null-hypothesis SFTA, FCBF, SLT, BLogR, FRST, t-SNE, confusion matrix, ROC curve

Contents

List of Acronyms/Abbreviations

AUC	Area Under Curve
BPNN	Back Propagation Neural Network
CAD	Computer-Aided Diagnosis
CART	Classification And Regression Tree
CC	Carnio-Caudal
CDF	Cumulative Distributed Function
CLAHE	Contrast Limited Adaptive Histogram Equalization
CT	Computed Tomography
DDSM	Digital Database for Screening Mammography
DM	Detail coefficient Matrix
DOST	Discrete Orthonormal S-Transform
DWT	Discrete Wavelet Transform
FCBF	Fast Correlation-based Filter
FD	Feature Descriptor
FDM	Feature Descriptor Matrix
FM	Feature Matrix
FN	False Negative
FNR	False Negative Rate
FP	False Positive
FPR	False Positive Rate
FRST	Fast Radial Symmetry Transform
FT	Fourier Transform
GLCM	Gray-Level Co-occurrence Matrix
IARC	International Agency for Research on Cancer
IRMA	Image Retrieval and Medical Applications
KNN	K Nearest Neighbor
LMT	Logistic Model Tree
MCC	Matthews Correlation Coefficient
MIAS	Mammographic Image Analysis Society
MLO	Medio-Lateral Oblique
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error

NB	Naive Bayes
NF	Normalized Feature
NGLCM	Normalized Gray-Level Co-occurrence Matrix
NPV	Negative Predictive Value
PET	Positron Emission Tomography
PPV	Positive Predictive Value
RF	Random Forest
ROC	Receiver Operating Characteristics
ROI	Region-of-Interest
SFM	Significant Feature Matrix
SFTA	Segmentation-based Fractal Texture Analysis
SLT	Slantlet Transform
SNE	Stochastic Neighbor Embedding
SVM	Support Vector Machine
TM	Training Model
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate
WHO	World Health Organization

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

Biomedical image processing has encountered striking development, and has been an interdisciplinary research field attracting expertise from applied mathematics, computer science, engineering, statistics, physics, biology, and medicine. By the expanding utilization of direct digital imaging frameworks for medical diagnostics, digital image processing turns out to be more and more imperative in health care [1]. Digital medical images display living tissue, organs, or body parts and composed of individual pixels to which discrete brightness or color values are assigned. In the digital biomedical image processing, the physiological structures can be processed and manipulated to visualize hidden characteristic diagnostic features that are difficult to see with film-based imaging methods. Medical image reconstruction and processing require specialized knowledge of a specific medical imaging modality that is used to acquire images. Medical imaging utilizes the techniques to create images of the interior parts of human body and processes for clinical diagnosis, treatment and disease monitoring [1, 2]. The imaging modality means the mode of image acquisition of interior body parts as shown in Figure 1.1. Different imaging modalities are:

X-ray Imaging: In this imaging modality, low-energy X-rays are passed through the body parts and then detected by the detector and image is formed by the analysis of the output of detector with the help of photographic film or digital equipment. The film is exposed to the detected X-rays after passing through the body, will have bright areas (little exposure), gray areas (more exposure) or nearly black areas (heavy exposure) depending upon the amount of X-rays having penetrated in various parts of the body. This modality is used for the diagnosis of breast cancer (mammography), osteoporosis, etc.

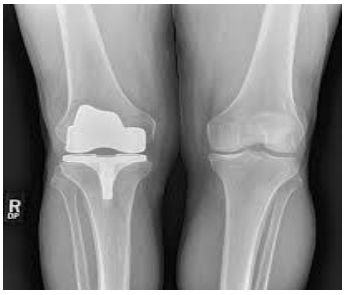
Computed Tomography (CT): In computed tomography, multiple images are acquired as the X-ray tube is moved in an arc above the stationary patient and digital detector. It combines multiple computer-processed X-ray images taken from different angles to produce cross-sectional images of a particular area of a scanned body. This technique is not applicable for soft tissues. Computed tomography is based on the general principle that a finite set of measurements of transmitted X-ray between pairs of points on the surface of an object is sufficient to reconstruct a transverse slice representing the distribution of internal scatterers and absorbers. As light does not travel through human soft tissues in straight lines, imaging technique such as x-ray computed tomography is not applicable. Also soft tissue contrast is very limited compared with CT. The CT method is mostly used for the diagnosis of brain tumors, kidney, liver, lung diseases, etc.

Magnetic Resonance Imaging (MRI): MRI is an imaging technique that includes three main types of equipment, a radio transmitter and receiver, and a computer. It uses a magnetic field and pulses of radio wave energy to make images of organs and structures inside the patient's body. MRI is often divided into structural MRI and functional MRI (fMRI). Structural imaging investigates the structure of the brain and can be used for the diagnosis of large scale intracranial disease, such as tumor, and injury. Functional imaging reveals the activity in certain brain regions by detecting changes in metabolism, blood flow, regional chemical composition, and absorption. The MRI method is very effective for soft tissues. This modality is used for the diagnosis of brain tumors, abdomen organs, osteoporosis, etc.

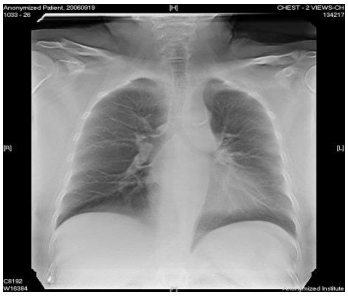
Ultrasonography: It is a medical imaging modality that is based on reflection of ultrasound waves. In this technique, an ultrasound wave travels through the tissue of the human body. At transitions between different muscles and fats, the sound wave is partly reflected and transmitted. The echo runtime indicates the distance between transducer and tissue border while the echo strength is related to material properties. Then, the same transducer is used to detect the echoes, and the image is formed from this pulse-echo signal. The limitations of ultrasonography depend on various factors on its field of view including patient cooperation and physique, difficulty imaging structures behind the bony structures or through organs filled with air, and its dependence on a skilled operator. The choice of frequency of sound wave is also plays a role to generate spatial resolutions of the image. The lower frequencies produce less resolution. Higher frequency sound waves have a smaller wavelength and thus are capable of reflecting or scattering from smaller structures. The ultrasonography modality is used for the diagnosis of prostate, urinary bladder,

uterus, kidney, etc.

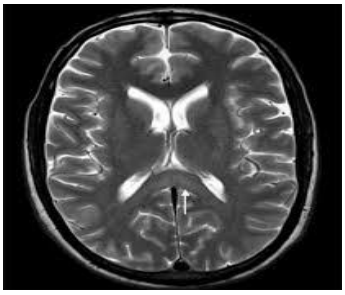
Positron Emission Tomography (PET): The PET imaging technique produces the 3D image of functional processes in the body. In this method, positron-emitting radionuclide tracer is introduced into the body on a biologically active molecule that emits gamma rays. The pairs of gamma rays are detected by the system, and 3D images of tracer concentration within the body are then constructed by computer analysis. This modality is used for the diagnosis of Huntington diseases, Alzheimer diseases, Parkinson diseases, early stage tumor detection, etc.



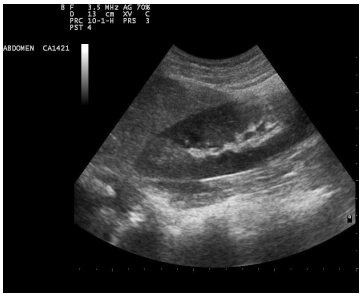
(a) X-ray of knee



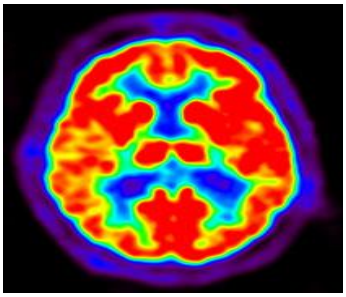
(b) CT of chest



(c) MRI of brain



(d) Ultrasonography of kidney



(e) PET of brain

Figure 1.1: Images of various body parts formed by different imaging modalities.

In this thesis we have investigated on mammograms for early detection of breast cancer. Our subsequent discussion is confined to the topic of research. The chapter is organized as follows:

1.1 Breast Cancer

Across the globe, the most widely recognized cause of cancer related death among women is due to breast cancer. International Agency for Research on Cancer (IARC) of World Health Organization (WHO) has released a press report on 12 December 2013 related to worldwide cancer incidence, mortality and prevalence [3]. According to

this report, 1.7 million women were diagnosed with breast cancer and among them, 522,000 patients died in the year 2012. Since the 2008 assessment, the incident of breast cancer has raised by more than 20% and mortality rate has increased by 14%. This report demonstrates the sharp ascent in breast cancer among women in recent years. In India, the breast cancer is also weighed as the most common cancer among women. For the year 2012, about $144,937$ women were to be affected and $70,218$ patients died among them. It has been observed that one patient ceases to exist of each two newly diagnosed women [4, 5].

Breast cancer is the consequence of the uncontrolled growth of breast cells. The female breast is mainly comprised of lobules (milk-producing glands), ducts (milk passages that connect the lobules to the nipple), fatty and connective tissue surrounding the ducts and lobules, blood vessels, and lymphatic vessels as shown in Figure 1.2. Most breast cancers have their origin in the cells of the ducts, some in the cells of the lobules. The early stage of ductal cancer is referred to as in-situ, implying that the cancer remains confined to the ducts (ductal carcinoma in-situ). When it has invaded the surrounding fatty tissue and possibly has also spread to other organs, it is referred to as invasive [6]. It has been studied that, the recovery of the breast cancer as well as survival rate can be improved by the early detection through periodic screening.

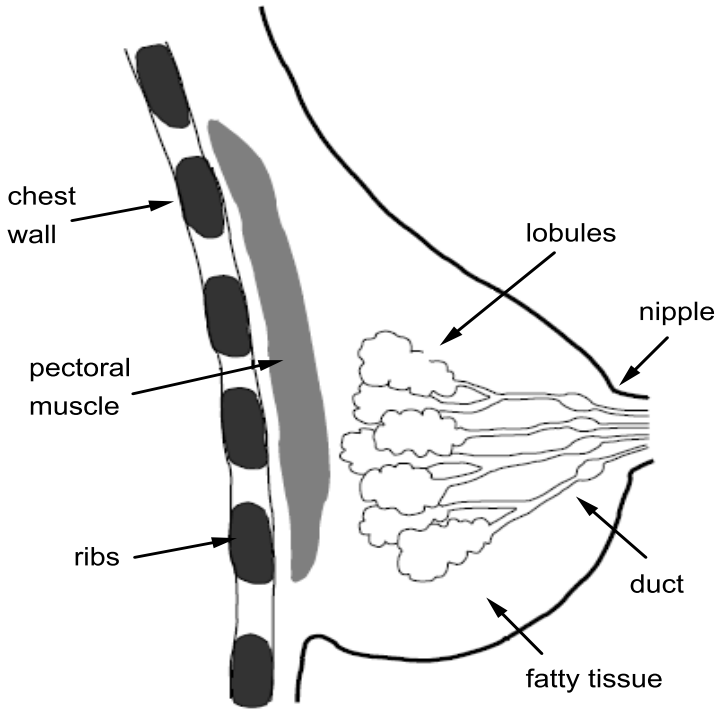


Figure 1.2: Side view of the anatomy and structure of the breast.

To combat the mortality rate due to breast cancer, early detection and treatment is an utmost necessity. Mammography is an efficacious, dependable, and cost-effective method for a precise detection of breast cancer in recent years [7]. Mammography is the procedure of utilizing low-energy X-rays for examination of breast to locate the suspicious lesions. In mammography, a beam of X-rays passes through each breast, where it is absorbed by tissue according to its density. The remaining rays go to a photographic film through the detector and produces a gray-level image after development. The outcome image is known as a film-based mammogram. Again the film-based mammogram can be made digital through film-digitizer. Also the output of the detector from the X-ray scanner directory goes to the digital equipment for development of digital mammogram. The process of digital mammography is described schematically in Figure 1.3. A digital mammographic image is shown in the Figure 1.4 that shows the projected structure of the internal breast. In common practice, there are two projections captured for each breast in mammography: one is Carnio-Caudal (CC) and other is Medio-Lateral Oblique (MLO) shown in Figure 1.5. In the MLO view, the view is taken obligatory during screening in which pectoral muscles appear, but in the CC view of mammogram, the view is taken from head down. In CC view of mammogram, the appearance of pectoral muscle is nil.

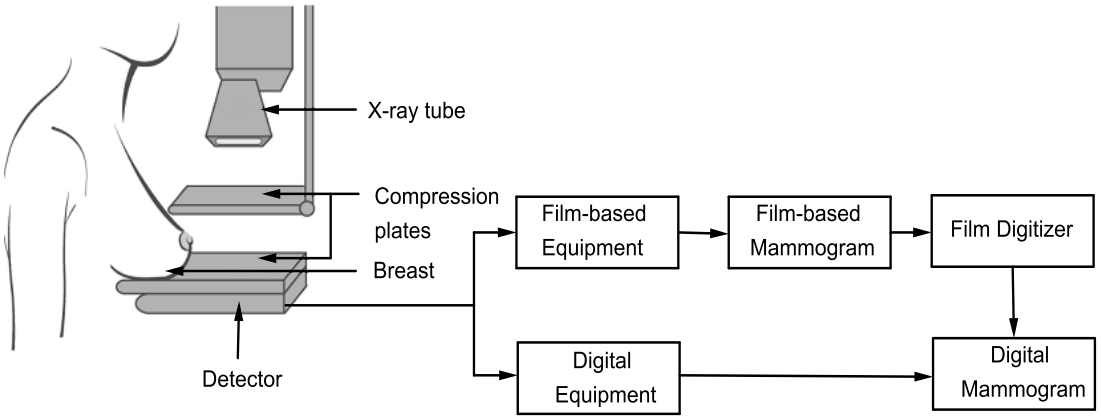


Figure 1.3: Digital mammography process.

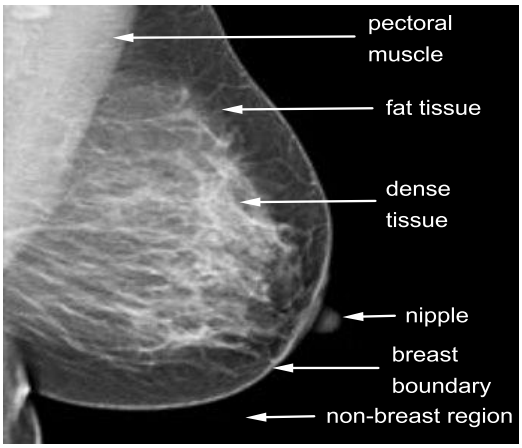
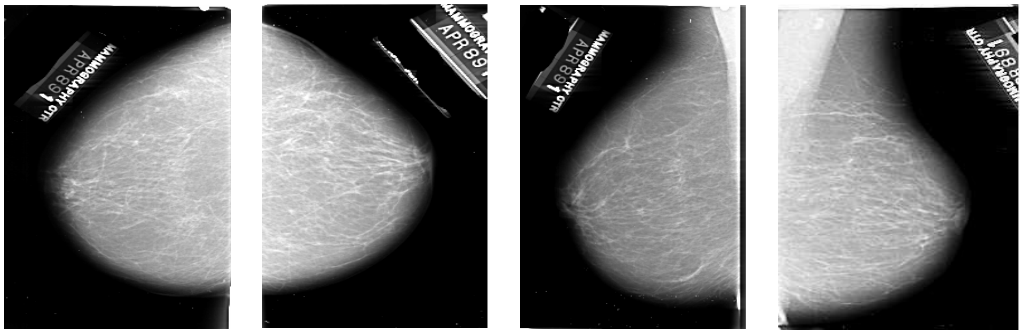


Figure 1.4: Digital mammographic image.



(a) Left CC view (b) Right CC view (c) Left MLO view (d) Right MLO view

Figure 1.5: Two types of view of the breast imaging.

Mammogram interpretation is a vital job for radiologists before suggesting patients for clinical diagnostic tests. However, human interpretation varies as it relies upon training and experience. Mammogram interpretation is a repetitive task which requires maximum attention for evasion of mis-interpretation. It has been noticed that $60 - 90\%$ of the biopsies of human anticipated cancers found benign later [8]. Therefore, Computer-Aided Diagnosis (CAD) system is at present an exceptionally popular and proficient method which analyzes the digital mammograms with the utilization of image processing and pattern recognition techniques.

1.2 Computer-Aided Diagnosis (CAD)

The CAD framework takes care of the abnormality identification issues automatically by collecting and analyzing the significant features from mammographic images. This

system helps radiologists for accurate interpretation of mammograms for the detection and classification of suspicious tissues present in the breast. The blend of CAD scheme and specialist's knowledge would significantly enhance the recognition exactness. The CAD system discriminates among three possible classes i.e., malignant, benign and normal. The CAD process mainly comprises two tasks: the features collected from the image and use of these features in the classification to arrive at a decision. As shown in Figure 1.6, the task of CAD involves several interrelated phases discussed below.

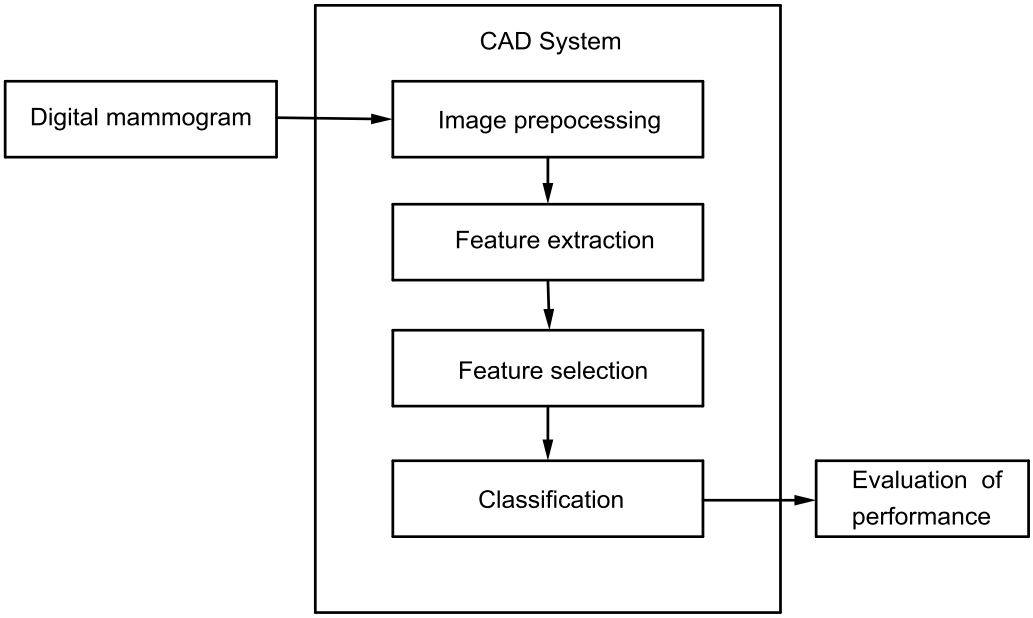


Figure 1.6: Framework of CAD system.

(a) **Image preprocessing:** It is sometimes necessary to modify the data either to correct the deficiencies in the acquired image due to limitations of image acquisition system. In addition, the Region-of-Interest (ROI) that contains the suspicious tissue is extracted from the mammogram by cropping procedure in this phase.

(b) **Feature extraction:** In this phase, features are generated from the mammographic ROIs to use them in the classification task.

(c) **Feature selection:** This task selects the significant features from available feature set that are fed to the classification task. These relevant features influence the efficacy of classification in the discrimination of mammogram classes.

(d) **Classification:** This phase uses a classifier to map a significant feature set to a class type. Such mapping can be specified during training phase to induce the mapping from a collection of feature vector known to be representative of the various classes among which discrimination is being performed (i.e., training set). Once formulated, the mapping can be used to assign an identification of a new unlabeled feature vector subsequently presented to the classifier.

1.3 Performance Measures Used

In the binary classification of abnormal–normal mammograms, the abnormal (cancerous) samples are denoted as the **positive class** while the normal samples are denoted as the **negative classes**. Similarly, for malignant-benign mammogram classification, malignant samples are considered as the **positive class** and benign samples are considered as the **negative classes**. The performance of the classifier is evaluated with the help of a confusion matrix as shown in Table 1.1 that summarizes the number of samples predicted correctly or incorrectly by the classifier [9, 10].

Table 1.1: Confusion Matrix for binary classification system.

Actual class	**Predicted class**	
	Positive	Negative
Positive	True positive (TP)	False negative (FN)
Negative	False positive (FP)	True negative (TN)

To evaluate the performance of the classifier, several performance measures can be used with the help of the entries of the confusion matrix:

(a) The **true positive rate** (TPR) or **sensitivity** (S_n) is defined as the fraction of positive samples predicted correctly by the model, i.e.,

$$TPR = TP/(TP + FN). \tag{1.1}$$

(b) The **false positive rate** (FPR) is defined as the fraction of negative samples predicted as a positive class, i.e.,

$$FPR = FP/(TN + FP). \tag{1.2}$$

(c) The **true negative rate** (TNR) or **specificity** (S_p) is defined as the fraction of negative samples predicted correctly by the model, i.e.,

$$TNR = 1 - FPR \quad \text{or} \quad TNR = TN/(TN + FP). \tag{1.3}$$

(d) The **false negative rate** (FNR) is defined as the fraction of positive samples predicted as a negative class, i.e.,

$$FNR = FN/(TP + FN). \tag{1.4}$$

(e) **Precision** (p) or **positive predictive value** (PPV) determines the fraction of samples that actually turns to be positive in the group the classifier has declared as a positive class and defined as,

$$p = TP/(TP + FP). \tag{1.5}$$

(f) **Recall** (r) measures the fraction positive samples correctly predicted by the classifier. It is equivalent to the TPR.

(g) The **negative predictive value** (NPV) determines the fraction of samples that actually turns to be negative in the group the classifier has declared as a negative class and is given by,

$$NPV = TN/(TN + FN). \tag{1.6}$$

(h) The **accuracy** (ACC) determines the proportion of the true results of the total number of samples tested. i.e,

$$ACC = (TP + TN)/(TP + FP + FN + TN). \tag{1.7}$$

(i) The **F1 score** (F_{score}) is the measure of test accuracy and defined as the weighted average of the precision (p) and recall (r), i.e.,

$$F_{score} = (2 \times p \times r)/(p + r). \tag{1.8}$$

(j) The **Matthews correlation coefficient** (MCC) determines the quality of the binary classification. It is defined as a correlation coefficient between the observed and predicted binary classification and given as,

$$MCC = \frac{((TP \times TN) - (FP \times FN))}{\sqrt{((TP + FP)\,(TP + FN)\,(TN + FP)\,(TN + FN))}}. \tag{1.9}$$

The MCC returns the value of -1, 0 and $+1$. A coefficient of $+1$ represents a perfect prediction, 0 no better than random prediction and 1 indicates total disagreement between prediction and observation.

The evaluation of a classifier performance can also be accomplished by means of **Receiver Operating Characteristics (ROC)** curves [8]. It is a two-dimensional plot of **true positive rate (sensitivity)** versus **false positive rate** (1-**specificity**) in vertical and horizontal axes respectively as shown in Figure 1.7. The area under the ROC curve referred by an index AUC is an important factor for evaluating the classifier performance. The value of AUC is 1.0 is a perfect performance of the classifier.

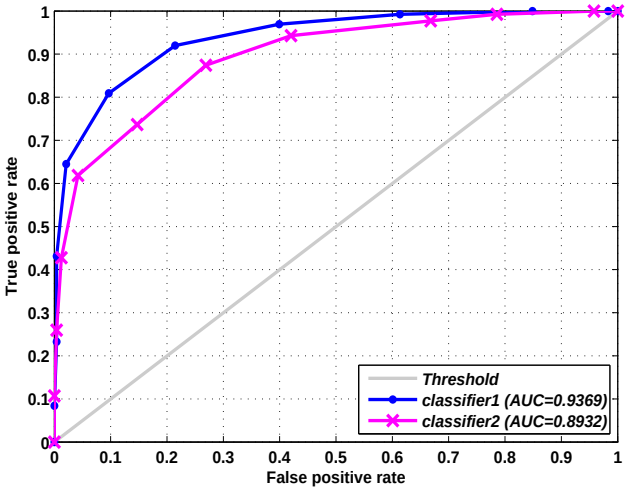


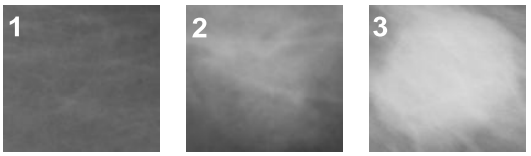
Figure 1.7: Typical ROC curves for two different classifiers in the classification of mammograms.

1.4 Database Used

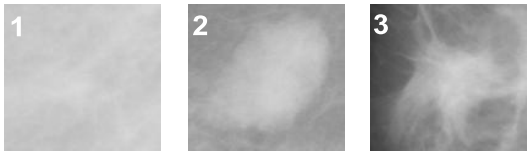
To validate the proposed schemes, mammographic images are taken from two databases namely, Mammographic Image Analysis Society (MIAS) database [11] and Image Retrieval and Medical Applications (IRMA) project [12]. The MIAS database is built by Suckling *et al.* and openly available for scientific research. The mammographic image database in IRMA project is made by Deserno *et al.*, who collected images from several other databases including Digital Database for Screening Mammography (DDSM). Both MIAS and IRMA provide appropriate information based on types of background tissues, and the class of abnormalities present in the mammograms. The class of abnormality consists of abnormal–normal class, and again based on the severity of abnormality; the abnormal class is divided into two sub-classes such as malignant and benign. The MIAS database contains 322 images, which are categorized according to tissue types like fatty, fatty-glandular and dense-glandular.

In IRMA project, the database is divided into 12 and 20 class problems. In 12 class problem, the mammograms are categorized according to tissue density, and each category is divided into three classes; normal, benign and malignant. In 20 class problem, the mammograms are of two categories of different types of lesions. The 12 class database consists of mammograms of four tissue types; almost entirely fatty, scattered fibro glandular, heterogeneously dense and extremely dense. This database consists of 2796 images out of which 2576 images are from DDSM database. Figures 1.8 and 1.9 show various regions-of-interest (ROIs) containing different classes of abnormality.

We have considered all 322 images from MIAS database for our experiments from this database. Out of 322 images, 207 images are normal, 115 images are abnormal; again among abnormal images the number of benign and malignant types are 64 images and 51 respectively. Also, a total of 1000 DDSM images from 12 class problem have been taken, out of which 500 images are normal and 500 images are abnormal. The abnormal class consists of 236 benign images and 264 malignant images. Each mammographic ROI has been taken of size 128×128 pixels used in the feature extraction phase to find the feature elements.



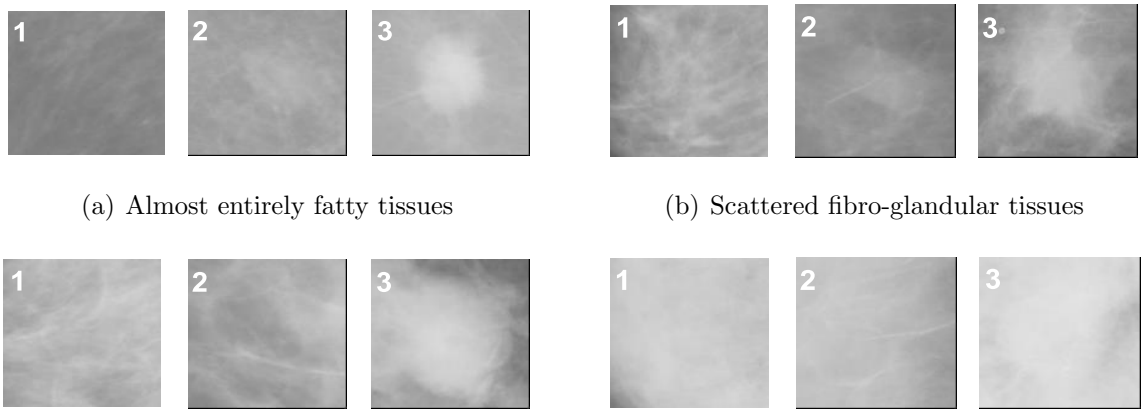
(a) Fatty tissues



(b) Fatty-glandular tissues

(c) Dense-glandular tissues

Figure 1.8: Mammographic ROIs of MIAS database. The sub-figures indicate different types tissues present in mammograms. The labels 1, 2 and 3 of ROIs represent normal, benign and malignant classes respectively.



(a) Almost entirely fatty tissues

(b) Scattered fibro-glandular tissues

(c) Hetereogeneously dense tissues

(d) Extremely dense tissues

Figure 1.9: Mammographic ROIs of DDSM database from IRMA project. The sub-figures indicate different types tissues present in mammograms. The labels 1, 2 and 3 of ROIs represent normal, benign and malignant classes respectively.

1.5 Related Work

Many researchers have worked to develop the automated recognition system for earlier screening of breast cancer. Dhawan *et al.* have proposed a mammogram classification scheme to predict the malignancy property of the tissues [13]. They have defined two categories of correlated gray-level image structure features for classification of difficult-to-diagnose cases. The first category of features includes second-order histogram statistics-based features representing the global texture and the wavelet decomposition based features representing the local texture of the microcalcification area of interest. The second category of features represents the first-order gray-level histogram based statistics of the segmented microcalcification regions, size, number, and distance of the segmented microcalcification cluster. Various features in each category were correlated with the biopsy examination results of 191 difficult-to-diagnose cases for selection of the best set of features representing the complete gray-level image structure information. The selection of the best features was performed using the multivariate cluster analysis as well as a genetic algorithm (GA)-based search method. The selected features have been used for classification using Back-Propagation Neural Network (BPNN) and parametric statistical classifiers. ROC analysis has been performed to compare the neural network-based classification with linear and K-Nearest Neighbor (K-NN) classifiers. The performance index value of the classification, AUC of 0.81 has been yielded by

neural network classifier.

Wei *et al.* have achieved AUC of 0.96 through ROC analysis in the classification of 168 abnormal–normal mammograms by using multiresolution texture features [14]. In their method, wavelet transform has been used to decompose the mammographic ROI to collect different detail coefficients and consequently, texture features were extracted from these coefficients. Linear discriminant models have been used to select effective features from the global, local, or combined feature spaces were established to maximize the separation between masses and normal tissue. Liu *et al.* have used linear phase non-separable two-dimensional wavelet transform to extract features from mammographic ROIs. They have found accuracy rate of 84.2% on true positive detection in the classification of mammograms from MIAS database by using binary classification tree [15]. Ferrari *et al.* have proposed a classification approach based on the multiresolution analysis of mammographic images [16]. The method utilizes Gabor wavelets to find the linear directional components of mammograms. The most relevant directional elements are selected using KL transform. The scheme has achieved an average classification accuracy of 74.4% on MIAS database using Bayesian linear classifier. Zhen *et al.* have designed an algorithm that comprises many artificial intelligent strategies and discrete wavelet transform (DWT) for detection of abnormalities in mammograms [17]. The categorization of mammograms as cancerous or normal has been performed by the use of tree-type classification technique. The algorithm has been validated using 322 mammograms of MIAS database and a performance result concerning sensitivity of 97.3% has been obtained.

M. Masotti has developed a method to extract the features by multiresolution analysis of mammograms using ranklet-based transform [18]. A classification performance index value, $AUC = 0.978$ has been obtained in the classification of abnormal–normal tissues for DDSM database. Mavroforakis *et al.* have proposed a method to characterize the breast tissue based on the texture analysis of mammograms. They have employed a fractal analysis to analyze the textural features and achieved 83.9% of performance score through SVM classifier [19]. Martins *et al.* have applied Gray-Level Co-occurrence Matrix (GLCM) to extract the features from mammographic images [20]. The forward selection technique has been employed to select the most significant features. Then, a Bayesian neural network has been used to evaluate the ability of these features to predict the class for each tissue sample into malignant, benign and normal. The method was tested on a set of 218 tissues samples of MIAS database, 68 benign and 51 malignant and 99 normal, and a classification accuracy of 86.84% has been achieved. Sakellaropoulos *et al.* have

used wavelet-based feature analysis for differentiating masses, of varying sizes, from normal dense tissue on mammographic images [21]. The images analyzed was from DDSM database consists of 166 ROIs containing spiculated masses (60), circumscribed masses (40) and normal dense tissue (66). A set of ten multiscale features, based on intensity, texture and edge variations, were extracted from the ROIs sub-images provided by the wavelet transform. Logistic regression analysis was employed to determine the optimal multiscale features for differentiating masses from normal dense tissue. The classification accuracy in differentiating circumscribed masses from the normal dense tissue is comparable with the corresponding accuracy in differentiating spiculated masses from normal dense tissue, achieving AUC values of 0.895 and 0.875, respectively.

Rashed *et al.* have obtained an average accuracy of 84.16% in the prediction of malignancy of mammograms from MIAS database [22]. Texture features were extracted from mammographic ROIS by decompositions based on three different wavelets, Daubechies-4, Daubechies-8, and Daubechies-16. The Euclidean distance has been used to design the classifier based on calculating the distance between the feature vectors of testing ROIs and the precomputed class core vector. Pereira *et al.* have proposed a method in which spatial gray-level dependence matrix of the wavelet transformed mammograms has been used to derive the texture features [23]. These texture features were utilized to classify the mammograms as malignant or benign with the help of non-parametric K-NN classifier. Different mammograms from DDSM database were used in their experiment. The AUC values of 0.973, 0.607, and 0.617 have been achieved for discriminating the abnormal–normal ROIs, malignant–benign microcalcification, and malignant–benign masses, respectively. Khademi *et al.* have utilized a shift-invariant wavelet transform to define the texture features of the mammographic images in their proposed method for classification of mammograms [24]. Gray level co-occurrence matrices are found for a variety of directions in the wavelet domain, and homogeneity and entropy were extracted which produces a shift, scale, and semi-rotational invariant feature set. Exhaustive feature selection was used with both a K-NN and LDA classifier, to find the best classification performance. They found the optimum classification accuracy of 72.5% by using LDA classifier. Dong *et al.* have used Gabor filter for the classification of normal and abnormal mammograms and achieved an average of 80% precision with selected features [25]. Dua *et al.* have developed a method to classify the mammograms using a unique weighted association rule based classifier [26]. In their method, texture components were extracted from segmented parts of the image and discretized for

rule discovery. Association rules were derived between various texture components extracted from segments of images and employed for classification based on their intra- and inter-class dependencies. These rules were then employed for the classification of mammograms collected from MIAS database, and an accuracy of 89% has been achieved.

Prathibha *et al.* have used multiscale wavelet transformation for extraction of texture features from the mammographic images. They have obtained the classification performance as AUC of 0.946 in ROC analysis to classify normal and abnormal mammograms of MIAS database by using the statistical classifier [27]. Verma *et al.* have used BI-RADS descriptor features to classify the malignant and benign mammograms utilizing the proposed Soft Clustered Based Direct Learning (SCBDL) classifier [28]. They have achieved an accuracy of 97.5% on DDSM database. Moayedi *et al.* have developed a scheme for automatic mass classification of mammograms by using contourlet transform for extraction of features [29]. A genetic algorithm has been utilized in their scheme to choose most discriminative texture feature set from the available extracted features. They have accomplished 96.6% of classification accuracy on MIAS dataset with the assistance of Successive Enhancement Learning (SEL) weighted Support Vector Machine (SVM). Cao *et al.* have proposed a mammogram classification scheme based on forty-two features including shape, intensity, texture, age etc., extracted from each segmented mass [30]. The Support Vector Machine (SVM) has been employed for the characterization of the mammograms as malignant or benign using DDSM database and the AUC of 0.948 has been achieved. Buciu *et al.* have developed a method to discriminate malignant, benign and normal mammograms for the early detection of breast cancer [31]. A Gabor wavelet has been applied to get features from mammograms in different orientations and frequencies. Principal Component Analysis (PCA) has been utilized to reduce the dimension of extracted feature set, and a proximal SVM has been used to classify the dataset. The scheme is evaluated on MIAS database and performance results in terms AUC values of 0.79 and 0.78 are attained in the classification of abnormal-normal and malignant-benign mammograms respectively.

Mutaz *et al.* have developed a method in which the textural features were extracted from ROI using GLCM [32]. Utilizing these features, they have discriminated the malignant and benign mammograms with the help of neural network and achieved the sensitivity of 91.67% and specificity of 84.17% on DDSM database. Fraschini has used discrete wavelet transform and neural network to classify the mammograms [33]. The performance index value of AUC of 0.91 has been yielded

using DDSM database in the analysis of ROC curve. Tahmasbi *et al.* have designed a mammogram diagnostic approach by utilizing the Zernike moments as feature descriptors [34]. The Multi Layer Perception (MLP) technique has been employed to classify the mammogram as malignant or benign and the *AUC* of 0.976 has been obtained on MIAS database. Biswas *et al.* have proposed a two-layered model for identification of architectural distortions in mammograms [35]. In the first layer of their model, a multiscale filter bank has been intended to generate texture descriptors from mammographic Region-of-Interest (ROI). The inferred features are represented as a set of textural primitives by the mixture of Gaussian distributions. An Expectation-Maximization (EM) algorithm has been employed to learn these texture patterns. They have achieved classification accuracies of 82.5% and 88.3% on MIAS and DDSM database respectively. The ROC analysis of classification has additionally been completed, and *AUC* values of 0.83 and 0.87 have been found on similar platforms. Tsai *et al.* have developed an efficient algorithm for the diagnosis of the breast cancer based on the mammographic image reconstruction and identification of microcalcification [36]. For this purpose, wavelet transform and Renyis information theory have been used to distinguish the suspicious ROI from normal tissues. The scattered regions of microcalcification have been reconstructed by utilizing a morphology dilation and majority voting rule. The scheme uses forty-nine feature descriptors namely shape inertia, compactness, eccentricity and Gray-Level Co-occurrence Matrix (GLCM) to specify the patterns of the suspicious microcalcification clusters. PCA has been employed to select the most significant descriptors for achieving the optimal results in the classification task that was performed by BPNN. The proposed scheme has been applied to the real clinical patients at National Cheng-Kung University Hospital, Taiwan, and a sensitivity value of 97.19% was obtained.

Jona *et al.* have used GLCM to extract the features from the mammographic images [37]. They have optimized the feature set by employing a hybrid particle swarm optimization and genetic algorithm, and obtained 94% of classification accuracy by using SVM to classify the abnormal and normal mammograms on MIAS database. Ramos *et al.* have explored on the abnormal-normal mammogram set classification using different methods namely ridgelet transform, GLCM and DWT for extraction of features [38]. The best significant feature set has been selected by utilizing Genetic Algorithm (GA). A maximum classification result has been obtained through Random Forest with the help of DWT and GA that gives an *AUC* value of 0.90 using DDSM database. Eltoukhy *et al.* have proposed a scheme for classification of mammograms

in which multiresolution techniques, wavelet and curvelet transform have been used to extract the features from mammographic ROIs [39]. The most significant features were selected by applying statistical t-test method upon the available derived feature set. A 5-fold cross-validation technique has been used with the help of SVM for the classification of mammograms from MIAS database. The optimal classification accuracies of 95.98% and 97.30% have been achieved for abnormal–normal and malignant–benign class respectively, using curvelet transform. Muštra *et al.* have proposed a mammogram classification method for the detection of abnormalities by breast density measurement [40]. They have observed the breast density as textures and used GLCM method to extract the textures taking into account gray-scale features of first and second order. Two databases, MIAS, and KB-FER have been tested by this method, and an optimal result have been found for BI-RADS two category case. A maximum classification accuracy of 91.6% has been achieved on the MIAS database by using Best-First Backward feature selection method and Naive Bayes classifier. Similarly, an accuracy of 97.2% has been obtained with the use of Best-First Forward feature selection method and K-NN classifier on KBD-FER digital mammography database of the University Hospital Dubrava, Zagreb, Croatia. Nanni *et al.* have proposed a mammogram classification system based on the Local Ternary Pattern (LTP) features [41] and found the AUC of 0.97. A Neighborhood Preserving Embedding (NPE) method has been used to produce the high variance features that are further provided to the classifier. An SVM has been employed to classify the mammogram as malignant or benign using DDSM database.

Görgel *et al.* have proposed a scheme to classify the mammogram using spherical wavelet transform (SWT) for extraction of features and SVM as the classifier [42]. In their proposed method, a local seed region growing algorithm has been used to detect ROIs of mammograms. The proposed scheme achieves 96% and 93.59% accuracy in mass–non-mass classification and malignant–benign classification respectively when using the Istanbul University (I.U.) database with k-fold cross-validation. Nascimento *et al.* have developed a scheme that uses DWT to extract the features and a polynomial classifier to discriminate the malignant–normal, benign–normal and malignant–benign mammogram sets [43]. Classification performance measures concerning AUC values of 0.98, 0.95 an 0.96 have been achieved for the respective mammogram sets using DDSM database. Kumar *et al.* have proposed a method based on the combination of DWT and Stochastic Neighbor Embedding technique for benign and malignant mammogram classification [44]. They have used Stochastic Neighbor Embedding technique to reduce wavelet coefficients of mammograms and SVM as

classifier. The method has achieved classification accuracies of 93.39% and 92.10% to classify normal–abnormal and benign–malignant mammograms, respectively. Oral *et al.* have used first order and second order textural feature to classify the mammograms as abnormal or normal. Principal component analysis (PCA) has been used in their method to reduce the dimension of feature spaces and an accuracy of 91.1% is achieved on MIAS database by multi layer perception (MLP) classifier [45]. Liu *et al.* have investigated on the classification of malignant–benign mammograms using selected geometry and texture features [46]. Maximum performance results with respect to the accuracy of 94% and AUC of 0.9615 with a leave-one-out scheme on DDSM database have been demonstrated. The optimum results have been accomplished by using the SVM based Recursive Feature Elimination (SVM-RFE) procedure with a Normalized Mutual Information Feature Selection (NMIFS) method.

Ganesan *et al.* have found a maximum accuracy of 92.48% by applying one-class classification on the set of mammograms provided by the Singapore Anti-Tuberculosis Association CommHealth (SATA) [47]. A trace transform functional has been used in the scheme to extract the features from mammograms. A Gaussian Mixture Model (GMM) has been engaged for the classification of the malignant-benign mammograms. Reyad *et al.* have proposed a scheme to extract features from mammograms by using different strategies namely Local Binary Pattern (LBP), statistical measure and multiresolution frameworks [48]. Texture descriptors and statistical features were derived by LBP and statistical methods respectively, whereas multiresolution features were extracted by DWT and contourlet transform. SVM has been utilized for the classification of abnormal–normal mammograms from DDSM database by using these extracted features. A classification accuracy of 98.43% has been achieved using statistical or LBP features. Subsequently, an improved accuracy of 98.63% has been accomplished by using the combination of both LBP and statistical features that outperform the contourlet and wavelet transform based method. Diaz *et al.* have proposed an approach in which, the morphological algorithms are applied to detect the microcalcification in the mammograms [49]. An SVM with Gaussian kernel has been used to distinguish the mammograms as abnormal or normal on MIAS database utilizing a set of spatial, texture and spectral features and achieved the AUC of 0.976. A mammogram classification scheme is designed by the Kim *et al.* to discriminate the spiculated malignant masses from normal tissues and the AUC of 0.956 has been obtained on DDSM database [50] . In this approach, region-based stellate features are determined by computing the statistical characteristics of three subregions, namely, core, inner, and outer parts of an ROI. The SVM has been employed for classification

using relevant set of features chosen by AdaBoost learning.

Görgel *et al.* have proposed a Spherical Wavelet Transform (SWT) based mammogram classification method for automatic detection of breast cancer [51]. The scheme extracts shape, boundary, and gray-level based feature of wavelet from mammographic ROIs. SVM has been employed to classify the benign–malignant masses which attains an accuracy of 91.4% on Istanbul University hospital database, Turkey and 90.1% on MIAS database. Li *et al.* have found an accuracy of 85.96% for the classification of malignant-benign mammograms using DDSM database [52] and their scheme deals on the analysis of texton based mammogram textures with multiple subsampling strategies. Each of the subsampling strategies catches a discriminating structure used in the classification phase. A K-NN classifier has been employed to attain the expected optimum accuracy. Rouhi *et al.* have proposed a scheme to discriminate mammogram mass type as benign or malignant [53]. In the first method of the scheme, segmentation has been performed using an automated region growing utilizing a threshold obtained from trained Artificial Neural Network (ANN). In the second method of the scheme, a Cellular Neural Network (CNN) has been utilized for segmentation using Genetic Algorithm (GA). Intensity, textural, and shape features were extracted from segmented ROIs by thresholding, GLCM and Zernike moments, respectively. GA has been used to select relevant features from the set of extracted features. ANN has been employed to classify the mammograms as benign or malignant. Experiments have been carried out on MIAS and DDSM databases, and optimal accuracy values of 96.47% and 90.6% have been achieved respectively.

Korkmaz *et al.* have proposed a diagnostic method to classify the mammograms as malignant, benign or normal [54]. In this methodology, a set of texture features including sum average, difference variance, kurtosis, skewness, entropy inverse difference moment, contrast, local homogeneity, cluster prominence and maximum probability are extracted and utilized. An mRMR (minimum-Redundancy-Maximal-Relevance) technique has been used to select significant values of the features. The mammograms are classified with the help of KL (Kullback-Leibler) classifier using the DDSM database and an accuracy (ACC) of 93.8% has been achieved. Jiang *et al.* have developed a CBIR (Content-Based Image Retrieval)-based CAD for correct identification of the mammographic ROI as a mass or normal by utilizing SIFT features with the help of a vocabulary tree [55]. In their approach, weighted majority vote technique has been applied to classify the mammograms collected from the DDSM database, and an accuracy (ACC) of 90.8% is obtained. Dhahbi *et al.* have used the curvelet transform

and moment theory in succession to extract two types of features namely, Curvelet Level Moment (CLM), and Curvelet Band Moment (CBM) from mammograms [56]. A *t-test* ranking technique has been applied to select most relevant feature sets. The K-NN is used to classify the mammograms from MIAS and DDSM databases into two classes, abnormal–normal and malignant–benign. The accuracy values of 91.27% (abnormal–normal), 81.35% (malignant–benign) has been achieved for MIAS database. Similarly, the values are 86.46% and 60.43% for DDSM database has been found in their methodology. Murat Karabatak has proposed a new weighted Naive Bayesian classifier to characterize the mammograms as malignant or benign [57]. He has used the Wisconsin breast cancer database that includes 699 records and each record has nine number of features and achieved the classification accuracy of 96.02%.

Xie *et al.* have presented a CAD system in which a total of 32 gray-level and texture features are extracted from mammograms in the feature extraction phase [58]. A combination of SVM and ELM (Extreme Learning Machine) has been used for the elimination of insignificant features. The ELM, which is a single hidden layer feed-forward network, has been employed to classify mammograms by utilizing the optimal subset of relevant features. They have achieved the accuracy (ACC) of 96.02% and AUC of 0.9659 in the classification of malignant and benign mammograms on MIAS database. Oliveira *et al.* have proposed a method to classify mammographic mass or non-mass regions using the taxonomic indices as texture features and found an accuracy (ACC) of 98.88% [59]. The taxonomic diversity and distinctness indexes are computed with the use of phylogenetic tree considering two spatial approaches namely, internal and external masking. An SVM has been used to classify the mammograms from DDSM database utilizing the computed taxonomic indices. Zhang *et al.* have proposed method to discriminate the malignant masses from benign masses [60]. The fractional Fourier transform has been employed to obtain the unified time–frequency spectrum coefficients which are reduced by principal component analysis (PCA). The have achieved sensitivity (S_n) of 92.22%, specificity (S_p) of 92.10%, and accuracy (ACC) of 92.16% using SVM as classifier on MIAS database.

1.6 Motivation

It has been observed from the literature study that the relevant features play a vital role in the successful classification of mammograms as normal, benign or malignant. Texture based features are predominant in the existing schemes and mostly when multiresolution transform and its variants for classification, neural network and SVM

have been mostly used. The existing schemes have been validated either on MIAS or DDSM but not on both. Considering the existing literature and importance of the topic, it has been realised that there exists an abundant scope to suggest new features and feature reduction schemes along with improved classifiers to enhance performances.

1.7 Research Objectives

The prime objective is to reduce the variability in judgments among radiologists by providing an accurate diagnosis of cancer using digital mammograms. Therefore, the objectives are narrowed down to

1. develop features using Segmentation-based Fractal Texture Analysis (SFTA), Discrete Orthonormal S-Transform (DOST), Slantlet Transform (SLT), and Fast Radial Symmetry Transform (FRST),
2. develop hybrid features using Discrete Wavelet Transform (DWT), and Gray-Level Co-occurrence Matrix (GLCM),
3. select significant features using null hypothesis with statistical *t-test*, Fast Correlation-Based Filter (FCBF), Bayesian Logistic Regression (BLogR), and t-distributed Stochastic Neighbor Embedding (t-SNE) method, and
4. devise classifiers using Back-Propagation Neural Network (BPNN or FNN), Support Vector Machine (SVM), AdaBoost and Random Forest (AdaBoost-RF), LogitBoost and Random Forest (LogitBoost-RF), and Logistic Model Tree (LMT).

1.8 Classifier Used

In order to validate the efficacy of the proposed feature and feature selection techniques, various classifiers are devised and used employing Back-Propagation Neural-Network (BPNN or FNN), Support Vector Machine (SVM), ensemble classifiers like AdaBoost and LogitBoost using Random Forest, and Logistic Model Tree (LMT). The achieved results have been compared among the devised classifiers as well as with other standard classifiers namely, Naive Bayes (NB) and K-Nearest Neighbor (K-NN) for the validation of proposed work. The description of each classifier are given below in brief.

1.8.1 Back-Propagation Neural-Network (BPNN or FNN)

Artificial neural network is a powerful parallel dynamic system consisting of multiple simple and interconnected processing units (nodes), that performs tasks like the biological brains. The nodes in a neural network architecture are commonly known as neurons. In the architecture of the neural network, each input node is connected via a weighted link to the output node. The weighted link is used to emulate the strength of the synaptic connection between neurons. A neural network can perform the necessary transformation operation automatically with the aid of neuron's state response to their input information. These networks are trained with a set of samples known as the training set. The network is trained by learning the values of its internal parameter from the training set so that, an input leads to a specific output.

A feed-forward Back-Propagation three-layered Neural Network (BPNN or FNN) as depicted in Figure 1.10 is one of the most common and efficient network structures used for classification in the feature space. This network has an intermediary layer known as hidden layer present with input and output layer. The hidden layer is composed of H hidden nodes. A set of R selected significant feature vectors ($x_i,\ i = 1, 2, ..., R$) are input to BPNN for the classification. The output with reduced error is to be expected for better performance. For this purpose, BPNN possesses two phases in each iteration: forward phase and backward phase. During the forward phase, the weights obtained from the previous iteration are used to compute the output value of each neuron in the network. The computation progresses in the forward direction. During the backward phase, the weights are updated in the reverse direction. The errors for neurons at current layer are used to estimate the errors for neurons at the previous layer.

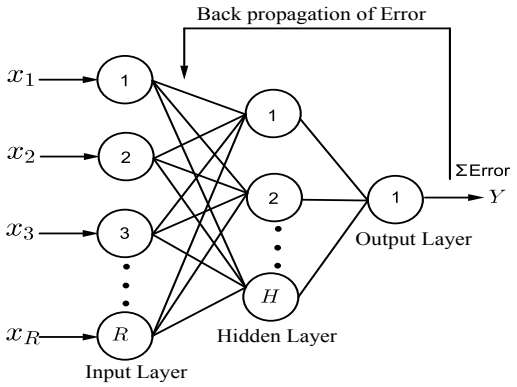


Figure 1.10: Model of a 3-layered feed-forward BPNN or FNN.

1.8.2 Support Vector Machine (SVM)

The Support Vector Machine is employed as the classifier that is based on the statistical learning theory shows promising empirical results in the characterization of mammograms [61]. In SVM, the feature vector x of an object is classified by looking at the sign of a linear scoring function $\langle x, w\rangle$. The objective of learning is to estimate the parameter $w \in \mathbb{R}^d$ in such a way that the score is positive if the x belongs to the positive class and negative otherwise. Thus, a margin is imposed between positive and negative class instances in the formulation of SVM. The parameter w is estimated by fitting the scoring function to a training dataset of K instances $\{x_i, y_i\},\ i = 1, 2, ..., K$. The $y_i \in \{-1, +1\}$ is the class label of the corresponding instance vector x_i. Then a loss function is computed that measures the fit quality, and given as,

$$\ell_i < w, x >= \max\left\{0, 1 - y_i \left\langle w, x\right\rangle\right\}. \tag{1.10}$$

Fitting the training data is usually insufficient. In order for the scoring function generalize to future data as well, it is usually preferable to trade off the fitting accuracy with the regularity of the learned scoring function $\langle x, w\rangle$. Regularity in the standard formulation is measured by the norm of the parameter vector $\|w\|^2$. Averaging the loss on all training instances and adding to it the regularizer weighed by a parameter λ yields the regularized objective function that is given by,

$$E\left(w\right) = \frac{\lambda}{2}\|w\|^2 + \frac{1}{K}\sum_{i=1}^{K} \max\left\{0, 1 - y_i \left\langle w, x\right\rangle\right\}. \tag{1.11}$$

This objective function is convex for which there exists a single global optimum. So far only the linear scoring function $\langle x, w\rangle$ have been considered. Implicitly, however, this assumes that the objects (images) to be classified have been encoded as vectors x in such a way that makes linear classification possible. This encoding step can be made explicit by introducing the feature map $\Phi\left(x\right) \in \mathbb{R}^d$. Including the feature map, a non-linear scoring function in x is yielded and given as,

$$x \in \mathcal{X} \rightarrow \left\langle \Phi\left(x\right), w\right\rangle \tag{1.12}$$

where $\mathcal{X}$ is the input sample space. The relation of feature maps to similarity functions is formalized by the notion of a kernel, a positive definite function $k\left(x, x'\right)$ measuring the similarity of a pair of objects. A feature map defines a kernel by,

$$k\left(x, x'\right) = \left\langle \Phi\left(x\right), \Phi\left(x'\right)\right\rangle. \tag{1.13}$$

In original SVM, two parallel planes are formed in a way that each plane is closest to one of the two populations belonging to two classes, and two planes are as far apart as possible. In generalized eigenvalue proximal SVM, two optimal nonparallel planes are generated which achieves the enhanced classification performances [62]. In nonparallel SVM, two non parallel planes are formed in the form of

$$w_1^T x - b_1 = 0 \quad \text{and} \quad w_2^T x - b_2 = 0 \tag{1.14}$$

To obtain the first plane,the following solution is derived from (1.14)

$$(w_1, b_1) = \arg\min_{(w,b)\neq 0} \frac{\left\|w^T X_1 - o^T b\right\|^2 \Big/ \|z\|^2}{\|w^T X_2 - o^T b\|^2 \Big/ \|z\|^2} \tag{1.15}$$

$$z \leftarrow \begin{bmatrix} w \\ b \end{bmatrix} \tag{1.16}$$

where, X_1 and X_2 are set of samples belong to class 1 and class 2, respectively. The parameter q is a vector of one of appropriate dimensions. Simplifying (1.15) gives

$$\min_{(w,b)\neq 0} \frac{\left\|w^T X_1 - o^T b\right\|^2}{\|w^T X_2 - o^T b\|^2} \tag{1.17}$$

A Tikhonov regularization term is included to decrease the norm of the variable z that corresponds to the first hyperplane in (1.14)

$$\min_{(w,b)\neq 0} \frac{\left\|w^T X_1 - o^T b\right\|^2 + t\|z\|^2}{\|w^T X_2 - o^T b\|^2} \tag{1.18}$$

where, t is a positive (or zero) Tikhonov factor. Equation (1.18) turns to the Rayleigh quotient in the following form of

$$z_1 = \arg\min_{z\neq 0} \frac{z^T P z}{z^T Q z} \tag{1.19}$$

where, P and Q are symmetric matrices in $\mathbb{R}^{(p+1)(p+1)}$ as

$$P \stackrel{def}{=} [X_1 - 0]^T [X_1 - 0] + tI \tag{1.20}$$

$$Q \stackrel{def}{=} [X_2 - 0]^T [X_2 - 0] + tI \tag{1.21}$$

Using the stationarity and boundedness properties of Rayleigh quotient, solution of (1.19 is deduced by solving a generalized eigenvalue problem as,

$$Pz = \lambda Q z, \quad z \neq 0 \tag{1.22}$$

where, z is eigenvector and λ is eigenvalue for both the hyperplanes correspond to populations of class 1 and 2 samples.

1.8.3 Ensemble Classifiers

In machine learning, ensemble methods make use of numerous base learning algorithms to come up with a better predictive model than that of any of the single learning algorithm. An ensemble classification model makes a set of base classifiers from training data and then perform classification taking a vote of each of the base classifier's predictions. The very inception of this algorithm is a fascinating procedure called Boosting. Boosting focuses on the training examples that are hard to classify, and it achieves this by iteratively change the distribution of the training instances. AdaBoost and LogitBoost algorithms are the most well-known form of the boosting procedure [63, 64]. These bear the most vital flexibility for adding many weak classifiers having high error rates to produce a combined hypothesis whose training error rate is small.

AdaBoost and Random Forest (AdaBoost-RF) Classifier

An AdaBoost algorithm has been used with the Random Forest classifier as base or weak learner for the mammogram classification. Random Forest is an ensemble classification technique proposed by Breiman that specially designed for decision tree classifier [65]. It utilizes the bagging procedure, where randomness is interposed into the model-building process by randomly choosing the samples, with replacement from the original training dataset. The bagging technique uses the uniform probability distribution to generate the bootstrapped samples that build each classifier throughout the entire mode-building process. This method combines the predictions generated by multiple decision trees where each tree is built based on the values of an independent set of random vectors and with the same distribution for all the trees in the forest. The learning error rate of the Random Forest depends on the number of input features used in each node of the decision tree. Thus, a decision tree is built using the training set and random vector. After many trees are generated, a classification decision hypothesis is fitted to a function based on the voting for the class.

AdaBoost algorithm is a popular version of boosting procedure. It has the high flexibility for adding many weak classifiers having high error rates to generate a combined hypothesis whose training error rate is small [63, 64, 66]. For the binary classification problem, an AdaBoost algorithm uses the training dataset $\mathcal{X} \rightarrow \{x_i, y_i\}$, $i = 1, 2, ..., K$ contains K number of instances. The attribute x_i of $\mathcal{X}$ is an instance associated with corresponding class label $y_i \in \{-1, +1\}$. The class labels -1 and $+1$ represent the positive and negative type in the mammographic class

set. With the use of the training dataset $\mathcal{X}$, the AdaBoost model handles the binary classifications using following factors.

1. Initially equal weights ($\mathcal{W}$) are assigned to all training instances.

2. In each round, a weak hypothesis, $h_n(x_i)$ of lower error rate is generated by n-th base classifier. The importance of each base classifier is dependent on its error rate defined as,
$$error_n \leftarrow \sum_{i:h_n(x_i)\neq y_i} \mathcal{W}_n(i). \tag{1.23}$$

3. The importance of the base classifier depends on the constant, c_n is given by,
$$c_n \leftarrow \frac{1}{2}\ln\left(\frac{1-error_n}{error_n}\right) \tag{1.24}$$

4. The weights of misclassified instances are increased by updating the weights of the training instances and given as,
$$\mathcal{W}_{n+1}(i) \leftarrow \frac{\mathcal{W}_n(i)}{Z_n} e^{-c_n y_i h_n(x_i)} \tag{1.25}$$
where Z_n is a normalizing constant.

5. The instances with higher weights are selected to train the classifier in the next round.

6. The final decision is obtained by the linear combination of the weak hypothesis generated in each round.

The detail description of the AdaBoost procedure with Random Forest classifier is given in the **Algorithm 1**.

LogitBoost and Random Forest (LogitBoost-RF) Classifier

A LogitBoost algorithm based on the Random Forest (RF) that acts as the base classifier has been employed to characterize mammograms into malignant, benign and normal type [63]. A LogitBoost algorithm is to be derived from the AdaBoost model by applying least squares regression cost function. In the LogitBoost algorithm, the same procedure has been followed with only one additional application of the least square cost function. In this model, each of the training examples is initialized by a specific weight that determines the probability of the corresponding sample being selected for the training set ($\mathcal{X}$) in the classifier. Initially, an equal weight is given to

Algorithm 1 Classification using AdaBoost-RF method.

Require: Dataset having K instances, $X = x_i$ for $i = 1, 2, ..., K$ with labels $y_i \in \{-1, +1\}$, N: Total number of iteration and T: Total number of trees.

Ensure: $classifier\_decision$

1: Initialize weight $W_1(i) \leftarrow 1/K, \ \forall i$
2: **for** $n \leftarrow 1$ to N **do**
3: **for** $t \leftarrow 1$ to T **do**
4: Generate a vector V_t with $W_n(i)$
5: $X_t \leftarrow \ bootstrap(X)$
6: $ctree_t \leftarrow \ buildtree(X_t, V_t)$
7: **return** hypothesis h
8: **end for**
9: Obtain class hypothesis, $h_n(x_i) \rightarrow y_i \in \{-1, +1\}$
10: Compute the error of $h_n(x_i)$,

$$error_n \leftarrow \sum_{i:h_n(x_i) \neq y_i} \mathcal{W}_n(i)$$

11: Set a constant $c_n \leftarrow \frac{1}{2} \ln \left(\frac{1-error_n}{error_n} \right)$
12: $\mathcal{W}_{n+1}(i) \leftarrow \frac{\mathcal{W}_n(i)}{Z_n} e^{-c_n y_i h_n(x_i)}$
13: **end for**
14: $h_{final}(x) \leftarrow \sum_{n=1}^{N} c_n h_n(x)$
15: $classifier\_decision \leftarrow sign[h_{final}(x)]$

all the training examples of the dataset with a probability estimation value of 0.5 to the corresponding training instance.

Next, the base learner is repeatedly trained on the weighted version of training examples for many rounds. In each round, a base learner decision function is generated and at the same time two parameters, working response and weight of the i-th instance are also computed. The learner decision fitting function is generated by a weighted least square regression of computed working response to the corresponding instance using the earlier assigned weight. Then the classification decision function gets updated in an additive manner of each classification function obtained from each round. Consequently, the probability estimation of the training example is also updated. Finally, after completion of all rounds, a classifier decision is obtained by taking the signum function of the accomplished updated classification function. The detailed description of the LogitBoost procedure with Random Forest as base classifier

has been given in **Algorithm 2**.

Algorithm 2 Classification using LogitBoost-RF method.

Require: Feature dataset $\mathcal{X}$ of K instances having a instance $x_i \in \mathcal{X}$ for $i = 1, 2, ..., K$ associated with class label $\mathcal{C}_i \rightarrow y_i \in \{-1, +1\}$, RF: Random Forest classifier, M: total number of iteration and T: total number of trees.

Ensure: *decision*: Induced decision of classification.

1: Initialize weights $\mathcal{W}_i \leftarrow 1/K$ for all i
2: set $F(x) \leftarrow 0$ and $p(x_i) \leftarrow 0.5$ $\{F(x)$: classifier decision function, $p(x_i)$: probability estimation of the instance x_i for binary class classification$\}$
3: **for** $m \leftarrow 1$ to M **do**
4: **for** $t \leftarrow 1$ to T **do**
5: Construct a vector V_t with weight $\mathcal{W}_{im}$
6: $\mathcal{X}_t \leftarrow bootstrap(\mathcal{X})$
7: $Ctree \leftarrow buildtree(\mathcal{X}_t, V_t)$
8: **end for**
9: $f_m(x_i) \leftarrow y_i \in \{-1, +1\}$ $\{m$-th base learner decision$\}$
10: Compute $z_i \leftarrow \frac{y_i - p(x_i)}{p(x_i)(1-p(x_i))}$ $\{z_i$: working response$\}$
11: Compute $\mathcal{W}_i \leftarrow p(x_i)(1 - p(x_i))$
12: Fit the learner decision $f_m(x_i)$ by a weighted least-squares regression of z_i to x_i using computed weight $\mathcal{W}_i$
13: Update $F(x) \leftarrow F(x) + \frac{1}{2} f_m(x_i)$
14: Update $p(x) = \frac{\exp\{F(x)\}}{\exp\{F(x)\} + \exp\{-F(x)\}}$
15: **end for**
16: $decision \leftarrow sign[F(x)]$

1.8.4 Logistic Model Tree (LMT)

The Logistic Model Tree (LMT) is a regression tree in which each node fits a LogitBoost function [67, 68]. The LMT consists of a tree structure that includes sets N and T of non-terminal nodes and terminal nodes, respectively. The tree is provided by the entire instance space, $\mathcal{X} \in SFM \rightarrow \{x_i, y_i\},\ i = 1, 2, ..., K$ contains K number of instances. The attribute $x_i \in sfv_i$ of $\mathcal{X}$ is an instance associated with corresponding class label $y_i \in C \rightarrow \{-1, +1\}$, where C is the number of classes. The class labels, -1 and $+1$ represent the negative and positive type in the mammographic ROIs class sets. Then, a disjoint subdivision of significant feature matrix (SFM) into regions $\mathcal{X}_t$ is made by the tree structure and expressed as,

$$\mathcal{X} = \bigcup_{t \in T} \mathcal{X}_t \tag{1.26}$$

where $\mathcal{X}_t \bigcap \mathcal{X}_{t'} = \phi,\ \forall\ t \neq t'$. Each formed region is represented by a leaf node $t \in T$, which fits a LogitBoost function. The LogitBoost function $f(t)$ develops an additive model of least-squares fitted to the given relevant dataset $\mathcal{X}$ for each class y and given as,

$$F_y(x) = \beta_0 + \sum_{i=1}^{R} \beta_i x_i \tag{1.27}$$

where R is the number of features present in the feature vector of each instance x, β_0 is the coefficient of initial feature component, β_i is the coefficient of the ith component in the feature vector. The class membership probabilities induced by the model is given by,

$$P(y\,|\mathcal{X} \to x) = \frac{\exp(F_y(x))}{\sum_{j=1}^{C} \exp(F_j(x))}. \tag{1.28}$$

Then, the model represented by the entire LMT is given by,

$$f(x) = \sum_{t \in T} f_t(x).I(x \in \mathcal{X}_t) \tag{1.29}$$

where $I(x \in \mathcal{X}_t) = \begin{cases} 1, & \text{if } x \in \mathcal{X}_t \\ 0, & \text{otherwise} \end{cases}$.

Now, the LMT is built by the following steps:

1. The tree is grown based on C4.5 approach [69], where each node fits a logistic model using LogitBoost algorithm.

2. A split is formed at the root of the data. Then, the tree is continuously growing by sorting the appropriate subsets of the data to the child nodes. Also, at each child node, a logistic regression model is built.

3. At each split, the logistic regressions of the parent node are passed to the child node.

4. The splitting process continues till it achieves a minimum information gain. The tree growing stops when there is no more split.

5. The final model in the terminal nodes accumulates all parent models that generates the probability estimates for each class.

6. After the entire tree has been built, CART-based pruning [70] is applied for reduction of the size of the tree that enhances the generalization of the model.

1.9 Thesis Organization

This thesis is organized into six different chapters including introduction. Each chapter presents the contributions specific to the feature development and selection scheme. The efficacy of the proposed schemes have been validated using several classifiers namely, Naive Bayes (NB), K-Nearest Neighbor (K-NN), Back-Propagation Neural Network (BPNN), Support Vector Machine (SVM), AdaBoost and Random Forest (AdaBoost-RF), LogitBoost and Random Forest (LogitBoost-RF), and Logistic Model Tree (LMT) on two standard databases, MIAS and DDSM. Each chapter is discussed below in a nutshell.

Chapter 2: Mammogram Classification using DWT and GLCM Features Followed by *t-test* Feature Selection

Two Dimensional Discrete Wavelet Transform (2D-DWT) and Gray-Level Co-occurrence Matrix (GLCM) is used in succession to extract the feature descriptors from the mammographic ROIs. To derive the relevant features from the feature matrix, *t-test* and *F-test* are utilized independently. The relevant features are used in different classifiers for classification of mammograms. The performance of classification is evaluated with respect to accuracy (ACC) and AUC of ROC curve. The accuracy measures are computed with respect to normal vs. abnormal and benign vs. malignant. It has been observed that, Back-Propagation Neural Network (BPNN) achieves better result among all classifiers. For MIAS database the optimal accuracy measures are 98.13% and 94.20% respectively, whereas for DDSM database they are 98.80% and 97.40%. Similarly, AUC parameters are of 0.9899, 0.9504 for MIAS, and 0.9945, 0.9761 for DDSM database.

Chapter 3: Mammogram Classification using SFTA Features with FCBF Feature Selection

This chapter presents an effective scheme to identify the abnormal mammograms in order to detect the breast cancer. The scheme utilizes the Segmentation-based Fractal Texture Analysis (SFTA) method to extract the texture features from the

mammograms. A Fast Correlation-Based Filter (FCBF) method is used to select feature subsets containing significant features, which are used for classification purpose. The 10-fold cross-validation has been made to obtain the optimal relevant feature subset. The scheme has been validated using different classifiers. A promising classification performances of $ACC = 98.76\%$, $AUC = 0.9901$ (abnormal–normal), and $ACC = 95.65\%$, $AUC = 0.9705$ (malignant–benign) have been achieved by SVM for MIAS database. The similar results of $ACC = 99.20\%$, $AUC = 0.9988$ (abnormal–normal), and $ACC = 98.00\%$, $AUC = 0.9967$ (malignant–benign) have been archived for DDSM database.

Chapter 4: Mammogram Classification using DOST Features followed by Null-hypothesis based Feature Selection

A Two-Dimensional Discrete Orthonormal S-Transform (2D-DOST) is used to extract the coefficients from the digital mammograms. A feature selection algorithm based on null-hypothesis with statistical *two-sample t-test* has been suggested to select most significant coefficients from large number of DOST coefficients. The selected coefficients are used in different classifiers for classification of mammograms. It has been observed that, the optimal results with respect to ACC and AUC are achieved by AdaBoost-RF classifier. The parameters are $ACC = 98.75\%$, $AUC = 0.9991$ (abnormal–normal), and $ACC = 98.26\%$ and $AUC = 0.9985$ (malignant–benign) for MIAS database. Similarly, for DDSM database the parameters are $ACC = 99.30\%$, $AUC = 0.9994$ (abnormal–normal), and $ACC = 98.80\%$, $AUC = 0.9992$ (malignant–benign).

Chapter 5: Mammogram Classification using Slantlet Features followed by BLogR for Feature Selection

This chapter presents an efficient scheme to characterize the type of digital mammogram as malignant, benign or normal in the early diagnosis of the breast cancer. The texture features are extracted by performing the Two-Dimensional Slantlet Transform (2D-SLT) on enhanced mammographic ROIs. The most significant features are selected from the available derived features by utilizing the Bayesian Logistic Regression (BLogR) method. To accomplish an adequate improved performance, the relevant features are balanced by the Gaussian distribution based balancing method prior to classification. The optimal performance measures are obtained by LogitBoost-RF classifier among all classifiers on similar platform. The proposed approach achieves the optimal accuracy results of 99.69% and 99.13% for abnormal–normal and the malignant–benign class set on MIAS database respectively.

The similar parameters of 99.80% and 99.40% are accomplished for DDSM database. The optimal AUC of value 1 with respect to ROC curve is achieved for all the class sets of both the databases.

Chapter 6: Mammogram Classification using Radial Symmetric Features followed by t-SNE Feature Selection

The Fast Radial Symmetry Transform (FRST) is performed on the mammographic Region-of-Interest (ROI) to derive the radially symmetric features. A t-distributed Stochastic Neighbor Embedding (t-SNE) method has been utilized to select most relevant features. The suggested scheme has been validated using various classifiers. Experimental results show an optimal classification performance that is achieved by LMT classifier with respect to accuracy (ACC) and area under ROC curve (AUC) value. The ACC measures are estimated concerning malignant vs. normal, malignant vs. benign, and benign vs. normal classes. For MIAS database, the accuracy measures are 99.61%, 99.13%, and 99.63% respectively, whereas, for DDSM database, they are 99.87%, 99.40%, and 99.73%. Similarly, the AUC values are 0.9997, 1, and 0.9998 for MIAS datbase. For DDSM database, the parameters are 1, 1, and 0.9968

Chapter 7: Conclusions and Future Work

This chapter presents the conclusions drawn from the proposed work with emphasis on the achievements and limitations. The scope for future research work has been discussed at the end.

Chapter 2

Mammogram Classification using DWT and GLCM Features Followed by *t-test* Feature Selection

---

The leading cause of death among cancer for women is the breast cancer. Early detection of breast cancer has been observed to improve the recuperation rates to a great extent. In this context, mammography is one of the most diagnostic tests for pre-screening the breast cancer. In most cases, experienced radiologists are responsible for interpreting and analyzing the mammograms. However, due to the possibility of human error, the result brings the variable judgments. In this regard, it is an extremely challenging and difficult task for radiologists to classify the suspicious lesion correctly in mammograms. Therefore, automatic classification of digital mammograms is required to support the decision making of radiologists. For accurate classification, the relevant features with an improved classifier play a vital role. In this chapter, we suggest a hybrid feature followed by a feature selection mechanism to classify mammograms as malignant, benign or normal . The feature extraction algorithm concentrates on the texture point in the mammographic image utilizing Two-Dimensional Discrete Wavelet Transform (2D-DWT) and Gray-Level Co-occurrence Matrix (GLCM) in succession on Region-of-Interest (ROI) to find out the feature descriptors of each detail coefficient at two-level resolution. Two statistical methods namely, *two-sample t-test* and *F-test* have been applied independently to

select significant features. Utilizing relevant features, several classifiers have been used to validate the proposed scheme. It has been observed that Back-Propagation Neural Network (BPNN) has showed the significant performance among all classifiers. The classification using BPNN has been explained elaborately in this chapter. The overall block diagram of the scheme is shown in Figure 2.1.

The rest of the chapter is organized as follows: The extraction of ROI form the mammographic image is described in Section 2.1. The fundamentals of 2D-DWT and GLCM methods are explained in Sections 2.2 and 2.3 respectively for completeness. The extraction of feature using 2D-DWT and GLCM method is described in Section 2.4. Section 2.5 outlines the feature selection followed by the classification. Section 2.6 describes the experimental results obtained on standard databases MIAS and DDSM. Section 2.7 gives the summary of overall work proposed in this chapter.

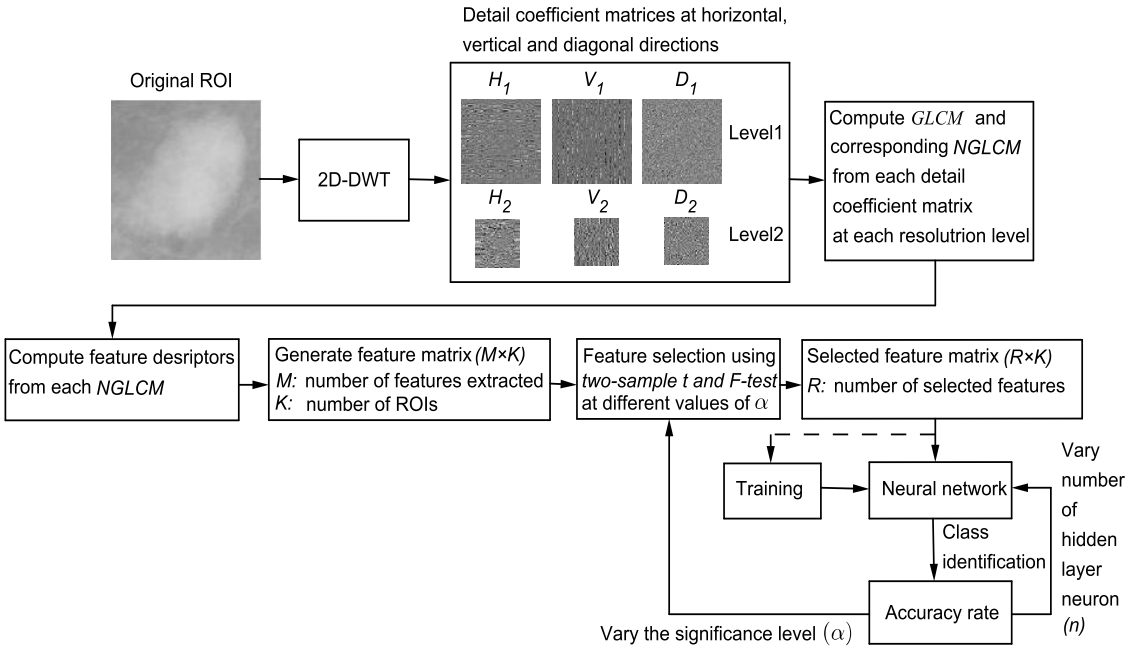
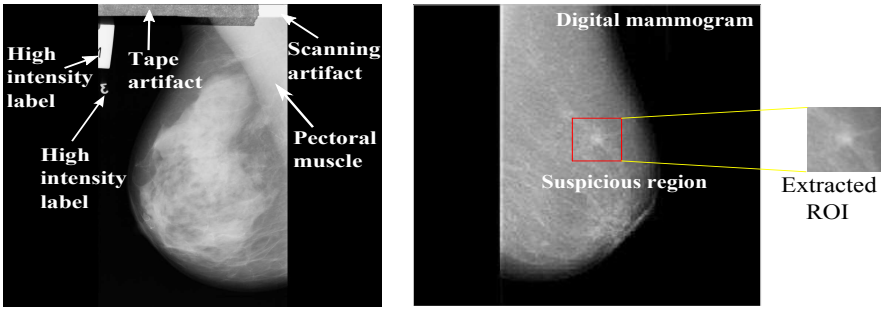


Figure 2.1: Block diagram of proposed scheme using 2D-DWT and GLCM.

2.1 Extraction of Region-of-Interest (ROI)

It might be noticed that digital mammographic image is composed of different types of noise, imaging artifacts, object background, and pectoral muscle as shown in Figure 2.2(a). All these regions are undesirable segments for the analysis of texture because of which; the full mammographic image is not suitable for feature extraction

and consequent characterization. Hence, a cropping operation has been applied on mammographic image to extract the Region-of-Interest (ROI) which contains the suspicious abnormality, excluding the undesirable parts of the image. This process is performed by referring the center of the abnormal area as the center of ROI and taking the approximate radius (in pixels) of a circle enclosing the abnormal area. Subsequently, the circular image is converted to a 128×128 rectangular image that encompassed the circle, and is shown in Figure 2.2(b). For the extraction of normal ROI, the same cropping procedure is performed on normal mammographic images with a random selection of the location. Thus, in this phase, the ROIs extracted are free from the background information and noises. The cropping process has been applied on the mammograms of MIAS database to extract ROIs based on the prior ground truth informations [11]. However, for the DDSM mammographic images, the ROIs are already given as mammographic patches by the IRMA project [12].



(a) Undesirable regions (b) ROI extraction

Figure 2.2: Mammogram with various undesirable regions and ROI extraction.

2.2 Multiresolution Analysis using 2D-DWT

In the multiresolution technique, the underlaying texture of mammographic ROIs are analyzed by zooming in and out process. Two-Dimensional Discrete Wavelet Transform (2D-DWT) decomposes the mammographic ROI into a number of sub-images in different resolution levels preserving the high and low frequency information. This property leads the wavelet to extract better texture information from the mammographic ROIs. Given a continuous, square integrable function $f(x)$, its wavelet transform is calculated as the inner product of $f(.)$ and a real valued wavelet function $(\psi(x))$ [71] given by,

$$W[f(s,\tau)] = \langle f, \psi^k_{s,\tau} \rangle = \int_{-\infty}^{\infty} f(x)\, \psi^k_{s,\tau}(x)\, dx \tag{2.1}$$

where $\psi^k_{s,\tau}(x) = \frac{1}{\sqrt{s}}\psi^k\left(\frac{x-\tau}{s}\right)$ is a wavelet family, $s \in Z$, τ and $k \in \{h, v, d\}$ are scale (resolution level), translation and orientation parameters respectively. The orientation parameters h, v and d represents to horizontal, vertical and diagonal directions respectively. Now the dyadic wavelet decomposition is achieved when $s = 2^j$ and $\tau = 2^j.n, \quad j, n \in Z$. Using the wavelet function $\psi(x)$ and scaling function $\varphi(x)$, the wavelet and scaling families are constructed as,

$$\psi^k_{j,n}(x) = \frac{1}{\sqrt{2^j}}\psi^k\left(\frac{x-2^j.n}{2^j}\right) \; and \; \varphi^k_{j,n}(x) = \frac{1}{\sqrt{2^j}}\varphi\left(\frac{x-2^j.n}{2^j}\right) . \tag{2.2}$$

These are orthonormal basis of sub-spaces and related to resolution 2^j. The wavelet atoms are defined by scaling and translating three mother atoms like ψ^h, ψ^v and ψ^d. These oriented mother atoms are computed as the tensor product of one dimensional $\psi(x)$ and $\varphi(x)$ given by,

$$\begin{array}{l} \varphi(x) = \varphi(x_1)\,\varphi(x_2), \psi^h(x) = \psi(x_1)\,\varphi(x_2), \\ \psi^v(x) = \varphi(x_1)\,\psi(x_2) \;\; and \;\; \psi^d(x) = \psi(x_1)\,\psi(x_2) \end{array} . \tag{2.3}$$

A Two-Dimensional Discrete Wavelet Transform is implemented using the combination of digital filter banks and down-samplers. The digital filter banks consist of high-pass (g) and low-pass (h) filters. In the configuration of DWT structure, the number of banks is set as per the desired resolution [72]. As the image is a $2D$ signal, separable wavelet functions compute the Discrete Wavelet Transform (DWT). The rows and columns of the image are separately undergone through the 1D wavelet transform to establish the 2D-DWT. As shown in Figure 2.3, the original image $A_{2^{j+1}}f$ at resolution 2^{j+1} is decomposed into four sub-band images in the frequency domain. Among them, three sub-band images, $D^h_{2^j}f$, $D^v_{2^j}f$, $D^d_{2^j}f$ are the detail images at resolution 2^j in horizontal, vertical, and diagonal directions respectively. The Fourth one is the approximation image, $A_{2^j}f$ found at coarse resolution. So the whole image $A_{2^{j+1}}f$ is represented as,

$$A_{2^{j+1}}f = D^h_{2^j}f + D^v_{2^j}f + D^d_{2^j}f + A_{2^j}f. \tag{2.4}$$

The decomposed sub-images are the representation of 2D orthogonal wavelet. Thus, the output of a wavelet decomposition of an image results into four orthogonal sub-band components like Low-Low (LL), Low-High (LH), High-Low (HL) and High-High (HH), that correspond to sub-images $D^h_{2^j}f$, $D^v_{2^j}f$, $D^d_{2^j}f$ and $A_{2^j}f$ respectively as shown in Figure 2.3.

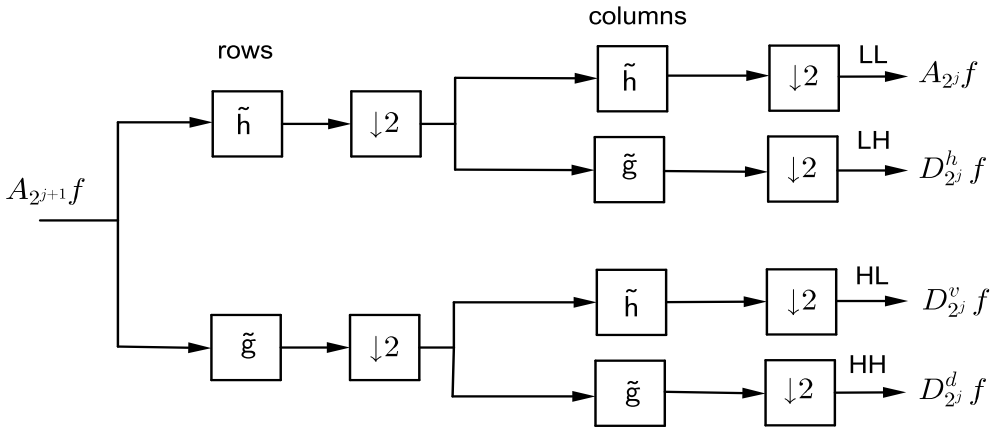


Figure 2.3: Wavelet decomposition using analysis filter banks.

2.3 Gray-Level Co-occurrence Matrix (GLCM)

The Gray-Level Co-occurrence Matrix ($GLCM$) is used to extract the texture in an image by doing the transition of gray level between two pixels. The $GLCM$ gives a joint distribution of gray level pairs of neighboring pixels within an image [73]. The co-occurrence matrix of the ROI is useful in classifying breast tissues by extracting descriptors from the matrix. For the computation of $GLCM$, first a spatial relationship is established between two pixels, one is the reference pixel, and the other is a neighbor pixel. This process forms the $GLCM$ containing different combination of pixel gray values in an image. Let $q\,(i,j)$ is the element of $GLCM$ of a given image f of size $M \times N$ containing the number of gray levels G ranging from 0 to $G-1$. The element $q\,(i,j)$ is defined as,

$$q\,(i,j) = \sum_{x=1}^{M} \sum_{y=1}^{N} \begin{cases} 1\,, \text{ if } f\,(x,y) = i \textit{ and } f\,(x+\Delta x,\, y+\Delta y) = j \\ 0, \text{ otherwise} \end{cases} \tag{2.5}$$

where (x,y) and $(x+\Delta x, y+\Delta y)$ are the locations of reference pixel and its neighboring pixel respectively. Each element of $GLCM$, $q\,(i,j\,|\Delta x, \Delta y)$ represents the relative frequency with which two pixels in a given neighborhood are separated by a distance $(\Delta x, \Delta y)$ having gray level values i and j respectively [74]. It can be represented as $q\,(i,j\,|D,\theta)$, where the parameter D is the distance of separation between two neighboring resolution cells with two pixels having intensities i and j in the image. The other parameter θ represents the direction of neighboring pixel with respect to the pixel of reference. The directionality used in $GLCM$ is shown in Figure 2.4. The

parameter D is also called as set distance as it specifies the distance of all neighboring resolution pairs in a set. For the texture calculation; the $GLCM$ must be symmetrical, and each entry of the $GLCM$ should be a probability value. For this purpose, a normalization process is followed. Each element of the n-dimensional Normalized Gray-Level Co-occurrence Matrix ($NGLCM$) is defined as,

$$p\,(i,j) = q\,(i,j)\,/\sum_{i=0}^{G-1}\sum_{j=0}^{G-1} q\,(i,j). \tag{2.6}$$

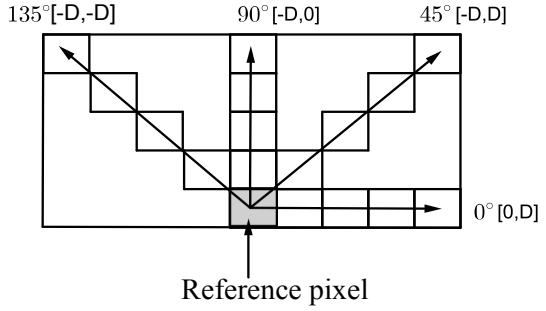


Figure 2.4: Directionality used in Gray-Level Co-occurrence Matrix.

The size of $GLCM$ is same as the number of gray levels of input image. The $GLCM$ is highly dependent on the parameters D and θ. Several matrices can be obtained with small changes in the parameter D and θ. For the digital mammograms, the distance parameter D is limited to integral multiples of the pixel size, and the value of a direction parameter θ can be 0°, 45°, 90° and 135°. Figure 2.5 describes the process of computation of $GLCM$ of a given test image intensity matrix. Here, the number of gray level is considered to be four and the offset values are taken as $[0,1]$, $[-1,0]$, $[-1,1]$, and $[-1,-1]$. The offset values represent set distance $D = 1$ in four possible neighbor pixel directions, $\theta = 0°$, 90°, 45° and 135° with respect to the reference pixel. It can be seen that the occurrence of resolution cells pair $(0,2)$ in the intensity matrix of input image is 4 in the horizontal direction ($\theta = 0°$) due to the symmetric property. Therefore, the element in the $(0,2)$ position of the horizontal $GLCM$ is 4 as shown in Figure 2.5(b). In the same manner other three $GLCM$s are computed. Figure 2.6 shows the Normalized Gray-Level Co-occurrence Matrices ($NGLCM$) where each cell in the matrices contains probability value. Each element of $NGLCM$ is computed by dividing 24 in case of horizontal and vertical directions, and 18 in case of left diagonal and right diagonal directions to each element of corresponding symmetrical $GLCM$.

	0	1	2	3
0	0	2	1	2
1	2	3	2	0
2	0	1	3	2
3	2	0	2	1

(a)

	0	1	2	3
0	0	1	4	0
1	1	0	3	1
2	4	3	0	3
3	0	1	3	0

(b)

	0	1	2	3
0	0	1	5	0
1	1	0	2	1
2	5	2	0	3
3	0	1	3	0

(c)

	0	1	2	3
0	0	0	0	3
1	0	0	2	1
2	0	2	6	0
3	3	1	0	0

(d)

	0	1	2	3
0	2	1	0	1
1	1	0	2	1
2	0	2	4	0
3	1	1	0	2

(e)

Figure 2.5: Computation of co-occurrence matrices. (a) Intensity values of input image with 4 gray levels. Different co-occurrence matrices ($GLCM$) for set distance $D = 1$ at four different directions such as (b) horizontal ($\theta = 0°$), (c) vertical ($\theta = 90°$), (d) right diagonal ($\theta = 45°$), (e) left diagonal ($\theta = 135°$).

	0	1	2	3
0	0	0.0417	0.1667	0
1	0.0417	0	0.1250	0.0417
2	0.1667	0.1250	0	0.1250
3	0	0.0417	0.1250	0

(a) $\theta = 0°$

	0	1	2	3
0	0	0.0417	0.2083	0
1	0.0417	0	0.0833	0.0417
2	0.2083	0.0833	0	0.1250
3	0	0.0417	0.1250	0

(b) $\theta = 90°$

	0	1	2	3
0	0	0	0	0.1667
1	0	0	0.1111	0.0556
2	0	0.1111	0.3333	0
3	0.1667	0.0556	0	0

(c) $\theta = 45°$

	0	1	2	3
0	0.1111	0.0556	0	0.0556
1	0.0556	0	0.1111	0.0556
2	0	0.1111	0.2222	0
3	0.0556	0.0556	0	0.1111

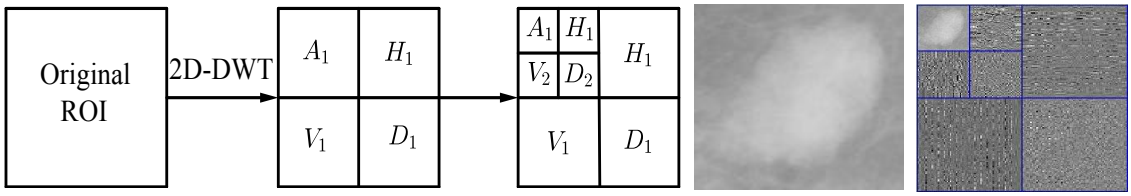
(d) $\theta = 135°$

Figure 2.6: $NGLCM$ of corresponding $GLCM$ in Figure 2.5 at different directions.

2.4 Feature Extraction using 2D-DWT and GLCM

In the discrete wavelet decomposition, the output detail images give the detail coefficients of the original image. It is found that, the approximation sub-image carries little energy due to which it is not taken into consideration for texture analysis of

mammographic ROI. But the wavelet detail coefficients provide the texture descriptors of the mammographic ROI. Using 2D-DWT, the three detail coefficient matrices at each resolution level are obtained, which represent horizontal, vertical, and diagonal sub-structures of the ROI as shown in Figure 2.7. Then the Gray-Level Co-occurrence Matrices are calculated at each resolution level by taking the absolute value of each coefficient in the corresponding matrices. For analysis of texture patterns of each ROI, the following five texture descriptors such as *energy*, *correlation*, *entropy*, *sum variance*, and *sum average* are computed using GLCM [73]. The expressions for different texture feature descriptors (FD) are given in Table 2.1. Now $p(i,j)$ is the (i,j)th entry of normalized $GLCM$. Let $p_x(i)$ is the i^{th} entry in the marginal probability matrix by summing the rows of $p(i,j)$, defined as, $p_x(i) = \sum_{j=1}^{G} p(i,j)$, where G is the number of distinct gray levels in the quantized ROI. Similarly, $p_y(j) = \sum_{i=1}^{G} p(i,j)$ and $p_{x+y}(k) = \sum_{\substack{i=1 \\ i+j=k}}^{G} \sum_{j=1}^{G} p(i,j)$, $k = 2,3,...,2G$. The steps associated for computation of feature matrix from 2D-DWT and GLCM are described in **Algorithm 3**.



(a) Wavelet decomposition at two resolution level (b) Original ROI (c) DWT of ROI

Figure 2.7: 2D-DWT of mammographic ROI.

In **Algorithm 3**, a 2D-DWT is applied on K mammographic ROIs to produce different detail coefficient matrices (DM) at r different directions such as horizontal, vertical and diagonal directions for l resolution levels. A co-occurrence matrix ($GLCM$) and its corresponding normalized co-occurrence matrix ($NGLCM$) are calculated from each DM in four directions ($p = 4$) i.e., at $\theta = 0°$, $45°$, $90°$, and $135°$ at a set distance D. Then all the feature descriptors (FD) mentioned in Table 2.1 are computed from each $NGLCM$ and combined to form a feature descriptor matrix (FDM). Thus, a feature matrix is generated by concatenating all the FDMs from all $NGLCM$s for K number of ROIs.

Table 2.1: Computation of feature descriptors for mammographic ROIs.

Feature Descriptor	Name	Computation
FD_1	Energy	$\sum_{i=1}^{G}\sum_{j=1}^{G}\{p(i,j)\}^2$
FD_2	Correlation	$\frac{\sum_{i=1}^{G}\sum_{j=1}^{G}(i,j)p(i,j)-\mu_x\mu_y}{\sigma_x\sigma_y}$
FD_3	Entropy	$-\sum_{i=1}^{G}\sum_{j=1}^{G}p(i,j)\log(p(i,j))$
FD_4	Sum variance	$\sum_{i=2}^{2G}(i-sum\,entropy)^2 p_{x+y}(i)$
FD_5	Sum average	$\sum_{i=2}^{2G} i p_{x+y}(i)$

where, μ_x, μ_y, σ_x and σ_y are the means and standard deviations of p_x and p_y, and $sum\,entropy = -\sum_{i=2}^{2G} p_{x+y}(i)\log\{p_{x+y}(i)\}$.

2.5 Feature Selection and Classification

The features extracted from the textures of ROIs are expressed as mathematical descriptions. This helps the classifier to distinguish the breast tissues as malignant, benign or normal. However, one major problem lies with the large number of features that is very difficult to determine which feature or combination of features achieves better classification accuracy rate [8]. Therefore, it is important to select a suitable and optimized set of features from a high dimensional feature matrix that has the ability to distinguish between different types of mammograms. In this scheme, two statistical methods such as *two-sample t-test* and *F-test* have been used independently to select the most significant features from the feature matrix. *Two-sample t-test* and *F-tests* are performed on two classes, and a test decision is returned for the null hypothesis that the data in two vectors v_1 and v_2 come from normal distributions with equal means. The objective of the test is to determine whether the data from two vectors v_1 and v_2 are related or not. In the proposed feature selection algorithm, a null hypothesis value, $h = 1$ indicates that the null hypothesis is incorrect and rejected. An incorrect null hypothesis implies that, data from two vectors v_1 and v_2 are different and independent. In the *two-sample t-test* and *F-test* method, the t and F values are computed as,

$$t = \frac{|\mu_{v_1} - \mu_{v_2}|}{\sqrt{\frac{(\sigma_{v_1})^2}{K_{v_1}} + \frac{(\sigma_{v_2})^2}{K_{v_2}}}} \quad \text{and} \quad F = \frac{S_{v_1}^2}{S_{v_2}^2} \tag{2.7}$$

Algorithm 3 Feature matrix generation using 2D-DWT and GLCM.

Require: K: total number of ROIs, l: resolution level, r: number of directions in which DM is to be computed, θ: direction , p: number of directions, D: set distance, s: number of feature descriptors, fv and FDM: feature descriptor vector and matrix respectively.

Ensure: $FM[M][K]$: feature matrix. Function $wavedec()$ performs wavelet transform of ROI. Function $detcoef()$ extracts three detail and approximation coefficient components from transformed ROI at lower resolution levels. And function $graycomatrix()$ computes $GLCM$ from DM.

$fv \leftarrow \phi$, $FDM \leftarrow \phi$
Initialize $l \leftarrow 2,\ r \leftarrow 3,\ p \leftarrow 4$, and $s \leftarrow 5$
$M \leftarrow l \times r \times p \times s$
for $i \leftarrow 1$ to K **do**
  Read ROI_i, $D \leftarrow 1$
  for $j \leftarrow 1$ to l **do**
    $TROI_i \leftarrow wavedec(ROI_i)$ $\{TROI_i$ is the wavelet transform of $ROI_i\}$
    for $d \leftarrow 1$ to r **do**
      $DM_{jd} \leftarrow detcoef(TROI_i)$
      for $k \leftarrow 1$ to p **do**
        $GLCM_{jd\theta_k} \leftarrow graycomatrix(DM_{jd}, \theta_k, D)$
        $NGLCM_{jd\theta_k} \leftarrow GLCM_{jd\theta_k}$ / sum(elements of $GLCM_{jd\theta_k}$)
        for $q \leftarrow 1$ to s **do**
          Compute FD_q from $NGLCM_{jd\theta_k}$ and $fv \leftarrow fv \bigcup FD_q$
          $FDM_{jd\theta_k} \leftarrow FDM \bigcup fv$
        end for
      end for
    end for
    $D \leftarrow D + 1$
  end for
end for
$FM \leftarrow$ concatenate (FDMs)

where K_{v_1} and K_{v_2} are the numbers of ROIs in two classes. Here, μ_{v_1} and μ_{v_2} are means, σ_{v_1} and σ_{v_2} are standard deviations, and S_{v_1} and S_{v_2} are the variances of two classes. The higher t and F values indicate more significant differences between the means of the two vectors. For a certain threshold t and F values, corresponding p_1 and p_2 values define probabilities of obtaining a t and F values more than the

threshold. A significance level, α defines the lower threshold for the p_1 and p_2 values. The value of α is in the range 0 and 1. As the α value decreases from one to zero, the selection of the number of features reduces. The selection of significant features has been described in **Algorithm 4**.

Algorithm 4 Feature selection using *two-sample t-test* and *F-test* method.

Require: $FM[M][K]$, $target[1][N]$, α: Significance level

Ensure: $SFM1[R][K]$ and $SFM2[R][K]$. R: Total number of selected features. Functions *ttest()* and *vartest()* compute the null hypothesis values of two vectors at different values of significance level, by *two-sample t* and *F-test* respectively.

1: Create two empty vectors v_1 and v_2
2: Initialize α, $0 < \alpha < 1$
3: **for** $i \leftarrow 1$ to M **do**
4: Clear contents of vector v_1 and vector v_2
5: **for** $j \leftarrow 1$ to K **do**
6: **if** $target\_class[j] = 1$ **then**
7: Append $FM[i][j]$ to v_1
8: **else**
9: Append $FM[i][j]$ to v_2
10: **end if**
11: **end for**
12: $h_1[i] \leftarrow ttest(v_1, v_2, \alpha)$
13: $h_2[i] \leftarrow vartest(v_1, v_2, \alpha)$
14: **for** $l \leftarrow 1$ to 2 **do**
15: **if** $h_k[i] = 1$ **then**
16: Append $FM[i][K]$ to SFM_k
17: **end if**
18: **end for**
19: **end for**

To validate the efficacy of the proposed hybrid DWT and GLCM features, various classifiers namely, Back-Propagation Neural Network (BPNN), Support Vector Machine (SVM), AdaBoost and Random Forest (AdaBoost-RF), LogitBoost and Random Forest (LogitBoost-RF), and Logistic Model Tree (LMT) have been used for classification. Moreover, two other classifiers such as Naive Bayes (NB) and K-Nearest Neighbor (K-NN) are also used to compare the classification accuracies utilizing the same proposed features. Among all the classifiers, BPNN has shown superior performance as compared to others. The artificial neural network structure

(Figure 1.10) has been experimentally optimized as a three layer network with 15 hidden neurons. Different performance measures of classification such as sensitivity (S_n), specificity (S_p), accuracy (ACC), and AUC of ROC analysis are studied with the different number of significant features by varying the α values and changing the number of neurons (H) in the hidden layer of neural network as shown in Figure 2.1.

2.6 Experimental Results and Analysis

In this scheme, simulation experiments have been carried out in MATLAB environment. For the evaluation of the performance, two image class sets are built from MIAS and DDSM databases and used in the experiments namely, abnormal–normal and malignant–benign. The abnormal and malignant type of ROIs are considered as positive class in the abnormal–normal and malignant–benign image class sets respectively.

2.6.1 Results for Feature Extraction

In this work, the symmetric biorthogonal 4.4 wavelet has been used to compute DWT of images. It has been observed that at $l = 2$, the 2D-DWT gives the suitable results on feature extraction. At each resolution level (j) the DWT results three detail coefficient matrices and thus a total of six detail coefficient matrices (DM) such as H_1, V_1, D_1 at $j = 1$, and H_2, V_2, D_2 at $j = 2$ are obtained in three different directions. Furthermore, four $GLCM$ and corresponding $NGLCM$ are computed from each detail coefficient matrix (DM) at each resolution level. The resolution level (j) of wavelet transform acts as the distance parameter (D) for $GLCM$ computation. The value of D has been taken 1 and 2 for resolution level $j = 1$ and $j = 2$ respectively. From each $NGLCM$, a total of five feature descriptors ($s = 5$) such as *energy*, *correlation*, *entropy*, *sum variance*, and *sum average* are extracted and consequently, form a feature descriptor matrix. Thus, for $l = 2$, $r = 3$, $p = 4$ and $s = 5$, a total 120 ($M = l \times r \times p \times s$) features are extracted from K number of ROIs. This M number of features are kept in rows with corresponding K number of ROIs in columns to generate a feature matrix, which is used in feature selection algorithm. Tables 2.2, 2.3, 2.4, and 2.5 show the values of different texture feature descriptors for different types of ROIs at each resolution level (j).

Table 2.2: Different values of various feature descriptors at $\theta = 0°$ with set distance $D = 1$ for $j = 1$ and $D = 2$ for $j = 2$.

Mammogram Database	Type of ROI	Detail coefficient	Feature Descriptors				
			FD_1	FD_2	FD_3	FD_4	FD_5
MIAS	Normal	H_1	0.2779	487.5895	2.0317	68.9951	9.7276
		H_2	0.0970	692.6964	2.8604	41.6123	8.1096
		V_1	0.1877	466.5697	1.9693	61.5468	9.0858
		V_2	0.0996	767.3355	2.8207	51.9270	8.8618
		D_1	0.2179	444.1351	1.8202	65.0145	9.2864
		D_2	0.2331	347.4291	2.2023	43.2433	7.9591
	Benign	H_1	0.2281	419.9608	1.9828	52.9509	8.6760
		H_2	0.0968	874.2193	2.7809	57.3901	9.2558
		V_1	0.3602	239.8866	1.6763	48.3035	8.0619
		V_2	0.0884	741.4346	2.8748	47.4457	8.5482
		D_1	0.2263	432.9716	1.8142	65.5525	9.3108
		D_2	0.1219	668.9592	2.5554	56.7644	9.0768
	Malignant	H_1	0.2096	669.7592	1.9406	88.8764	10.8185
		H_2	0.1253	663.9185	2.4817	56.4908	9.0556
		V_1	0.4441	216.5229	1.4351	51.0913	8.0865
		V_2	0.2024	498.4734	2.0593	65.2262	9.4488
		D_1	0.3583	258.8663	1.5051	53.2372	8.3248
		D_2	0.1205	645.2020	2.6058	53.9703	8.8808
DDSM	Normal	H_1	0.3516	280.7523	1.8023	46.6090	8.0955
		H_2	0.1532	844.4912	2.2438	91.2830	11.0730
		V_1	0.1731	459.2810	2.1313	55.5010	8.7921
		V_2	0.1101	653.7500	2.6209	51.4534	8.5875
		D_1	0.1450	549.6110	2.2607	57.1411	8.9572
		D_2	0.2361	325.8900	2.0760	47.0090	8.0912
	Benign	H_1	0.9296	219.4603	0.2129	98.9842	9.9898
		H_2	0.2355	294.7308	1.6366	34.1920	6.9669
		V_1	0.9139	213.1101	0.2764	95.1445	9.9783
		V_2	0.5363	184.5213	1.1469	47.6322	7.7436
		D_1	0.7223	148.4800	0.5889	54.3605	7.8384
		D_2	0.9639	130.3821	0.1285	61.9740	7.9811
	Malignant	H_1	0.9341	217.4412	0.1970	97.1415	9.9778
		H_2	0.3882	246.0833	1.3712	42.7977	7.5013
		V_1	0.9127	214.5800	0.2823	95.3071	9.9841
		V_2	0.2998	239.7766	1.5814	40.7870	7.4602
		D_1	0.3350	284.9154	1.2925	58.8471	8.5895
		D_2	0.7726	143.5648	0.5099	55.7352	7.8725

H_1, V_1, D_1: horizontal, vertical and diagonal detail coefficient matrices at $j = 1$
H_2, V_2, D_2: horizontal, vertical and diagonal detail coefficient matrices at $j = 2$
FD_1, FD_2, FD_3, FD_4 and FD_5 are feature descriptors defined in Table 2.1.

Table 2.3: Different values of various feature descriptors at $\theta = 90°$ with set distance $D = 1$ for $j = 1$ and $D = 2$ for $j = 2$.

Mammogram Database	Type of ROI	Detail coefficient	Feature Descriptors				
			FD_1	FD_2	FD_3	FD_4	FD_5
MIAS	Normal	H_1	0.2326	479.8930	2.1199	71.5657	9.7237
		H_2	0.0903	677.5727	2.8725	42.0656	8.1228
		V_1	0.1969	468.5895	1.9439	60.1313	9.0849
		V_2	0.1015	762.7643	2.7947	51.3033	8.8450
		D_1	0.2177	442.5444	1.8138	65.2780	9.2845
		D_2	0.2250	346.4965	2.2173	43.3453	7.9539
	Benign	H_1	0.2011	413.7083	2.0464	55.6596	8.6787
		H_2	0.0897	864.0826	2.8057	58.5130	9.2778
		V_1	0.4352	241.9732	1.5906	48.3573	8.0619
		V_2	0.1005	746.1557	2.8445	47.4000	8.5461
		D_1	0.2291	433.3259	1.8093	65.8895	9.3106
		D_2	0.1213	674.2864	2.5647	57.0645	9.0855
	Malignant	H_1	0.1930	664.6917	1.9807	92.7172	10.8194
		H_2	0.1234	637.6616	2.4486	56.7353	9.0446
		V_1	0.5052	216.8881	1.3503	50.3845	8.0858
		V_2	0.2093	506.7676	2.0243	64.7834	9.4423
		D_1	0.3559	258.9083	1.5028	53.5048	8.3264
		D_2	0.1279	634.4545	2.5620	53.8055	8.8830
DDSM	Normal	H_1	0.2907	276.4500	1.8987	47.2782	8.0948
		H_2	0.1745	819.9701	2.1207	98.9453	11.0805
		V_1	0.1951	461.4608	2.0561	54.6023	8.7895
		V_2	0.1123	675.1321	2.6809	48.5760	8.5925
		D_1	0.1451	548.7622	2.2601	57.3110	8.9543
		D_2	0.2515	324.4001	2.0657	47.2960	8.0953
	Benign	H_1	0.9139	213.2400	0.2881	95.4910	9.9834
		H_2	0.2197	270.7105	1.7934	34.3020	6.9534
		V_1	0.9259	219.8802	0.2182	96.9100	9.9857
		V_2	0.5974	194.7628	0.9838	50.0120	7.7476
		D_1	0.7195	148.5254	0.5898	54.3786	7.8380
		D_2	0.9639	130.3410	0.1285	61.9856	7.9811
	Malignant	H_1	0.9127	212.2101	0.2813	95.1182	9.9746
		H_2	0.3230	231.8769	1.5543	41.3890	7.4696
		V_1	0.9182	222.9300	0.2372	96.7241	9.9957
		V_2	0.3256	254.5963	1.4196	42.2070	7.4696
		D_1	0.3324	285.0214	1.2995	58.7192	8.5893
		D_2	0.7804	142.9786	0.4945	55.9823	7.8772

Table 2.4: Different values of various feature descriptors at $\theta = 45°$ with set distance $D = 1$ for $j = 1$ and $D = 2$ for $j = 2$.

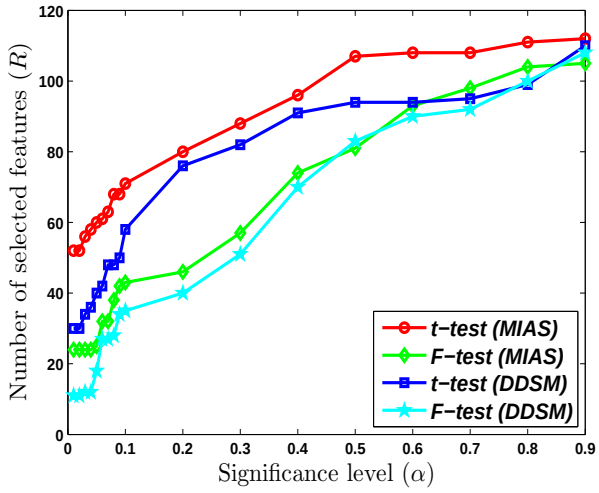
Mammogram Database	Type of ROI	Detail coefficient	Feature Descriptors				
			FD_1	FD_2	FD_3	FD_4	FD_5
MIAS	Normal	H_1	0.2264	480.5639	2.1465	70.6623	9.7249
		H_2	0.0879	681.0941	2.8792	42.3220	8.1304
		V_1	0.1864	467.7653	1.9784	60.9509	9.0844
		V_2	0.0926	766.4563	2.8347	51.7463	8.8526
		D_1	0.2174	443.6977	1.8201	64.9795	9.2851
		D_2	0.2095	348.9067	2.2437	43.2632	7.9568
	Benign	H_1	0.1974	415.2558	2.0749	54.6695	8.6772
		H_2	0.0906	858.3919	2.8054	58.5568	9.2685
		V_1	0.3516	241.0157	1.6931	47.6765	8.0624
		V_2	0.0898	748.6347	2.8727	47.1760	8.5424
		D_1	0.2249	434.0044	1.8205	65.2272	9.3105
		D_2	0.1199	668.5166	2.5641	56.9308	9.0802
	Malignant	H_1	0.1883	666.8222	2.0078	91.3533	10.8204
		H_2	0.1233	643.4498	2.4670	56.9562	9.0486
		V_1	0.4392	216.4788	1.4470	50.4909	8.0867
		V_2	0.2056	494.2931	2.0431	65.5455	9.4444
		D_1	0.3515	259.9105	1.5157	52.6011	8.3250
		D_2	0.1197	643.0346	2.6010	53.7630	8.8819
DDSM	Normal	H_1	0.2847	276.8162	1.9086	46.7880	8.0924
		H_2	0.1485	827.6910	2.2421	94.6770	11.0822
		V_1	0.1721	459.9882	2.1365	55.0801	8.7910
		V_2	0.0978	671.1874	2.7181	49.0870	8.5902
		D_1	0.1449	550.2200	2.2622	56.8473	8.9557
		D_2	0.2114	329.7618	2.1021	44.7035	8.0979
	Benign	H_1	0.9139	213.2245	0.2897	95.4060	9.9831
		H_2	0.2050	271.6973	1.8368	33.2631	6.9557
		V_1	0.9139	213.1512	0.2776	95.0765	9.9784
		V_2	0.5371	184.2814	1.1599	47.4170	7.7424
		D_1	0.7210	148.7421	0.5919	54.2596	7.8378
		D_2	0.9657	130.4118	0.1256	62.1110	7.9824
	Malignant	H_1	0.9123	212.2347	0.2843	95.0021	9.9747
		H_2	0.3243	231.1132	1.5688	41.1360	7.4945
		V_1	0.9123	214.6300	0.2834	95.1750	9.9846
		V_2	0.3078	239.7369	1.5993	40.3093	7.4633
		D_1	0.3352	285.1724	1.3013	58.1961	8.5885
		D_2	0.7873	143.3512	0.5003	55.8630	7.8764

Table 2.5: Different values of various feature descriptors at $\theta = 135°$ with set distance $D = 1$ for $j = 1$ and $D = 2$ for $j = 2$.

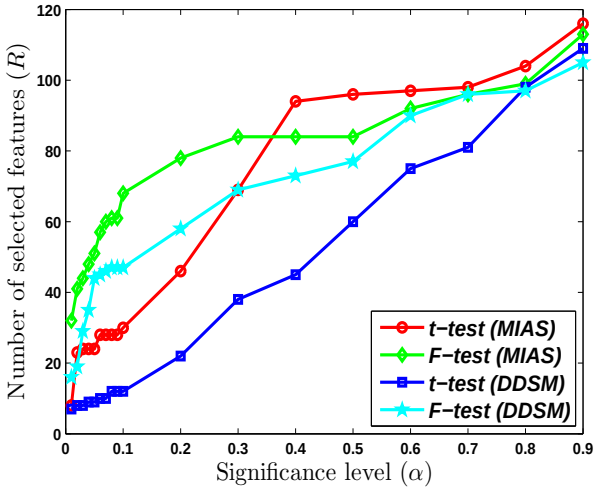
Mammogram Database	Type of ROI	Detail coefficient	Feature Descriptors				
			FD_1	FD_2	FD_3	FD_4	FD_5
MIAS	Normal	H_1	0.2269	480.7278	2.1466	70.6295	9.7249
		H_2	0.0900	684.9990	2.8778	42.1644	8.1296
		V_1	0.1861	467.8358	1.9786	60.8743	9.0849
		V_2	0.0954	760.9155	2.8241	51.8240	8.8542
		D_1	0.2164	443.6506	1.8209	65.1938	9.2854
		D_2	0.2122	347.3934	2.2333	43.1518	7.9576
	Benign	H_1	0.1970	414.9199	2.0684	54.9266	8.6774
		H_2	0.0910	859.0862	2.8037	58.3728	9.2670
		V_1	0.3523	240.9230	1.6918	47.7658	8.0624
		V_2	0.0901	743.2662	2.8825	47.8501	8.5424
		D_1	0.2239	433.9719	1.8191	65.1063	9.3110
		D_2	0.1183	672.4989	2.5709	56.7228	9.0802
	Malignant	H_1	0.1880	666.9464	2.0085	91.2963	10.8207
		H_2	0.1227	647.4172	2.4699	56.9847	9.0494
		V_1	0.4382	216.5055	1.4477	50.4490	8.0867
		V_2	0.2016	492.9544	2.0538	65.8033	9.4437
		D_1	0.3474	259.8169	1.5155	52.7306	8.3250
		D_2	0.1165	642.1973	2.6072	53.6483	8.8819
DDSM	Normal	H_1	0.2881	276.6600	1.9064	47.0070	8.0924
		H_2	0.1502	828.8966	2.2446	94.3992	11.0817
		V_1	0.1733	459.8334	2.1368	55.1521	8.7912
		V_2	0.0978	671.1200	2.7155	49.0154	8.5895
		D_1	0.1448	549.9213	2.2617	56.9170	8.9548
		D_2	0.2160	329.5301	2.1044	44.8295	8.0979
	Benign	H_1	0.9139	213.2773	0.2873	95.4790	9.9833
		H_2	0.2068	271.7451	1.8327	33.4358	6.9550
		V_1	0.9139	2132.0892	0.2766	95.1265	9.9782
		V_2	0.5364	184.2615	1.1616	47.362	7.7424
		D_1	0.7229	148.6612	0.5919	54.2346	7.8348
		D_2	0.9671	130.4906	0.1244	62.0631	7.9820
	Malignant	H_1	0.9123	212.2300	0.2835	95.0051	9.9747
		H_2	0.3285	232.2644	1.5734	40.9761	7.4945
		V_1	0.9123	214.5783	0.2840	95.1964	9.9838
		V_2	0.3088	239.6123	1.6006	40.2593	7.4645
		D_1	0.3358	285.1300	1.3011	58.2996	8.5887
		D_2	0.7916	143.4742	0.4983	55.8124	7.8766

2.6.2 Results for Feature Selection and Classification

During the experiments, different number of significant features (R) have been selected through *two-sample t-test* and *F-test* methods. Figure 2.8 shows the variation of the number of selected features (R) with respect the various values of significant level (α) for MIAS and DDSM databases. It has been observed that, the reduced number of selected features (R) is obtained at lower values of significance level (α) using both statistical methods. It is also observed that for same value of α, the dimension reduction is more in DDSM images as compared to MIAS images. The selected features are used in the classifier to find the optimal classification accuracy rate.



(a) Abnormal-normal class set



(b) Malignant-benign class set

Figure 2.8: Feature selection by *two-sample t-test* and *F-test* method.

After getting several sets of significant features, we conducted the classification experiments on both MIAS and DDSM dataset using a three-layered BPNN. In the experiment, 70% of the total dataset have been used for training. From the remaining dataset, 15% data are used for validation and rest 15% are used for testing purposes.

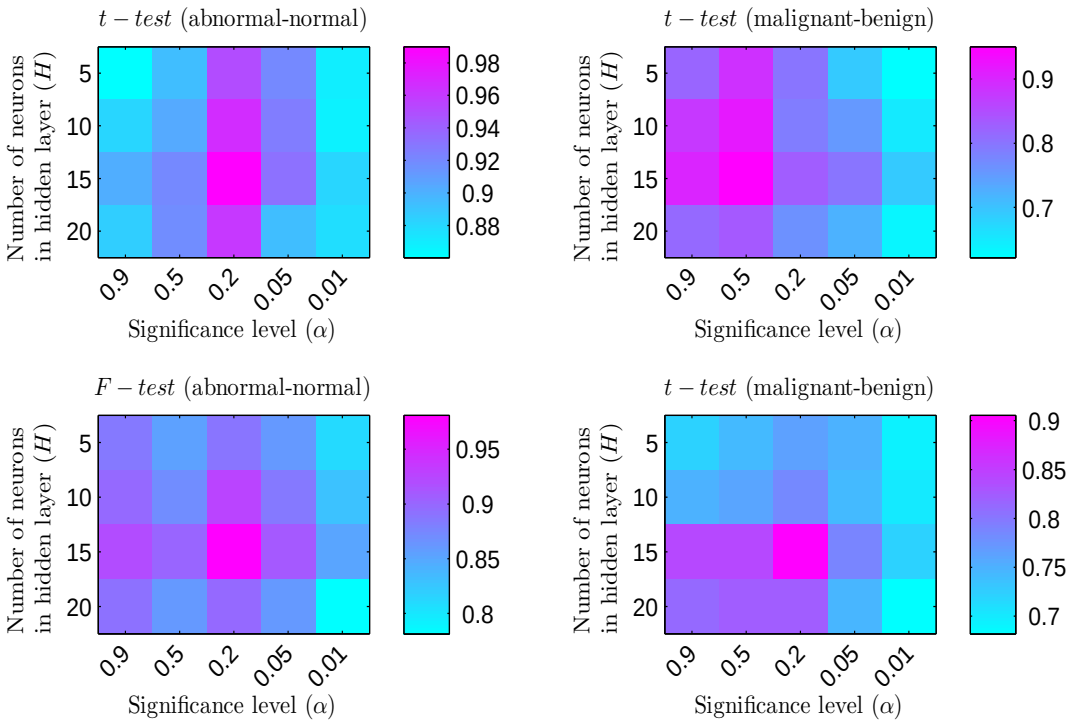
As mentioned in the proposed model (Figure 2.1), the magnitude of significance level (α) for feature selection and number of neurons in the hidden layer (H) of the BPNN influence the performance of the classifier. It is very difficult to find the best significant feature set through which the classifier achieves optimal performance. Therefore, several feature sets obtained at various values of significance level (α) are used in the classifier to find the optimum results. In fact, for the same value of α, the classifier achieves different performance results at the different number of hidden neurons (H).

In our experiments, the values of H have been chosen as 5, 10, 15 and 20 to investigate the best performance. It has been found that, at $H = 15$ with respect to different α, the classifier attains its best performance. Different performance measures, including sensitivity (S_n), specificity (S_p) and test classification accuracy (ACC) using two feature selection methods are presented in Table 2.6. It is observed that, the higher classification accuracy rates are obtained with *two-sample t-test* feature selection method for both the databases. These values are as 98.13% (abnormal–normal), 94.2% (malignant–benign) for MIAS database, and 98.8% (abnormal–normal), 97.4% (malignant–benign) for DDSM database.

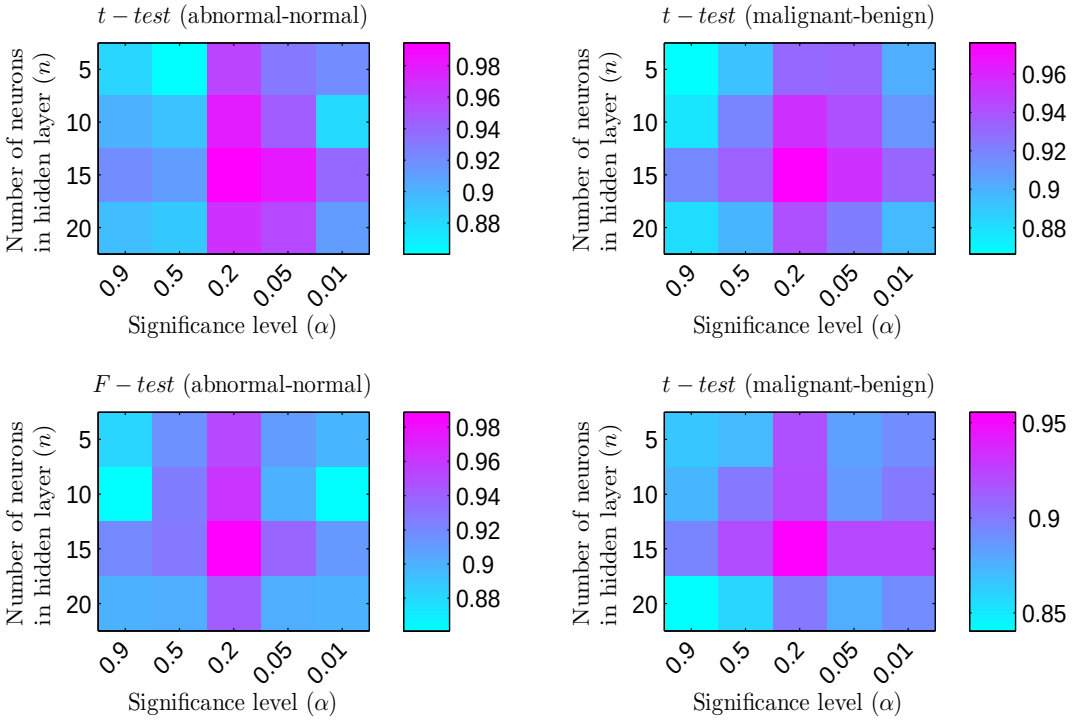
We have also evaluated the performance of two feature selection methods by comparing the obtained AUC values of ROC curves at different magnitudes of significance level (α) with respect to the different number of hidden neurons (H) in BPNN classifier. A heat-map has been used to demonstrate the comparison as shown in Figure 2.9. It is clearly observed that, the best values of AUC have been accomplished with the significance level (α) of 0.2 for classification of MIAS and DDSM datasets. One tenuous deviation in AUC is observed at $\alpha = 0.5$ for malignant–benign classification in MIAS data. This might be due to some irregular tissue pattern in mammograms.

Table 2.6: Different values of performance measures of the classifier using two feature selection methods with $H = 15$.

Mammogram Database	Selection method	α	Abnormal-normal				Malignant-benign			
				Measures (%)				Measures (%)		
			R	S_n	S_p	ACC	R	S_n	S_p	ACC
MIAS	***Two-sample t-test***	0.9	112	77.80	93.30	87.50	116	87.5	88.90	88.20
		0.5	107	88.2	93.30	91.70	**96**	**100**	**90.00**	**94.20**
		0.2	**80**	**100**	**97.00**	**98.13**	46	77.80	87.50	82.40
		0.05	60	85.70	97.10	93.80	24	75.00	77.80	76.50
		0.01	52	100	77.40	85.40	08	66.70	54.50	58.90
	Two-sample F-test	0.9	105	69.20	100	91.70	113	87.50	77.80	82.40
		0.5	81	75.00	93.80	87.50	84	100	66.70	82.40
		0.2	46	88.20	100	95.80	78	88.90	87.50	88.20
		0.05	25	82.60	96.00	89.60	51	87.50	66.70	76.50
		0.01	24	64.70	93.50	83.30	32	62.50	77.80	70.6
DDSM	***Two-sample t-test***	0.9	110	93.50	86.40	90.30	109	93.30	86.9	89.4
		0.5	94	91.40	83.30	87.90	60	94.10	90.40	92.10
		0.2	**76**	**100**	**97.90**	**98.80**	**22**	**100**	**94.70**	**97.40**
		0.05	40	100	95.10	97.60	09	92.30	96.00	94.70
		0.01	30	97.80	89.40	93.90	07	93.30	91.30	92.10
	Two-sample F-test	0.9	108	93.60	84.60	89.10	105	86.60	86.90	86.80
		0.5	83	93.30	89.40	91.50	77	93.30	86.90	89.40
		0.2	40	100	95.10	97.60	58	93.80	90.90	92.10
		0.05	18	95.60	92.10	93.90	44	86.70	91.70	89.40
		0.01	11	95.40	84.60	90.30	16	92.30	88.00	89.50



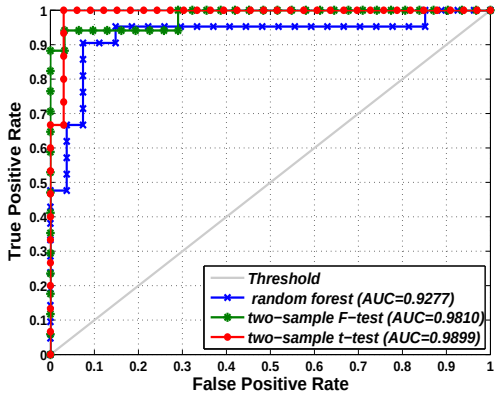
(a) MIAS Database



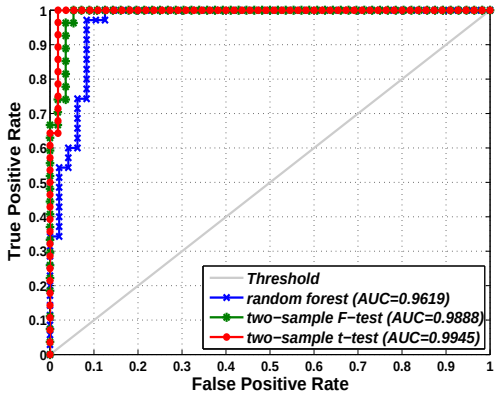
(b) DDSM Database

Figure 2.9: Heat-maps of AUC measurements using significant feature sets.

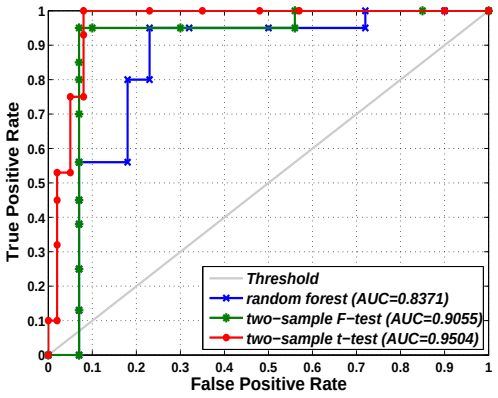
During our experiment, we have compared the performances achieved by the BPNN classifier along with statistical *two-sample t-test* and *F-test* method with random forest method [65]. ROC curves obtained using the proposed feature selection schemes, and the random forest method are shown in Figure 2.10. It has been inferred that the proposed schemes outperform the random forest method with respect to *AUC* measurements. Table 2.7 presents the comparison of the test accuracies and *AUC* measurements for *two-sample t-test* and *F-test*, and random forest technique. The maximum *AUC* values obtained by the BPNN and *t-test* method are 0.9899 and 0.9504 in MIAS, and 0.9945 and 0.9761 in DDSM database for both abnormal–normal and malignant–benign pattern classification. It has been clearly observed that the *two-sample t-test* has quite higher-performance values in comparison to other methods mentioned for both databases.



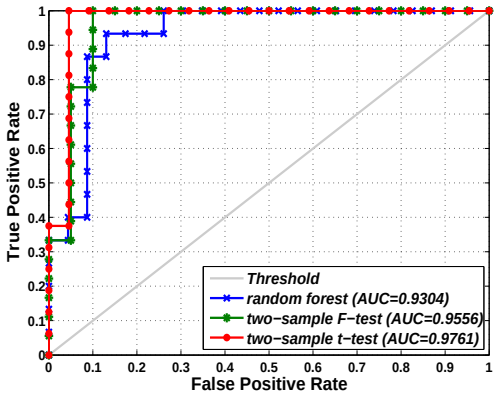
(a) Abnormal–normal (MIAS)



(b) Abnormal–normal (DDSM)



(c) Malignant–benign (MIAS)



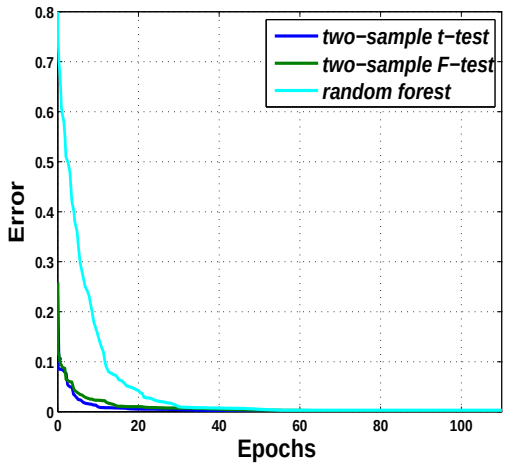
(d) Malignant–benign (DDSM)

Figure 2.10: Comparison of ROC curves for both image class sets using proposed feature selection schemes and random forest method with help of BPNN.

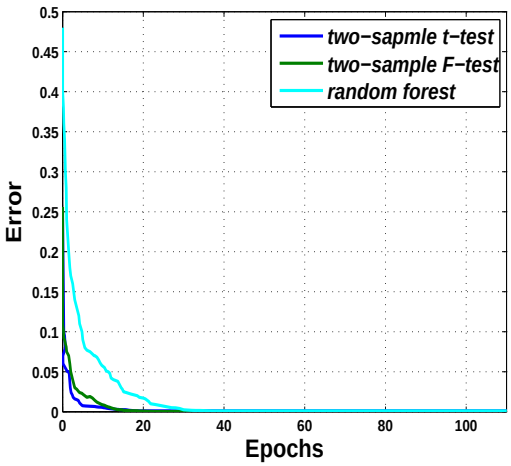
Table 2.7: Comparison of optimal test *ACC* and *AUC* measurements between proposed and random forest methods.

Mammogram Database	Image class set	Performance measures					
		ACC (%)			*AUC*		
		t-test	*F-test*	random forest	***t-test***	*F-test*	random forest
MIAS	Abnormal–normal	**98.13**	95.80	93.30	**0.9899**	0.9810	0.9277
	Malignant–benign	**94.20**	88.20	82.40	**0.9504**	0.9055	0.8371
DDSM	Abnormal–normal	**98.80**	97.60	92.80	**0.9945**	0.9888	0.9619
	Malignant–benign	**97.40**	92.10	89.50	**0.9761**	0.9556	0.9304

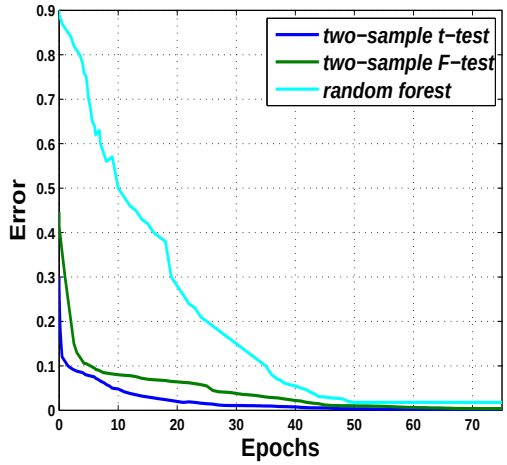
Further, a training error comparison has been made for the proposed scheme and random forest method as shown in Figure 2.11 to evaluate the training convergence as one of the performance indices. The training error of the classifier is expressed as mean squared error (MSE) values at multiple numbers of training iteration in BPNN classifier. The mean squared error of a predictor measures the average of the squares of the errors, i.e., the difference between the predicted and actual. Regarding this context, the mean squared error is the average squared difference between output classes generated by the classifier and existing actual classes. The BPNN adjusts the weights and biases of the network in order to minimize the mean squared error. The weights of hidden layer neurons are adjusted in direct proportion to the error in the neuron to which it is connected. The training error curves of *two-sample t-test* method shows that it converges faster than other selection methods for both abnormal–normal and malignant–benign mammogram classes.



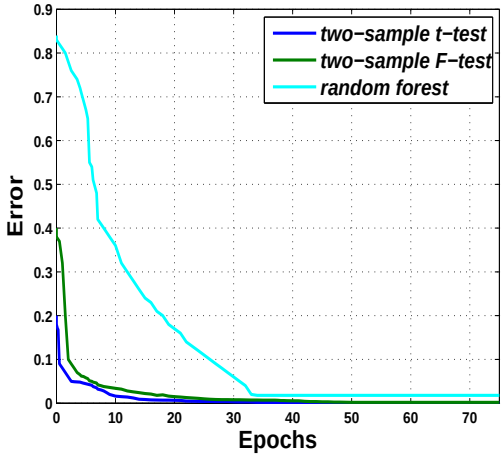
(a) Abnormal-normal class set (MIAS)



(b) Abnormal-normal class set (DDSM)



(c) Malignant-benign class set (MIAS)



(d) Malignant-benign class set (DDSM)

Figure 2.11: Training error comparison by BPNN using *two-sample t-test* and *F-test*, and random forest feature selection methods.

The comparison of accuracies achieved by the BPNN and other classifiers utilizing the relevant DWT and GLCM features selected by *t-test* is given in Table 2.8. It has been observed that the BPNN achieves the optimal accuracies using the hybrid DWT and GLCM features for both the MIAS and DDSM databases as compared to other classifiers.

Table 2.8: Comparison of accuracies ($ACC(\%)$) achieved by different classifiers utilizing the relevant features selected by *t-test* method.

Classifier	MIAS		DDSM	
	Abnormal–normal	Malignant–benign	Abnormal–normal	Malignant–benign
NB	91.67	82.35	95.18	89.47
K-NN	89.58	76.47	92.77	86.84
BPNN	**98.13**	**94.20**	**98.80**	**97.40**
SVM	85.40	76.47	90.36	84.21
AdaBoost-RF	90.56	88.52	96.38	95.00
LogitBoost-RF	91.26	90.11	97.15	96.68
LMT	93.75	88.23	96.38	92.10

Finally, a comparative analysis between the proposed scheme with other existing schemes has been made and shown in Table 2.9. It has been clearly observed that the proposed scheme performs better classification than other schemes with respect to different performance measures.

Table 2.9: Performance comparison by different approaches with the proposed scheme.

Approach	**Technique**	**Database**	**Performance measures**
Prathibha *et al.* (2010) [27]	DWT, ANN	MIAS	$AUC = 0.95$ (Abnormal–normal)
Buciu *et al.* (2011) [31]	Gabor wavelets and PCA, SVM	MIAS	$S_n = 97.56\%$, $S_p = 60.86\%$ $AUC = 0.79$ (Abnormal–normal)
			$S_n = 84.61\%$, $S_p = 80.0\%$ $AUC = 0.78$ (Malignant–benign)
Mutaz *et al.* (2011) [32]	GLCM, ANN	DDSM	$S_n = 91.6\%$, $S_p = 84.17\%$ (Malignant–benign)
Jona *et al.* (2012) [37]	GLCM, SVM	MIAS	$ACC = 94.0\%$ (Abnormal–normal)
Görgel *et al.* (2013) [42]	SWT, SVM	I.U. Database	$ACC = 96.0\%$ (Abnormal–normal)
			$ACC = 93.59\%$ (Malignant–benign)
Zhang *et al.* (2016) [60]	fractional Fourier transform, PCA, SVM	MIAS	$S_n = 92.22\%$, $S_p = 92.10\%$ $ACC = 93.59\%$ (Malignant–benign)
Proposed scheme	**DWT+GLCM +*t-test*+BPNN**	**MIAS**	$\boldsymbol{S_n = 100\%}$, $\boldsymbol{S_p = 97.00\%}$ $\boldsymbol{ACC = 98.13\%}$, $\boldsymbol{AUC = 0.9899}$ **(Abnormal–normal)**
			$\boldsymbol{S_n = 100\%}$,$\boldsymbol{S_p = 90.00\%}$ $\boldsymbol{ACC = 94.20\%}$, $\boldsymbol{AUC = 0.9504}$ **(Malignant–benign)**
		DDSM	$\boldsymbol{S_n = 100\%}$,$\boldsymbol{S_p = 97.90\%}$ $\boldsymbol{ACC = 98.80\%}$, $\boldsymbol{AUC = 0.9945}$ **(Abnormal–normal)**
			$\boldsymbol{S_n = 100\%}$, $\boldsymbol{S_p = 94.70\%}$ $\boldsymbol{ACC = 97.40\%}$, $\boldsymbol{AUC = 0.9761}$ **(Malignant–benign)**

2.7 Summary

In this chapter, an efficient mammogram classification scheme has been proposed to support the decision of radiologists. The scheme utilizes DWT and GLCM in succession to derive the hybrid features form mammograms. To select the relevant features from the feature matrix, both *t-test* and *F-test* have been applied independently. The efficiency of the proposed selection methods has been compared with random forest technique. To validate the efficacy of the suggested scheme, simulation has been carried out using several other classifiers namely, Back-Propagation Neural Network (BPNN), Naive Bayes (NB), K-Nearest Neighbor (K-NN), Support Vector Machine (SVM), AdaBoost and Random Forest (AdaBoost-RF), LogitBoost and Random Forest (LogitBoost-RF), and Logistic Model Tree (LMT) for both MIAS and DDSM databases. It has been observed that, *t-test* based relevant features achieve higher classification accuracy with the help of BPNN classifier. An accuracy of 98.13% and 94.20% have been obtained for abnormal–normal and malignant–benign respectively for MIAS database. The similar parameters are 98.80% and 97.40% achieved for DDSM databases. Furthermore, the competent schemes are also simulated in the similar platform, and comparative analysis with respect to accuracy (ACC) and AUC of ROC reveals that the suggested scheme outperforms other schemes.

Chapter 3

Mammogram Classification using SFTA Features with FCBF Feature Selection

---

All over the world, the causes of cancer-related deaths among women are due to breast cancer. Early detection of breast cancer improves the reduction in death rates. Currently, mammography is the most reliable radiological screening method for detection of the abnormality in the breast. In mammography, the X-ray images known as mammograms are analyzed for the abnormality detection. Reading of mammograms is a very important task for radiologists as they suggest patients for biopsy. However, the reading result varies among radiologists as it depends on experience. It has been observed that most of the predicted abnormal tissues by the radiologist found normal in the biopsy. To overcome this problem, Computer-Aided Diagnosis (CAD) method has been developed in order to help radiologists for accurate diagnosis. In a CAD of breast cancer, the crucial task is to find out the significant features from the mammograms to characterize them as malignant, benign or normal. This chapter presents an effective scheme to identify the abnormal mammograms in order to detect the breast cancer. The scheme utilizes the Segmentation-based Fractal Texture Analysis (SFTA) to extract texture features from the mammograms for the classification of mammograms. A fractal analysis has been applied to collect the qualitative information of texture features. A Fast Correlation-Based Filter (FCBF) method has been used to select feature subsets containing significant features, which are used for classification purpose. To validate the efficacy of proposed scheme

different classifiers have been used among which, Support Vector Machine (SVM) gives a better performance. In this work, a detail study of classification performed by SVM has been explained. The overall block diagram of the proposed scheme is shown in Figure 3.1.

The organization of the chapter is as follows: The extraction of feature using SFTA is described in Section 3.1. Section 3.2 explains the selection of significant features utilizing FCBF method. The classification and evaluation of performance is outlined in Section 3.3. The experimental results are presented in Section 3.4. The overall work proposed in this chapter is summarized in Section 3.5.

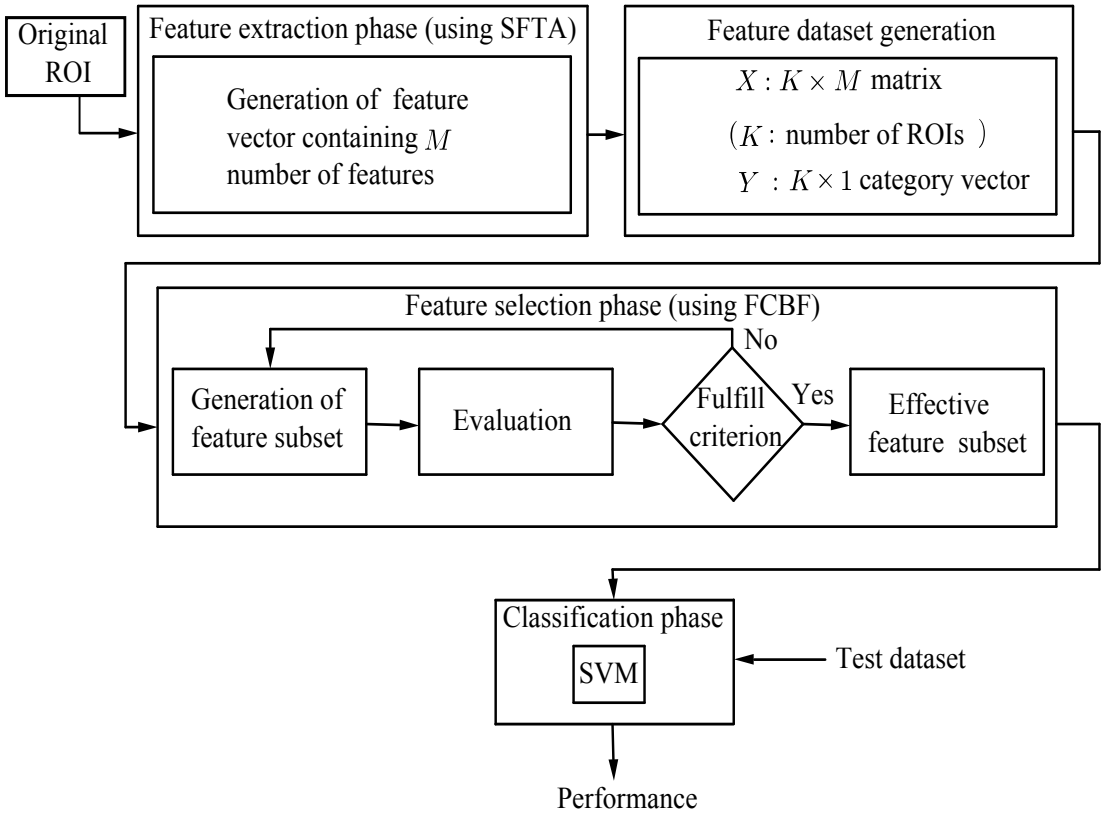


Figure 3.1: Block diagram of proposed scheme using SFTA and FCBF method.

3.1 Feature Extraction using SFTA

In Segmentation-based Fractal Texture Analysis (SFTA) of mammograms, the features are extracted from the mammographic Regions-of-Interest by decomposing them into a set of binary images from which fractal dimensions are computed [75]. These dimensions are used to describe the segmented texture patterns of the mammograms. For the decomposition of input gray-scale Regions-of-Interest

(ROIs), a two-threshold binary decomposition method is applied. In this type of decomposition, a set of n number of threshold values of gray-scale ROI is computed by applying multilevel Ostu algorithm recursively, where n is the user defined parameter. A set of binary Regions-of-Interest ($bROI$s) are computed from a gray-scale ROI by applying a two-threshold binary decomposition. In this method, pairs of lower and upper and threshold values (t_L, t_U) are selected from the set of threshold values, T and the gray-scale ROI is decomposed as,

$$bROI\,(x,y) = \begin{cases} 1, & \text{if } \; t_L < ROI(x,y) \leq t_U \\ 0 \end{cases} . \tag{3.1}$$

In two threshold binary decomposition of gray-scale ROI, a $2n$ number of $bROI$s is obtained by using all pairs of threshold values from the set T. After the binary image decomposition, three components such as fractal dimension, mean gray level, and region area of resulting $bROI$s are computed. The fractal dimension of $bROI$ is generated from a texture of its border image. The border image is the regions of boundaries of $bROI$ computed by using the set of 8-connected pixels. The border image has the value one at pixel (x, y) on the corresponding $bROI(x, y) = 1$, and having at least one neighboring pixel zero. Otherwise, the border image takes the value zero.

The fractal dimension (D) measures the degree of irregularities of an image. Various approaches are there to compute the fractal dimension of an image, but box-counting is the most common method in this regard [76]. The fractal dimension (D) of a $bROI$ is computed with the help of a grid of squares each having the size s and given as,

$$D = -\lim_{s \to 0} \frac{\log\,(N_s)}{\log\,(s)} \tag{3.2}$$

where N_s is the number of squares needed to cover the portion of $bROI$. Now the feature vector, V is constructed for each ROI using the series collection of the three components, fractal dimension (D), mean gray level ($\overline{GL}$), and region area (A) of resulting $bROI$s as described in **Algorithm 5**. Thus, the feature extraction algorithm gives the feature vector for every ROIs, and consequently; a dataset X is also obtained containing K number of ROIs and M number of features. The value of M is $6n$ where n is the number of thresholds given to the algorithm.

Algorithm 5 Feature extraction and dataset generation using SFTA.

Require: K: Total number of ROIs, n: Number of thresholds t_U, t_L: Upper and lower thresholds respectively, G_L: Maximum possible gray level of an ROI.

Ensure: $X[K][M]$: Matrix contains the feature data. Function $BinaryDecomp()$ decomposes gray-scale ROI to a set of $bROI$s.

Initialize the required value to n
$m \leftarrow 2 \times n$ and $M \leftarrow 3 \times m$
for $i \leftarrow 1$ to K **do**
  Get ROI_i; $T_i \leftarrow \text{Ostu}(ROI_i, n)$; $R \leftarrow |T_i|$
  for $p \leftarrow 1$ to $R - 1$ **do**
    $T_1 \leftarrow \{\{t_p, t_{p+1}\} : t_p, t_{p+1} \in T_i\}$
  end for
  for $q \leftarrow 1$ to R **do**
    $T_2 \leftarrow \{\{t_q, G_L\} : t_q \in T_i\}$
  end for
  $l \leftarrow 1$
  for $j \leftarrow 1$ to m **do**
    $bROI_{ij} \leftarrow BinaryDecomp(ROI_i, t_L, t_U) \forall \{\{t_L, t_U\} \in \{T_1, T_2\}\}$
    Compute D_{ij}, $\overline{GL_{ij}}$ and A_{ij} of $bROI_{ij}$
    $V_i[l] \leftarrow D_{ij}$; $V_i[l+1] \leftarrow \overline{GL_{ij}}$; $V_i[l+2] \leftarrow A_{ij}$
    $l \leftarrow l + 1$
  end for
  Append V_i to X
end for

3.2 Feature Selection using FCBF Method

Fast Correlation-Based Filter (FCBF) is a filter model of feature selection proposed by Yu *et al.* [77], based on the correlation between feature and class. The correlation can be measured by a desired property known as symmetry uncertainty. The symmetry uncertainty (SU) can measure the effectiveness of the feature and limits the biasing effect in favor of features with more values. The FCBF selects a set of relevant features S, that is highly correlated to the class with $SU \geq t$, where t is predefined threshold and SU is defined as,

$$SU(X, Y) = 2\left[\frac{IG(X|Y)}{H(X) + H(Y)}\right] \tag{3.3}$$

where $IG(X,Y)$ is the information gain, and $H(X)$, $H(Y)$ and $H(X|Y)$ are entropies. Information gain of a feature X and class Y is computed as,

$$IG(X,Y) = H(X) - H(X|Y). \tag{3.4}$$

Entropy measures the uncertainty of a variable. $H(X)$ measures the entropy of feature X, and $H(X|Y)$ measures the entropy of feature X after observing class Y. They are defined as,

$$H(X) = -\sum_i P(x_i)\log_2(P(x_i)), \tag{3.5}$$

$$H(X|Y) = -\sum_j P(y_j) - \sum_i P(x_i|y_j)\log_2(P(x_i|y_j)), \tag{3.6}$$

where $P(x_i)$ is the prior probabilities for all values of X, and $P(x_i|y_j)$ is the posterior probabilities of X given the values of Y. The features of mammographic ROIs and class labels are structured in the dataset X and category vector Y as shown in Figure. 3.2. Each row in the dataset X is a feature vector of all instances, $V_i = \{f_1, f_2, ..., f_M\}$, $1 \leq i \leq K$ and each column is a part of instance. The column vector contains the class labels of the instances. In our problem, the total number of classes are two, as all the instances are belonging to either normal or abnormal class.

$$X = \begin{array}{c} \\ ROI_1 \\ ROI_2 \\ \cdot \\ \cdot \\ \cdot \\ ROI_K \end{array} \begin{array}{c} \begin{array}{cccccc} f_1 & f_2 & \cdot & \cdot & \cdot & f_M \end{array} \\ \begin{bmatrix} x_{11} & x_{12} & \cdot & \cdot & \cdot & x_{1M} \\ x_{21} & x_{22} & \cdot & \cdot & \cdot & x_{2M} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{K1} & x_{K2} & \cdot & \cdot & \cdot & x_{KM} \end{bmatrix} \end{array} \begin{array}{l} \\ \leftarrow V_1 \\ \leftarrow V_1 \\ \cdot \\ \cdot \\ \cdot \\ \leftarrow V_K \end{array} \qquad Y = \begin{bmatrix} l_1 \\ l_2 \\ \cdot \\ \cdot \\ \cdot \\ l_K \end{bmatrix}$$

Figure 3.2: The structure of dataset X and category vector Y.

FCBF computes $SU_{i,l}$ value of every feature f_i ,$1 \leq i \leq M$ in the dataset X and compares with t value. Here $l \in Y$ is the class label of the mammographic ROIs. Then, the feature f_i is predominant if $SU_{i,l} \geq t$ ($\forall f_i \in S$, $1 \leq i \leq M$, $SU_{i,l} \geq t$). A feature f_j with $SU_{j,i} \geq SU_{i,l}$ $\forall f_j \in S$, $i \neq j$ does not exist in the dataset, and if it exists, then it is redundant to f_i. If two features are redundant, then FCFB applies a heuristic to

remove one of them from the feature set. The heuristic is based on removing one of the redundant features that is less relevant to the class labels. Finally, a set S has been selected, which contains the most significant and non redundant features. Further, various feature subsets S_{ij}, $S_{ij} \in S_i$, $1 \leq i \leq K$, are constructed from selected feature set using different combination of relevant features, where j represents the number of relevant features used for combination. These subsets are used in the classification phase for the characterization of abnormal mammograms.

3.3 Classification and Evaluation of Performance

A comparative classification of mammograms is carried out by using the reduced set of features. Here we have used a Support Vector Machine (SVM) as classifier with 10-fold cross-validation. The goal of machine learning classifier is to build a model which makes accurate predictions on the training set. But, the training dataset accuracy is not a good indication of better performance of the classifier; however, it depends on how well the classifier will perform when classifying the new data outside the training dataset. Therefore, the classifier needs some effective measure to provide adequate accuracy when it will be deployed. For this purpose, cross-validation process is used to provide a much truer accuracy of the classifier. In the cross-validation process, the dataset is divided into a large training set and a smaller validation set. The classifier is then trained on the training set and use the validation set to measure the accuracy. For the division of dataset into training and validation set, K fold cross-validation method is used. The cross-validation process is continued in K rounds. In each round, one fold is selected for validation and remaining $(K-1)$ folds are combined and used for training purpose. Then the classifier is trained by the training dataset and accuracy is measured on the validation data for each round. After K rounds, the average of all obtained accuracies is calculated to get the final cross-validation accuracy.

During the training period, the classifier is fed with separate selected feature subsets of MIAS database. A number of cross-validation accuracies are obtained corresponding to the different feature subsets. The optimal feature subset is determined on the basis of high cross-validation accuracy. The testing dataset from both MIAS and DDSM database is provided to classifiers namely Support Vector Machine (SVM), Back-Propagation Neural Network (BPNN), AdaBoost and Random Forest (AdaBoost-RF), LogitBoost and Random Forest (LogitBoost-RF), and Logistic Model Tree (LMT) for an independent measurement of performances. In addition,

Naive Bayes (NB) and K-Nearest Neighbor (K-NN) classifiers are also used to compare the obtained accuracies utilizing the same proposed features for both the databases. The performance of the proposed scheme has been assessed with the assistance of various parameters namely, sensitivity (S_n), specificity (S_p), positive predictive value (PPV), negative predictive value (NPV), classification accuracy (ACC), and AUC value of ROC curve.

3.4 Experimental Results and Discussion

For extraction of features, SFTA method is applied to all mammographic ROIs of size of 128×128 pixels of MIAS and DDSM databases. The feature extraction algorithm uses the various numbers of threshold values starting from 2 onwards. For each image, using n number of threshold values, the algorithm gives $2n$ number of binary images and subsequently gives $6n$ number of features. Suppose, using 10 threshold values, the algorithm extracts a total of 60 features from a mammographic ROI. Thus, various numbers of threshold values can be employed in extracting the feature in the algorithm. However, our objective in feature extraction algorithm is to find out the optimal number of threshold values, which are to be used to extract the feature sets from which relevant subsets can be extracted. For this purpose, all the extracted feature sets are given to the SVM classifier, and a 10-fold cross-validation operation is performed for a number of rounds.

In the experiment, the whole dataset is partitioned into 10 folds. In each round, nine folds are used together for training of the classifier and remaining one fold is used as validation purpose. This procedure continues for 10 rounds with different validation sets and 10 number of validation accuracies are obtained. Then a final cross-validation accuracy is found by averaging the obtained accuracies in each round. Thus, the SVM classifier determines several cross-validation accuracies for various feature sets of MIAS dataset using the different number of threshold values as shown in Figure. 3.3. From Figure. 3.3, it has been observed that, the validation accuracies of the classifier are optimal employing the number of threshold values as, $n = 6, 7, 8, 9$ and 10. For the number of threshold values, $n = 2$ to 5, the accuracy rate rises sharply and from $n = 11$ and onwards, the accuracy rates falls slowly and become constant but less than the optimal accuracy values. Therefore, we have taken feature sets corresponding to the number of threshold values from $n = 6$ to 10 for effective feature subset selection. Various feature vectors containing different numbers of features for all optimal numbers of threshold values are shown in Table 3.1. In FCBF method,

different relevant feature sets selected from the respective feature vectors are also shown in Table 3.1.

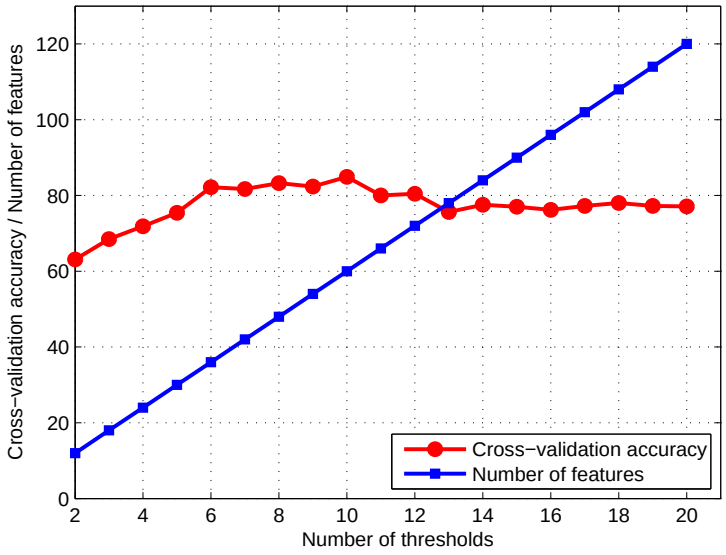


Figure 3.3: Number of features with their respective cross-validation accuracies obtained using different number of threshold values (MIAS database).

Table 3.1: Selected feature sets containing relevant features by FCBF method (MIAS database).

Number of threshholds (n)	Extracted feature vector (V_n)	Total Number of features (f)	Selected feature set with relevant features (SF_n)
6	V_6	36	$S_6 = \{f_{35}, f_{10}, f_{34}\}$
7	V_7	42	$S_7 = \{f_{41}, f_{10}, f_{16}\}$
8	V_8	48	$S_8 = \{f_{47}, f_{13}, f_{19}\}$
9	V_9	54	$S_9 = \{f_{53}, f_{13}, f_{43}\}$
10	V_{10}	60	$S_{10} = \{f_{59}, f_{16}, f_{49}\}$

Further, various feature subsets $SF_{i,j}$, $SF_{i,j} \in SF_i$ containing optimal relevant feature combinations are given to the classifier for adequate discrimination of

abnormal–normal mammograms. In our experiment, a 10-fold cross-validation operation has been performed using the relevant feature subsets in the training phase of the classifier. Table 3.2 shows various selected feature subsets with their respective average cross-validation accuracies. From the experiment, it is observed that, relevant feature subset, S_{102} gives optimal cross-validation accuracy among all feature subsets.

Table 3.2: Feature subsets containing different combination of relevant features with corresponding cross-validation accuracies.

Selected feature set (SF_n)	Number of effective features used (p)	Selected feature subset (SF_{np})	Cross-validation accuracy (%)
S_6	1	$S_{61} = \{f_{35}\}$	87.25
	2	$S_{62} = \{f_{35}, f_{10}\}$	86.25
	3	$S_{63} = \{f_{35}, f_{10}, f_{34}\}$	85.92
S_7	1	$S_{71} = \{f_{41}\}$	86.30
	2	$S_{72} = \{f_{41}, f_{10}\}$	86.94
	3	$S_{73} = \{f_{41}, f_{10}, f_{16}\}$	85.84
S_8	1	$S_{81} = \{f_{47}\}$	92.01
	2	$S_{82} = \{f_{47}, f_{13}\}$	96.53
	3	$S_{83} = \{f_{47}, f_{13}, f_{19}\}$	88.37
S_9	1	$S_{91} = \{f_{53}\}$	91.83
	2	$S_{92} = \{f_{53}, f_{13}\}$	95.36
	3	$S_{93} = \{f_{53}, f_{13}, f_{43}\}$	85.41
S_{10}	1	$S_{101} = \{f_{59}\}$	92.34
	2	$\boldsymbol{S_{102} = \{f_{59}, f_{16}\}}$	**98.18**
	3	$S_{103} = \{f_{59}, f_{16}, f_{49}\}$	90.94

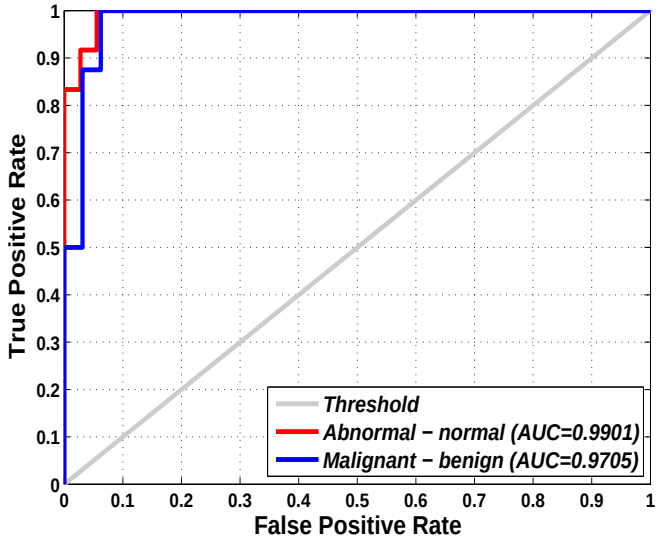
Now, using the optimal feature subset, different test samples are randomly selected from both the MIAS and DDSM databases and given to different classifiers for subsequent classification and performance evaluation. Different performance measures achieved by the SVM and other classifiers utilizing the relevant features selected by FCBF method on the same platform is given in Table 3.3. It has been observed that, SVM achieves better classification accuracy (ACC) of 98.76% (abnormal–normal) and 95.65% (malignant–benign) for MIAS database. The similar parameters of 99.20% (abnormal–normal) and 98.00% (malignant–benign) are achieved for DDSM database.

Table 3.3: Comparison of performances of various classifiers using optimal relevant feature set, S_{102}.

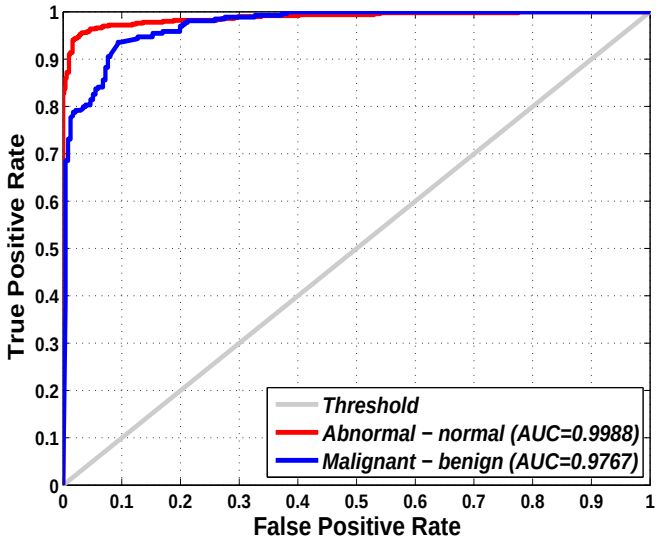
Database	Class set	Classifier	Measures of performance (%)				
			S_n	S_P	PPV	NPV	ACC
MIAS	Abnormal–normal	NB	96.62	85.22	92.17	93.33	92.55
		K-NN	99.03	69.57	85.42	97.56	88.51
		BPNN	98.07	88.70	93.98	96.23	94.72
		SVM	99.52	97.39	98.56	99.12	**98.76**
		AdaBoost-RF	96.62	93.91	96.62	93.91	95.65
		LogitBoost-RF	95.65	99.13	99.50	92.68	96.89
		LMT	99.03	95.65	97.62	98.21	97.83
	Maligant–benign	NB	85.94	94.12	94.83	84.21	89.57
		K-NN	93.75	78.43	84.51	90.91	86.96
		BPNN	90.63	90.20	92.06	88.46	90.43
		SVM	96.88	94.12	95.38	96.00	**95.65**
		AdaBoost-RF	95.31	86.27	89.71	93.62	91.30
		LogitBoost-RF	96.88	88.24	91.18	95.74	93.04
		LMT	95.31	92.16	93.85	94.00	93.91
DDSM	Abnormal–normal	NB	99.00	88.20	89.35	98.88	93.60
		K-NN	98.00	82.00	84.48	97.62	90.00
		BPNN	98.20	92.40	92.82	98.09	95.30
		SVM	98.40	100	100	98.43	**99.20**
		AdaBoost-RF	99.00	95.20	95.38	98.96	97.10
		LogitBoost-RF	99.20	96.00	96.12	99.17	97.60
		LMT	99.40	97.20	97.26	99.39	98.30
	Malignnt–benign	NB	93.22	89.02	88.35	93.63	91.00
		K-NN	92.37	85.98	85.49	92.65	89.00
		BPNN	95.34	90.53	90.00	95.60	92.80
		SVM	99.58	96.59	96.31	99.61	**98.00**
		AdaBoost-Rf	99.15	90.91	90.70	99.17	94.80
		LogitBoost-RF	97.88	92.42	92.03	97.99	95.00
		LMT	98.73	95.83	95.49	98.83	97.20

Further, ROC curves are computed by the SVM to study the efficiency of classification performance. Different test ROC curves with corresponding AUC values

of the classification using different optimal feature subset S_{102} for both MIAS and DDSM database are shown in Figure. 3.4. The optimal AUC values are 0.9901 (abnormal–normal) and 0.9705 (malignant–benign) for both MIAS database. For DDSM database, the similar parameters are 0.9988 and 0.9967.



(a) MIAS



(b) DDSM

Figure 3.4: Comparison of ROC curves achieved by optimal classifier (SVM).

Finally, the proposed scheme has been compared with other existing schemes and is shown in Table 3.4. It is observed that the proposed scheme performs better than other schemes in the characterization of mammograms.

Table 3.4: Performance comparison of other schemes with proposed scheme for classification of mammograms.

Scheme	**Technique**	**Database**	**Measurement**
Eltoukhy *et al.* (2012) [39]	Curvelet transform, statistical *t-test*, SVM	MIAS	$ACC = 95.98\%$ (Abnormal–normal) $ACC = 97.30\%$ (Malignant–benign)
Ramos *et al.* (2012) [38]	DWT, Random forest	DDSM	$AUC = 0.90$ (Abnormal–normal)
Jona *et al.* (2012) [37]	GLCM, PSO, SVM	MIAS	$ACC = 94.00\%$ (Abnormal–normal)
Oral *et al.* (2013) [45]	First and second order textural feature, PCA, MLP	MIAS	$ACC = 91.10\%$ (Abnormal–normal)
Görgel et al. (2013) [42]	SWT, SVM	I.U. Database	$ACC = 96.00\%$ (Abnormal–normal)
Proposed scheme	**SFTA+FCBF+SVM**	**MIAS**	$\boldsymbol{ACC = 98.76\%}$ $\boldsymbol{AUC = 0.9901}$ **(Abnormal–normal)** $\boldsymbol{ACC = 95.65\%}$ $\boldsymbol{AUC = 0.9705}$ **(Malignant–benign)**
		DDSM	$\boldsymbol{ACC = 99.20\%}$ $\boldsymbol{AUC = 0.9988}$ **(Abnormal–normal)** $\boldsymbol{ACC = 98.00\%}$ $\boldsymbol{AUC = 0.9967}$ **(Malignant–benign)**

3.5 Summary

In this chapter, an efficient scheme has been presented for classification of mammographic images as malignant, benign or normal to support the radiologist in the interpretation of digital mammograms. The scheme utilizes SFTA method to extract the features from the digital mammograms. An efficient feature selection technique FCBF has been used to select the most significant feature set from the extracted features. A promising classification performances of $ACC = 98.76\%$, $AUC = 0.9901$ (abnormal–normal), and $ACC = 95.65\%$, $AUC = 0.9705$ (malignant–benign) have been achieved by SVM for MIAS database. The similar results of $ACC = 99.20\%$, $AUC = 0.9988$ (abnormal–normal), and $ACC = 98.00\%$, $AUC = 0.9967$ (malignant–benign) have been archived for DDSM database. The yielded results are compared with that of other classifiers namely, Naive Bayes (NB), K-Nearest Neighbor (K-NN), Back-Propagation Neural Network (BPNN), AdaBoost and Random Forest (AdaBoost-RF), LogitBoost and Random Forest (LogitBoost-RF), and Logistic Model Tree (LMT) on the similar platform. It has been observed that the results achieved by the SVM utilizing the proposed SFTA features is optimum. A rigorous comparative analysis has been made with other existing schemes with respect to accuracy (ACC) and AUC. It has been observed that the suggested scheme outperforms its competent schemes.

Chapter 4

Mammogram Classification using DOST Features followed by Null-hypothesis based Feature Selection

---

The breast cancer is currently one of the major reasons for increased death rate among women. Early detection through periodic screening improves the chance of recovery in breast cancer. For a reliable early detection, mammography is an efficient method in which digital mammograms are analyzed [7]. Digital mammograms are the scanned X-ray images of breasts. Interpretation of mammograms is a very important task for radiologists as they refer patients for biopsy. However, interpretation of mammograms varies among radiologists as it depends on training and experience. This leads to different judgments by different radiologists. It has been observed that, $60-90\%$ of initially suspected malignant lesions by radiologists were found benign later [8]. Therefore, avoidance of misinterpretation is highly desirable. Currently, Computer-Aided Diagnosis (CAD) is a very popular and efficient method which analyzes the digital mammograms and helps radiologists in mammogram interpretation to detect the suspicious lesions as well as their type. Regarding this responsibility, one important step is to extract a set of significant features from the mammographic ROIs that can classifies malignant, benign or normal mammograms. In this chapter Two-Dimensional Discrete Orthonormal S-Transform (2D-DOST) has been utilized to extract the features from the digital mammograms.

A feature selection algorithm based on null-hypothesis with statistical *two-sample t-test* method has been suggested to select most significant coefficients from a large number of DOST coefficients. The selected coefficients are used as features in the classification of mammographic images as malignant, benign or normal. Several classifiers have been employed to characterize the mammographic ROI as malignant, benign or normal. The detail description of classification by AdaBoost and Random Forest (AdaBoost-RF) has been given as it gives the significant result. The overall block diagram of the proposed scheme is shown in Figure 4.1.

The chapter is organized as follows: The extraction of feature using 2D-DOST method is described in Section 4.1. Section 4.2 outlines the feature selection followed by the classification. Section 4.3 describes the experimental results obtained on the MIAS database . Section 4.4 summarizes the overall work proposed in this chapter.

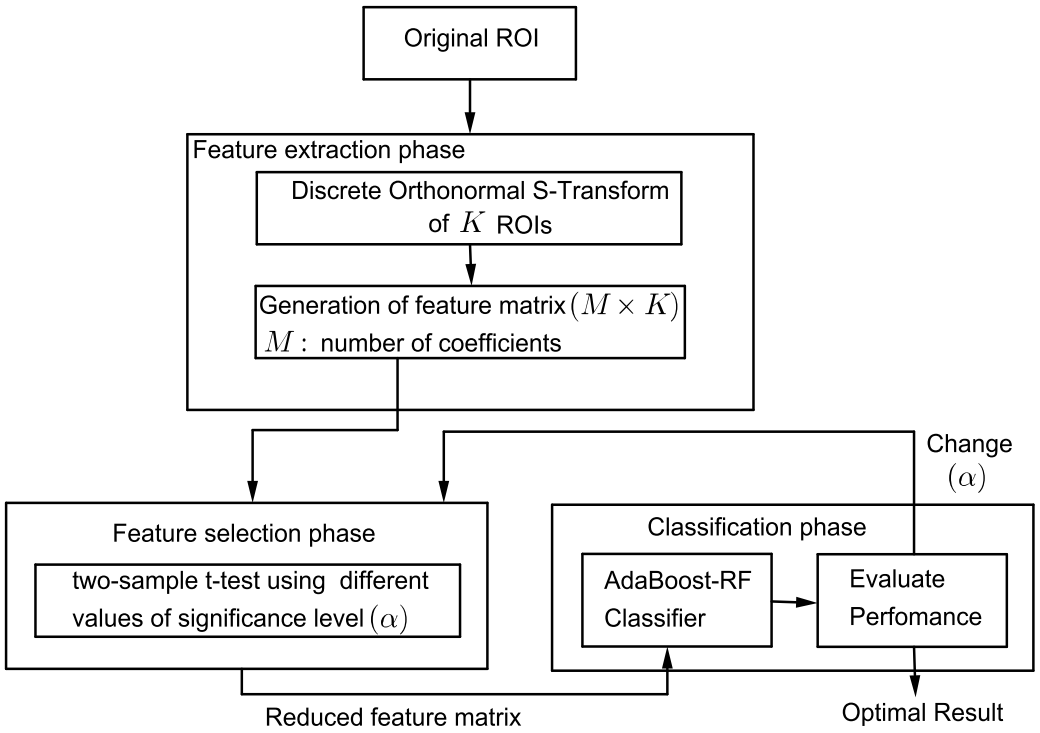


Figure 4.1: Block diagram of proposed scheme using 2D-DOST.

4.1 Extraction of Features using 2D-DOST

Wavelet Transform has been used in the extraction of features from mammograms in Chapter 2. A Two-Dimensional Discrete Wavelet Transform (2D-DWT) is a multi-resolution decomposition method in which an original image $A_{2^{j+1}}f$ at

resolution 2^{j+1} is decomposed to three detail images $D^h_{2^j} f$, $D^v_{2^j} f$, $D^d_{2^j} f$ at resolution 2^j in horizontal, vertical, and diagonal directions respectively. It also gives an approximation image $A_{2^j} f$ at coarse resolution. The detail and approximation images are the wavelet coefficient matrices in which each coefficient is considered as a feature of the original image.

The proposed scheme uses a Two-Dimensional Discrete Orthonormal S-Transform (2D-DOST) which is a multi-scale technique to extract the pixel-by-pixel texture features of a mammographic image. The 2D-DOST is based on the S-Transform, which is a time-frequency representation closely related to continuous wavelet transform [78]. The S-Transform is advantageous for the analysis of mammographic images as it preserves the phase information using linear frequency scaling. However, the major limitation of S-Transform is its high time and space complexity due to its redundant nature. To eliminate these limitations, 2D-DOST uses an orthonormal set of basis functions. Therefore, 2D-DOST has less computational and storage complexity as compared to S-Transform. The 2D-DOST of a mammographic ROI, $f(x, y)$ of size $N \times N$ can be obtained using a dyadic sampling scheme given by the following steps,

1. Perform Two-Dimensional Fourier Transform ($2D\text{-}FT$) on the image $f(x, y)$ of size $N \times N$ to obtain Fourier samples, $F(u, v) \leftarrow 2D\text{-}FT[f(x, y)]$.

2. Partition $F(u, v)$ and determine the number of points in that partition.

3. Compute the square root of the number of points and multiply it with $F(u, v)$ to get a result.

4. Apply an inverse $2D\text{-}FT$ on the result to get the DOST description of the image $f(x, y)$, which is termed as voice image and given by,

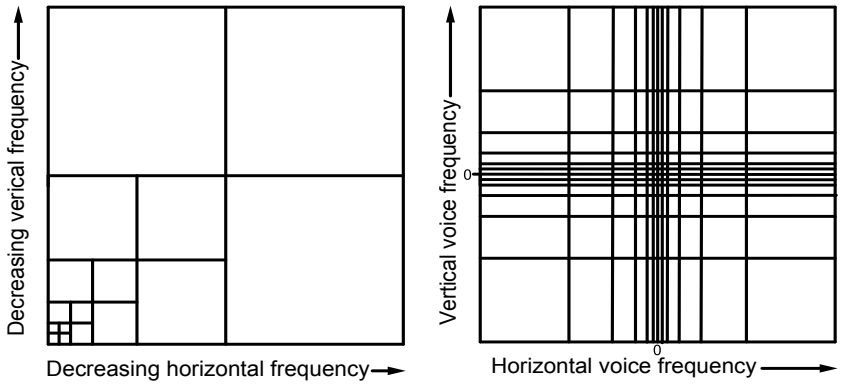
$$S\left(x', y', \upsilon_x, \upsilon_y\right) = \frac{1}{\sqrt{2^{p_x+p_y-2}}} \times \sum_{u=-2^{p_x-2}}^{2^{p_x-2}-1} \sum_{v=-2^{p_y-2}}^{2^{p_y-2}-1} F\left(u+\upsilon_x, v+\upsilon_y\right) \times e^{2\pi i\left(\frac{ux'}{2^{p_x-1}}+\frac{vy'}{2^{p_y-1}}\right)} \tag{4.1}$$

   where $\upsilon_x = 2^{p_x-1} + 2^{p_x-2}$ and $\upsilon_y = 2^{p_y-1} + 2^{p_y-2}$ are horizontal and vertical voice frequencies.

5. A rectangular voice image is obtained having $2^{p_x-1} \times 2^{p_y-1}$ points same as in the original image as shown in Figure 4.2(b).

In 2D-DWT, horizontal, vertical and diagonal detail coefficients of an image are obtained for each order as shown in Figure 4.2(a). In DOST, each pixel $p(x, y)$ within the image gives voice frequencies (υ_x, υ_y) with $2^{p_x-1} \times 2^{p_y-1}$ bandwidth. Subsequently, the pixel-wise local spatial frequency description in 2D-DOST is computed as,

1. Select an arbitrary pixel at coordinate (x, y) within the image.
2. Compute the value of the voice image S, $\forall(\upsilon_x, \upsilon_y)$ in the frequency order (p_x, p_y) of the location (x, y) at $S\,[x/N \times 2^{p_x-1}, y/N \times 2^{p_y-1}]$.
3. Build a local spatial frequency domain having size $2\log_2 N \times 2\log_2 N$ by iterating over all values (p_x, p_y) for each pixel of the image.



(a) DWT (b) DOST

Figure 4.2: A six order partition of DWT and DOST using dyadic sampling scheme.

The frequency domain contains the positive and negative components from DC, $(\upsilon_x, \upsilon_y) = (0, 0)$ to the Nyquist frequency N_f, $(\upsilon_x, \upsilon_y) = (N/2, N/2)$. Thus, all the components in the frequency domain are mapped to the M-space frequency coefficients. In this way, a $N \times N$ ROI generates $N \times N$ DOST coefficients and each coefficient is included in the feature vector (FV). Combination of K number of FVs is represented in a feature matrix FM. The detail feature extraction process is described in **Algorithm 6**. The FM becomes an input to the feature selection phase.

4.2 Selection of Features and Classification

In feature selection phase, an optimal set of relevant features are selected from the extracted feature matrix. Here, a statistical null-hypothesis test using *two-sample t-test* method [79] is used for selection of features. The null-hypothesis test is carried

Algorithm 6 Feature matrix generation using 2D-DOST.

Require: K : Total number of ROIs taken for the experiment.

Ensure: $FM[M][K]$: Feature matrix. Function $dost()$ computes DOST coefficients of ROIs and function $resize()$ sets the dimension of each ROI as per required.

Create an empty matrix $CM[N][N]$ and an empty vector FV {CM is used as DOST coefficient matrix and FV is used as feature vector}
Initialize N in terms of pixel, $i \leftarrow 1$
$M \leftarrow N \times N$ {A total number of features is to be extracted from an ROI}
for $k \leftarrow 1$ to K **do**
  Get ROI_k
  $ROI_k \leftarrow resize(ROI_k, N)$
  $CM_k[N][N] \leftarrow dost(ROI_k)$
  for $p \leftarrow 1$ to N **do**
    for $q \leftarrow 1$ to N **do**
      $FV_k[i][1] \leftarrow CM_k[p][q]$
      $i \leftarrow i + 1$
    end for
  end for
  Reset $i \leftarrow 1$
  for $m \leftarrow 1$ to M **do**
    $FM[m][k] \leftarrow FV_k[m][1]$
  end for
end for

out on two normally distributed populations of samples, say v_1 and v_2 containing benign and malignant feature data respectively. The test decision specifies whether the null-hypothesis to be correct or incorrect, which in turn triggers the data from two populations are significantly different or not. The incorrect null-hypothesis is rejected and specifies that the data from two populations are significantly different from each other and independent. Whereas, a correct null-hypothesis is failed to reject and there is no significant difference between the data from two populations.

Let, instances of two populations $b_i \in v_1, i = 1, 2, ...n_1$ and $m_j \in v_2, j = 1, 2, ...n_2$ are feature vectors. The corresponding means and standard deviations of two populations v_1 and v_2 are μ_{v_1}, μ_{v_2}, and σ_{v_1} and σ_{v_2} respectively. Now, the null-hypothesis test is performed in the following steps.

1. Specify the desired value of significance level (α) between 0 and 1. The

significance level is the probability of null-hypothesis to be incorrect.

2. Compute the statistic as,

$$t = (|\mu_{v_1} - \mu_{v_2}|) \Bigg/ \sqrt{\left(\frac{\sigma_{v_1}^2}{n_1} + \frac{\sigma_{v_2}^2}{n_2}\right)}. \tag{4.2}$$

3. Calculate the degrees of freedom as,

$$d = \left(\frac{\sigma_{v_1}^2}{n_1} + \frac{\sigma_{v_2}^2}{n_2}\right)^2 \Bigg/ \left(\frac{\sigma_{v_1}^4}{n_1^2\,(n_1-1)} + \frac{\sigma_{v_2}^4}{n_2^2\,(n_2-1)}\right). \tag{4.3}$$

4. Compute the p-value using the cumulative distributed function of *t-test* statistics as,

$$p = \int_{-\infty}^{t} \frac{\Gamma\left(\frac{d+1}{2}\right)}{\sqrt{\pi \times d} \times \Gamma\left(d/2\right)} \times \left(1 + \frac{t^2}{d}\right)^{-\left(\frac{d+1}{2}\right)} \tag{4.4}$$

where Γ is a Gamma function $(\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx)$. The p-value is the probability of the *t-test* with degrees of freedom d given that the null-hypothesis is correct.

5. Set the decision value for the null-hypothesis test as,

$$h = \begin{cases} 1, & if\ p \leq \alpha \\ 0, & if\ p > \alpha \end{cases}. \tag{4.5}$$

6. For $h = 1$, null-hypothesis is incorrect and rejected for the specified value of α.

We have taken the label values -1 and $+1$ for representing the sample as negative and positive class respectively. The *target* vector contains the label values of all ROI samples which are used in the scheme. With the help of the *target* vector, the two populations v_1 and v_2 are generated. From the null-hypothesis testing, the returned decision value is a vector and defined as, $h_m \in \{0, 1\}$, $m = 1, 2, ..., M$, where M represents the total number of extracted features. Then, a feature $f_m \in FM$ is to be selected as a relevant one, if and only if h_m equals to 1. Thus, all the selected relevant features are collected from the feature matrix FM to form a significant feature matrix SFM for K number of ROIs. The total number of reduced feature(s) denoted as R is decided according to the value of α specified in the hypothesis testing. Further, a training dataset X is created for K number of ROIs using the SFM and *target* vector,

which is used in the classifier to design an effective classifier model. **Algorithm 7** describes the feature selection process in detail.

Algorithm 7 Feature selection using statistical null-hypothesis with *t-test* method.

Require: $FM[M][K]$, $target[1][K]$, α : Significance level

Ensure: $SFM[R][K]$: Significant feature matrix. R: Total number of reduced features. Function $nhtest()$ computes statistical null-hypothesis decision value using two vectors at different values of α utilizing *two-sample t-test* method.

1: Create two empty vectors v_1 and v_2
2: Initialize α with $0 < \alpha < 1$ and $i \leftarrow 1, j \leftarrow 1, l \leftarrow 1$
3: **for** $m \leftarrow 1$ to M **do**
4: Clear contents of vector v_1 and vector v_2
5: **for** $k \leftarrow 1$ to K **do**
6: **if** $target[i] = 1$ **then**
7: $target[k] = 1$
8: $v_2[1][i] \leftarrow FM[m][k]$
9: $i \leftarrow i + 1$
10: **else**
11: $v_1[1][j] \leftarrow FM[m][k]$
12: $j \leftarrow j + 1$
13: **end if**
14: **end for**
15: Reset $i \leftarrow 1$ and $i \leftarrow 1$
16: $h[i] \leftarrow nhtest(v_1, v_2, \alpha)$
17: **if** $h[m] = 1$ **then**
18: **for** $k \leftarrow 1$ to K **do**
19: $SFM[l][k] \leftarrow FM[m][k]$
20: **end for**
21: $l \leftarrow l + 1$
22: **end if**
23: **end for**

For the validation of the proposed scheme, various classifiers namely, Naive Bayes (NB), K-Nearest Neighbor (K-NN), Back-Propagation Neural Network (BPNN), Support Vector Machine (SVM), Logistic Model Tree (LMT), AdaBoost and Random Forest (AdaBoost-RF), and LogitBoost and Random Forest (LogitBoost-RF) are used by utilizing the relevant DOST features for both MIAS and DDSM databases.

However, the detail classification experiment has been described using AdaBoost-RF method as it achieves better result. The performances of classification are evaluated with the help of parameters such as, true positive rate (TPR), false positive rate (FPR), F_{score}, Matthews correlation coefficient (MCC), and AUC value in the analysis of ROC curves.

4.3 Experimental Results and Analysis

In this work all the mammographic ROIs of MIAS and DDSM databases have been cropped of size 128×128 pixels which are used in the feature extraction experiment. In feature extraction phase (**Algorithm 6**), DOST coefficients of size 128×128 are extracted from each ROI and a feature matrix is built by keeping all corresponding coefficient of each ROI in rows and ROI indices in columns. Next, the significant features are selected using the feature selection algorithm (**Algorithm 7**). The significant feature matrices (SFMs) are generated by using different values of significance level (α). Using these SFMs and the class vector ($target$), a number of datasets are generated and used in the classifier.

We have employed a 10-fold cross-validation technique for each experiment for a number of rounds. In 10-fold cross-validation experiment, the whole dataset is partitioned into 10 number of folds. In each round, nine folds are combined to form one set and remaining one fold is made as another set. Thus, two disjoint sets are formed containing 10% and 90% data that are used separately for training and validation process respectively. This process is repeated for 10 times with random selection of training and testing data by the classifier. For classification, we have taken Random Forests with $10, 20, 40, 80$, and 100 trees with maximum depth of two which is used as the base learner in the AdaBoost algorithm. It has been observed that, the best performance is achieved using a Random Forest with 20 trees. Thus, with optimal structure of the classifier, a number of datasets having various sizes are used for the classification of mammograms.

The various values of classification accuracies (ACC) obtained using different values of α are articulated in Table 4.1. It may be noticed that, optimum accuracy values are achieved with p-value less than $\alpha = 7 \times 10^{-4}$ for both MIAS and DDSM database. The classification performances of the Adaboost-RF classifier with that of other classifiers have been compared and given in Table 4.2.

Table 4.1: Comparative analysis of classification accuracies at different values of α.

Mammogram Database	Significance level (α)	*ACC* (%)	
		Abnormal–normal	Malignant–benign
MIAS	6×10^{-3}	90.99	88.7
	5×10^{-3}	93.47	94.8
	4×10^{-3}	94.09	92.2
	3×10^{-3}	95.34	94.8
	2×10^{-3}	96.89	95.6
	1×10^{-3}	97.51	96.5
	8×10^{-4}	97.51	97.4
	$\mathbf{7 \times 10^{-4}}$	**98.75**	**98.3**
	6×10^{-4}	96.89	95.7
	5×10^{-4}	77.95	68.7
DDSM	6×10^{-3}	92.1	90.4
	5×10^{-3}	94.1	91.6
	4×10^{-3}	95.1	92.8
	3×10^{-3}	96.8	94.4
	2×10^{-3}	97.1	94.0
	1×10^{-3}	97.8	96.8
	8×10^{-4}	98.6	97.6
	$\mathbf{7 \times 10^{-4}}$	**99.3**	**98.8**
	6×10^{-4}	94.7	92.0
	5×10^{-4}	92.6	88.8

It has been observed that, the AdaBoost-RF classifier performs better than other classifiers. The optimum performances are $ACC = 98.75\%$ (abnormal–normal) and $ACC = 98.26\%$ (malignant–benign) for MIAS database. Similarly for DDSM database, the parameters are $ACC = 99.30\%$ (abnormal–normal) and $ACC = 98.80\%$ (malignant–benign). The other parameters such as F_{score}, MCC and AUC are also maximum at that optimal $\alpha = 7 \times 10^{-4}$. At this value of significance level, the root relative square errors are 0.1509 (abnormal–normal) and 0.2390 (malignant–benign) for MIAS database. Similarly, for DDSM database, error values are 0.0895 (abnormal–normal) and 0.1194 (malignant–benign). It has been observed that these error values are minimum than that of at other values of α.

Table 4.2: Comparison of performances of various classifiers using optimal relevant feature set (at $\alpha = 7 \times 10^{-4}$).

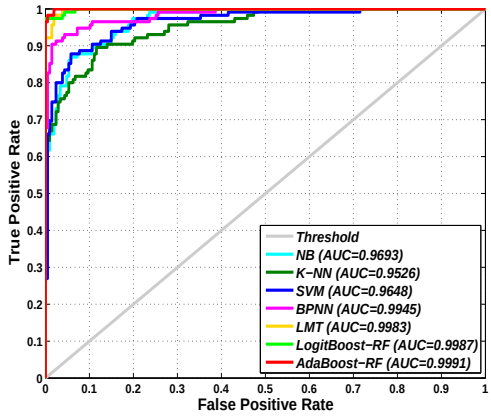
Database	Class set	Classifier	Measures of performance			
			F_{score}	MCC	AUC	ACC (%)
MIAS	Abnormal–normal	NB	0.875	0.809	0.9693	91.30
		K-NN	0.838	0.754	0.9562	88.81
		BPNN	0.922	0.885	0.9945	97.72
		SVM	0.881	0.816	0.9648	91.61
		AdaBoost-RF	**0.982**	**0.973**	**0.9991**	**98.75**
		LogitBoost-RF	0.973	0.960	0.9987	98.13
		LMT	0.946	0.919	0.9983	96.27
	Maligant–benign	NB	0.870	0.797	0.9050	89.56
		K-NN	0.771	0.696	0.8030	83.47
		BPNN	0.917	0.862	0.9920	93.04
		SVM	0.603	0.545	0.6823	74.78
		Adaboost-RF	**0.980**	**0.965**	**0.9985**	**98.26**
		LogitBoost-RF	0.970	0.947	0.9948	97.39
		LMT	0.950	0.912	0.9930	95.65
DDSM	Abnormal–normal	NB	0.942	0.884	0.9772	94.20
		K-NN	0.921	0.842	0.9494	92.10
		BPNN	0.971	0.942	0.9921	97.10
		SVM	0.951	0.902	0.9826	95.10
		AdaBoost-RF	**0.993**	**0.986**	**0.9994**	**99.30**
		LogitBoost-RF	0.986	0.972	0.9990	98.60
		LMT	0.978	0.956	0.9968	97.80
	Malignnt–benign	NB	0.910	0.809	0.9718	90.40
		K-NN	0.883	0.804	0.8953	89.20
		BPNN	0.948	0.890	0.9919	94.40
		SVM	0.684	0.586	0.7597	75.20
		AdaBoost-RF	**0.988**	**0.976**	**0.9992**	**98.80**
		LogitBoost-RF	0.977	0.952	0.9987	97.60
		LMT	0.960	0.923	0.9612	96.00

The fold-wise results in terms of confusion matrix for optimal dataset (at $\alpha = 7\times10^{-4}$) in 10-fold cross-validation experiment is also presented in Table 4.3.

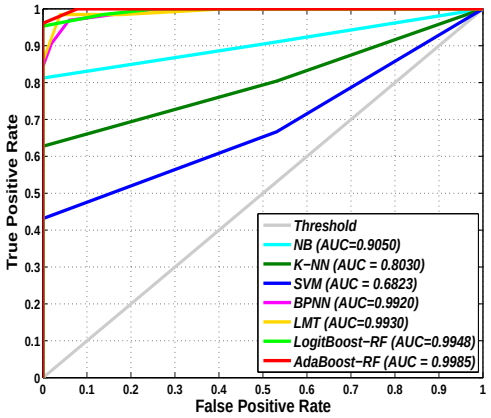
Table 4.3: Optimal confusion matrices for both MIAS and DDSM databases (fold-wise) at $\alpha = 7 \times 10^{-4}$.

Database	Folds	Class set											
		Normal-abnormal						Benign-malignant					
		Training instances	Testing instances	TP	FP	TN	FN	Training instances	Testing instances	TP	FP	TN	FN
MIAS	Fold 1	289	33	11	0	21	1	103	12	6	0	6	0
	Fold 2	289	33	12	0	21	0	103	12	4	0	7	1
	Fold 3	289	33	12	0	21	0	103	12	5	0	7	0
	Fold 4	289	33	11	0	21	1	103	12	5	0	7	0
	Fold 5	291	31	11	0	20	0	103	12	6	0	6	0
	Fold 6	290	32	11	0	21	0	104	11	4	0	6	1
	Fold 7	290	32	11	1	20	0	104	11	5	0	6	0
	Fold 8	291	31	11	0	20	0	104	11	5	0	6	0
	Fold 9	290	32	11	0	21	0	104	11	5	0	6	0
	Fold 10	290	32	11	0	20	1	104	11	4	0	7	0
DDSM	Fold 1	900	100	50	0	48	2	450	50	26	0	24	0
	Fold 2	900	100	50	0	50	0	450	50	26	0	24	0
	Fold 3	900	100	50	0	50	0	450	50	25	0	24	1
	Fold 4	900	100	50	0	49	1	450	50	26	0	24	0
	Fold 5	900	100	50	0	50	0	450	50	26	1	23	0
	Fold 6	900	100	50	0	49	1	450	50	26	0	23	1
	Fold 7	900	100	50	0	50	0	450	50	26	1	23	0
	Fold 8	900	100	50	0	48	2	450	50	26	0	23	1
	Fold 9	900	100	50	0	50	0	450	50	27	0	23	0
	Fold 10	900	100	50	0	49	1	450	50	26	0	23	1

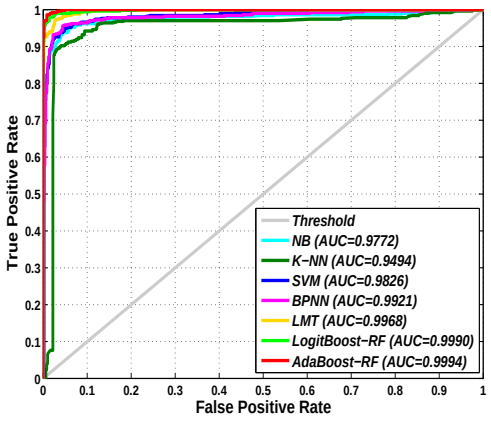
The *AUC* values of ROC curves achieved by the AdaBoost-RF classifier with that of other classifiers have also been compared and presented in the Figure 4.3. The optimal *AUC* values are 0.9991 (abnormal–normal) and 0.9985 (malignant–benign) obtained by the AdaBoost-RF classifier for MIAS database. For DDSM database, similar values are 0.9994 (abnormal–normal) and 0.9992 (malignant–benign).



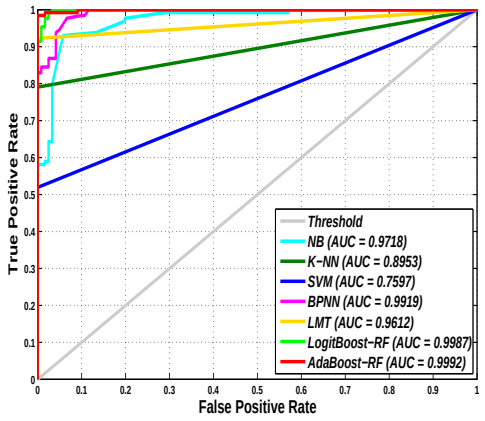
(a) Abnormal-normal (MIAS)



(b) Malignant-benign (MIAS)



(c) Abnormal-normal (DDSM)



(d) Malignant-benign (DDSM)

Figure 4.3: ROC curves obtained by different classifiers using relevant features at optimum α of 7×10^{-4}.

Finally, Table 4.4 presents the comparative analysis of various performance measures of the present scheme with the existing approaches. It may be observed that the suggested scheme outperforms its competent ones.

Table 4.4: Performance comparison by different approaches with the proposed scheme for classification of mammograms.

Approach	**Technique**	**Database**	**Measurement**
Verma *et al.* (2010) [28]	BI-RADS descriptor, SCBDL classifier	DDSM	$ACC = 97.5\%$ (Malignant–benign)
Buciu *et al.* (2011) [31]	Gabor wavelets, PCA and SVM	MIAS	$AUC = 0.78$ (Malignant–benign)
Görgel *et al.* (2013) [42]	SWT, SVM	I.U. database	$ACC = 96.0\%$ (Abnormal–normal) $ACC = 93.59\%$ (Malignant–benign)
Nascimento *et al.* (2013) [43]	DWT, Polynomial classifier	DDSM	$AUC = 0.96$ (Malignant–benign)
Xiaoming *et al.* (2014) [46]	Geometry and texture features, SVM-RFE with NMIES filter	DDSM	$AUC = 0.9615$ (Malignant–benign)
Ganesan *et al.* (2014) [47]	Trace transform, GMM	SATA	$ACC = 92.48\%$ (Malignant–benign)
Proposed Scheme	**DOST +Null-hypothesis +AdaBoost-RF**	**MIAS**	$\boldsymbol{ACC = 98.75\%}$ $\boldsymbol{AUC = 0.9991}$ **(Abnormal–normal)** $\boldsymbol{ACC = 98.26\%}$ $\boldsymbol{AUC = 0.9985}$ **(Malignant–benign)**
		DDSM	$\boldsymbol{ACC = 99.30\%}$ $\boldsymbol{AUC = 0.9994}$ **(Abnormal–normal)** $\boldsymbol{ACC = 98.80\%}$ $\boldsymbol{AUC = 0.9992}$ **(Malignant–benign)**

4.4 Summary

In this chapter an efficient scheme has been proposed to classify mammographic images as malignant, benign or normal to support the early detection of breast cancer. The scheme utilizes DOST method to extract features from the mammographic images. Null-hypothesis with *two-sample t-test* has been proposed to select the most discriminant features from high dimensional feature matrix. Several classifiers namely, Naive Bayes (NB), K-Nearest Neighbor (K-NN), Back-Propagation Neural Network (BPNN), Support Vector Machine (SVM), Logistic Model Tree (LMT), AdaBoost and Random Forest (AdaBoost-RF), and LogitBoost and Random Forest (LogitBoost-RF) have been used to classify the mammograms utilizing relevant feature set. The classification algorithm with selected relevant features achieves the best performance at significance level, $\alpha = 7 \times 10^{-4}$. The results achieved by AdaBoost-RF with respect to accuracy (ACC) and AUC are optimal in comparison to other classifiers. The parameters are $ACC = 98.75\%$, $AUC = 0.9991$ (abnormal–normal), and $ACC = 98.26\%$ and $AUC = 0.9985$ (malignant–benign) for MIAS database. Similarly, for DDSM database the parameters are $ACC = 99.30\%$, $AUC = 0.9994$ (abnormal–normal), and $ACC = 98.80\%$, $AUC = 0.9992$ (malignant–benign). A comparative analysis has been made with other existing schemes with respect to accuracy (ACC) and AUC. It has been observed that the suggested scheme outperforms its competent schemes.

Chapter 5

Mammogram Classification using Slantlet Features followed by BLogR for Feature Selection

---

Breast cancer continues to be a significant public health problem in the world. It is viewed as one of the most frequent mortality causes among women. Early detection is the key for enhancing breast cancer anticipation. Mammography is at present the best method for reliable and early detection of breast cancer. On the other hand, it is difficult for radiologists to provide both exact and uniform assessment for a large number of mammograms generated in widespread screening. Computer-Aided Diagnosis (CAD) of digital mammograms replaces conventional screening of breast cancer. The CAD framework enhances diagnostic accuracy as well as the reproducibility of mammographic interpretation. In this chapter, an efficient scheme is proposed to characterize the type of digital mammogram as malignant, benign or normal. A Contrast Limited Adaptive Histogram Equalization (CLAHE) technique is utilized to enhance the mammographic ROI that contains the suspicious region of the breast. A Two-Dimensional Slantlet Transform (2D-SLT) has been employed to extract the texture features from the mammographic images. Bayesian Logistic Regression (BLogR) method has been utilized for the selection of most discriminatory feature element that represents the pattern of mammogram class and minimizes the effort of the classification along with accuracy improvement. However, in most of the cases, the formed relevant feature dataset lacks balance in the number of instances to each class. This lacking of balance degrades the performance of the

classifier due to over-fitting. To accomplish an adequate improved performance, the relevant features are balanced by the Gaussian distribution based balancing method prior to classification. The classification phase uses several classifiers to map the processed relevant feature vector to a class of the mammogram. It has been observed that LogitBoost and Random Forest (LogitBoost-RF) classifier achieves the optimal performance among all classifiers. Therefore the detail explanation of the classification performed by LogitBoost-RF classifier has been given in this chapter. The overall block diagram of the proposed scheme has been shown in Figure 5.1.

The chapter is organized as follows: The enhancement of the mammographic ROI is described in Section 5.1. The overviews of 2D-SLT and BLogR methods are explained in Sections 5.2 and 5.3 respectively. Section 5.4 outlines feature extraction and selection procedure. The balancing of relevant features and subsequent classification are explained in Sections 5.5 and 5.6 respectively. Section 5.7 describes the results obtained on standard databases. Section 5.8 summarizes the overall work proposed in this chapter.

5.1 Enhancement of ROIs

The tissues present in the digital mammographic ROI possess very little contrast. Hence, the ROI image is very poor quality and needs enhancement prior to feature extraction. In this work, a CLAHE technique is applied to enhance the ROIs. The CLAHE technique computes the histogram of intensities in a contextual region centered at each pixel. Then, it sets a value for the intensity of the pixel according to the rank of that pixel in the local histogram within the display range [80]. It is a refinement of adaptive histogram equalization (AHE) where, the ordinary histogram is modified to induce the enhancement by imposing a user-specified maximum intensity level. During the enhancement with CLAHE, the original ROIs are partitioned into many non-overlapping contextual square regions of equal sizes. For each block of the image, the histogram is computed and equalized. To equalize the histogram, the given gray-scale function is converted to a uniform density function by estimating the cumulative distributed function (CDF) [81].

Consider P and G be the number of pixels and gray-scales respectively in each block. Let $h_i(n)$ is the histogram of i-th block for $n = \{0, 1, 2, ..., G-1\}$. The CDF scaled by $(G-1)$ for a gray-scale mapping is defined as,

$$F_i(n) = \frac{(G-1)}{P} \sum_{k=0}^{n} h_i(k) \tag{5.1}$$

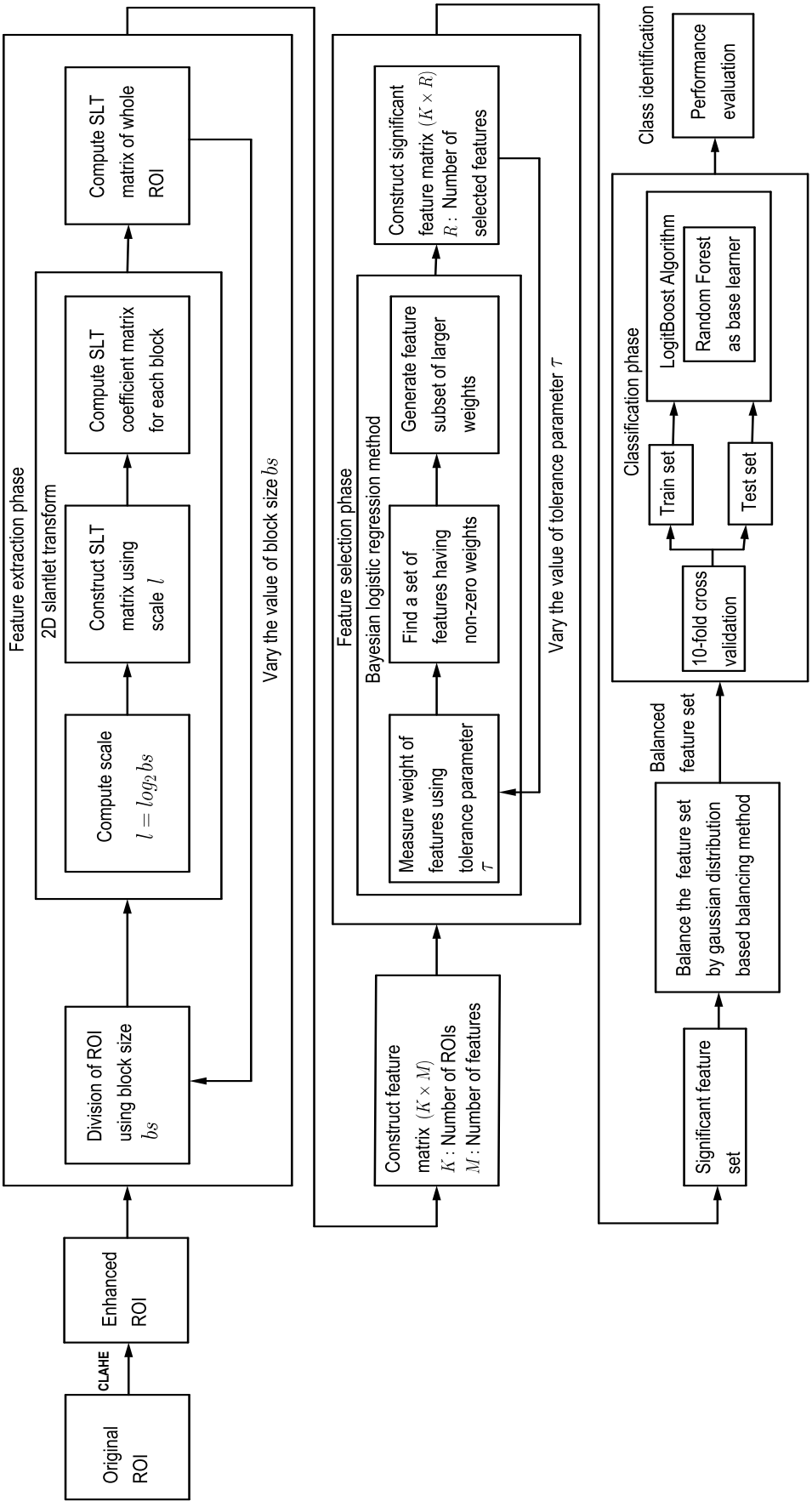
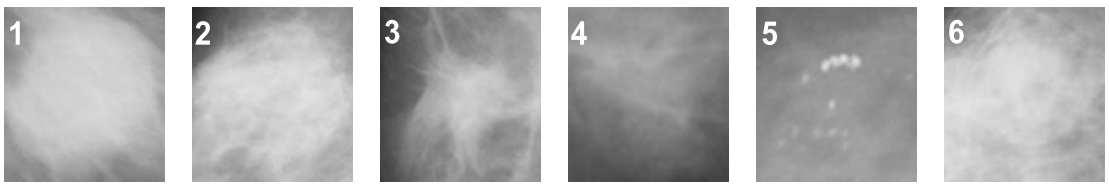


Figure 5.1: Block diagram of proposed scheme using 2D-SLT and BLogR method.

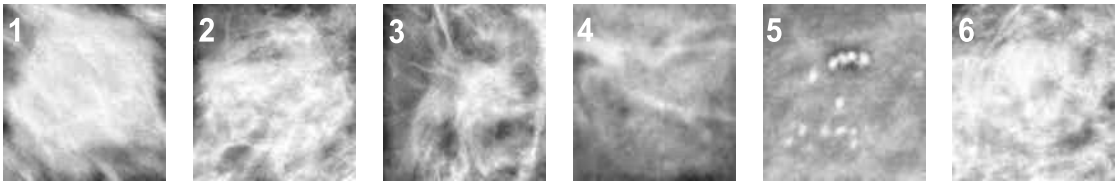
The histogram equalization has a limitation that the block contrast is increased to its maximum. Hence, the contrast of a block is set to a desired level by limiting the maximum slope $S_{\max}$ of the CDF in (5.1). For this purpose, a threshold parameter called as clip limit is set for clipping the histograms. Now, consider the clip limit is c with a clip factor, a (in %). Then, c is given as

$$c = \frac{P}{G}\left(1 + \frac{a}{100}\left(S_{\max} - 1\right)\right) \tag{5.2}$$

In every mapping the value of maximum slope ranges from 1 to $S_{\max}$ as the clipping factor (a) changes between 0 and 100. For X-ray images, $S_{\max}$ is generally set to 4. Each histogram of all blocks is redistributed in such a way that its height does not go beyond the clip limit. Thus, all the histograms are modified by limiting the maximum number of counts for each gray-scale to a clip limit c. Next, a bilinear interpolation method is utilized to combine all the neighborhood blocks for removing the boundaries, which are induced artificially. Now, the gray-scale values of the mammogram are altered according to the modified histogram. The size of a contextual block is taken as 4×4 to partition the mammogram image for enhancement. The total number of blocks optimally used in the experiment depends on the type of input mammogram. The clip limit that specifies the contrast enhancement limit is taken as 0.01, which gives the best results in the present case. For the histogram, the number of bins used in building a contrast enhancing transformation is limited to 32. The uniform distribution is used for the flat histogram of mammogram blocks, which results in an optimal output. The enhanced mammographic ROIs of the original ROIs are shown in Figure 5.2.



(a) Original mammographic ROIs

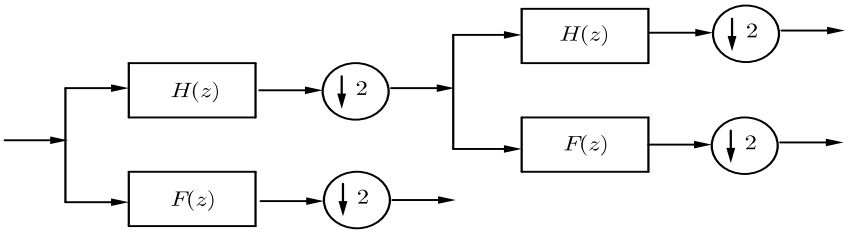


(b) Corresponding enhanced ROIs

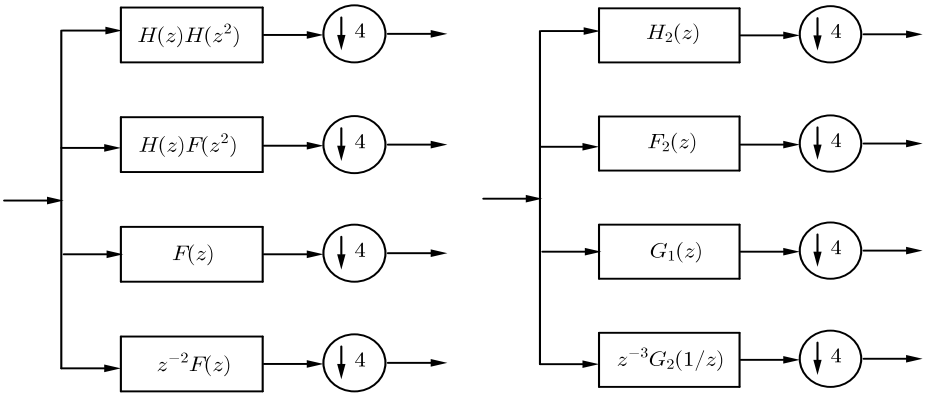
Figure 5.2: Enhancement of mammographic ROIs using CLAHE technique.

5.2 Two-Dimensional Slantlet Transform

The Two-Dimensional Slantlet Transform (2D-SLT) has been proposed by Selesnick utilizing the lengths of the discrete time basis function and their moments in such a way, to the point that both time localization and smoothness properties are achieved [82]. It is similar to orthogonal Discrete Wavelet Transform (DWT) with an enhanced time localization and two zero moments. The architecture of the SLT is based on parallel filter bank structure where distinctive filters are used for every scale rather than the iteration of filters for every level (Figure 5.3).



(a) Two-scale filter bank



(b) Equivalent form of filter bank (c) Two-scale filter SLT bank structure

Figure 5.3: Two-scale filter bank with its equivalent form and corresponding SLT filter bank structure.

The SLT filter banks are made out of filters $g_i(n)$, $f_i(n)$ and $h_i(n)$ with a scale i. The length of filters for scale i will be $2i$. The number of channels in a l-scale SLT filter bank is $2l$. The filter $h_i(n)$ is a low pass filter and $f_l(n)$ is the adjacent filter to it. Both filters $h_l(n)$ and $f_l(n)$ follow a down sampling by 2^l. The remaining $2l-2$ channels are channels are filtered by $g_i(n)$ and its shifted time-reverse for $i = 1, ..., l-1$ and followed by down sampling by 2^{i+1}. The filters $g_i(n)$, $f_i(n)$ and $h_i(n)$ are linear over the two intervals, $n \in \{0, ..., 2^i - 1\}$ and $n \in \{2^i, ..., 2^i - 1\}$. The filter $g_i(n)$ is

described by four parameters, $a_{0,0}$, $a_{0,1}$, $a_{1,0}$ and $a_{1,1}$, and expressed as,

$$g_i(n) = \begin{cases} a_{0,0} + a_{0,1}n, & for\, n = 0, ..., 2^i - 1 \\ a_{1,0} + a_{1,1}(n - 2^i), & for\, n = 2^i, ..., 2^{i+1} - 1 \end{cases} \tag{5.3}$$

The parameters $a_{0,0}$, $a_{0,1}$, $a_{1,0}$ and $a_{1,1}$ are computed as,

$$\begin{aligned} a_{0,0} &= \tfrac{(s_0+t_0)}{2}, \quad a_{1,0} = \tfrac{(s_0-t_0)}{2}, \\ a_{0,1} &= \tfrac{(s_1+t_1)}{2}, \quad a_{1,1} = \tfrac{(s_1-t_1)}{2} \end{aligned} \tag{5.4}$$

where $s_0 = -s_1\left(\frac{m-1}{2}\right)$, $s_1 = 6\sqrt{\frac{m}{(m^2-1)(4m^2-1)}}$, $t_0 = \left(\frac{s_1(m+1)}{3-mt_1}\right)\left(\frac{m-1}{2m}\right)$, $t_1 = 2\sqrt{\frac{3}{m(m^2-1)}}$ and $m = 2^i$.

The filters $f_i(n)$ and $h_i(n)$ are defined in terms of eight parameters $b_{0,0}$, $b_{0,1}$, $b_{1,0}$, $b_{1,1}$, $c_{0,0}$, $c_{0,0}$, $c_{0,1}$, $c_{1,0}$ and $c_{1,1}$, and expressed as,

$$h_i(n) = \begin{cases} b_{0,0} + b_{0,1}n, & for\, n = 0, ..., 2^i - 1 \\ b_{1,0} + b_{1,1}(n - 2^i), & for\, n = 2^i, ..., 2^{i+1} - 1 \end{cases} \tag{5.5}$$

$$f_i(n) = \begin{cases} c_{0,0} + c_{0,1}n, & for\, n = 0, ..., 2^i - 1 \\ c_{1,0} + c_{1,1}(n - 2^i), & for\, n = 2^i, ..., 2^{i+1} - 1 \end{cases} \tag{5.6}$$

The parameters are computed as,

$$\begin{aligned} b_{0,0} &= u\tfrac{v+1}{2m}, \quad b_{1,0} = u - b_{0,0} \\ b_{0,1} &= \tfrac{u}{m} \quad b_{1,1} = -b_{0,1} \\ c_{0,1} &= q(v - m), \; c_{1,1} = -q(v + m) \\ c_{1,0} &= c_{1,1}\tfrac{v+1-2m}{2} \; c_{0,0} = c_{0,1}\tfrac{v+1}{2} \end{aligned} \tag{5.7}$$

where $m = 2^i$, $u = \frac{1}{\sqrt{m}}$, $v = \sqrt{\frac{(2m^2+1)}{3}}$ and $q = \sqrt{\frac{3}{m(m^2-1)/m}}$.

The dimension of the orthogonal matrix is 2^l generated from a l-scale filter banks. In the matrix, the first and second rows correspond to the filters $h_l(n)$ and $f_l(n)$ respectively. Each of the remaining rows will be generated by the succession of $g_i(n)$, its time reverse and their shift by 2^{i+1} for $i = 1, ..., l-1$. The SLT matrix S computed from the two-scale SLT filter banks using 2D signal will have dimension 4 ($l = 2$, $2^l = 4$). Similarly for the three-scale SLT filter banks, the dimension of the S is 8. The SLT matrices of size (4×4) and (8×8) are shown in Figure 5.4.

0.5	0.5	0.5	0.5
0.6708	0.2236	-0.2236	-0.6708
-0.5117	0.8279	-0.1208	-0.1954
-0.1954	-0.1208	0.8279	-0.5117

(a) $S_{4\times4}$

0.3536	0.3536	0.3536	0.3536	0.3536	0.3536	0.3536	0.3536
0.5401	0.3858	0.2315	0.0772	-0.0772	-0.2315	-0.3858	-0.5401
-0.5062	-0.0874	0.3314	0.7502	-0.0793	-0.1078	-0.1362	-0.1646
-0.1646	-0.1362	-0.1078	-0.0793	0.7502	0.3314	-0.0874	-0.5062
-0.5117	0.8279	-0.1208	-0.1954	0	0	0	0
-0.1954	-0.1208	0.8279	-0.5117	0	0	0	0
0	0	0	0	-0.5117	0.8279	-0.1208	-0.1954
0	0	0	0	-0.1954	-0.1208	0.8279	-0.5117

(b) $S_{8\times8}$

Figure 5.4: Values of Slantlet matrices with dimensions 4 and 8.

5.3 Bayesian Logistic Regression Method

The Bayesian Logistic Regression (BLogR) method has been proposed by Cawley *et al.* [83] in the year 2006 is an improvement of the sparse logistic regression approach of Shevade *et al.* [84]. This algorithm is a parameterless technique in which the regularization parameter is integrated out analytically to avoid its optimization. The modified Bayesian Logistic Regression method outperforms the classical sparse logistic regression algorithm in distinguishing a subset of the most significant features from the larger set of extracted SLT features of mammographic ROIs.

Let a set $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$ of n samples in which $x_i \in \chi \subset \mathbb{R}^d$ denotes the i-th sample that is associated with a binary class label $\mathcal{C}_i \rightarrow y_i \in \{-1, +1\}$, which represents positive class ($\mathcal{C}_1$) and negative class ($\mathcal{C}_2$). Now, the classical logistic regression approach which estimates a posteriori probability of class membership based on the linear combination of input features and is given by,

$$prob\left(\mathcal{C}_i \left| x_i \right.\right) = \frac{1}{1 + \exp - \left\{ f\left(x_i\right) \right\}} \tag{5.8}$$

where

$$f\left(x_i\right) = \sum_{j=1}^{d} \alpha_j x_{ij} + \alpha_0. \tag{5.9}$$

The parameter α_0 and α_j is found by minimizing the negative log-likelihood. Considering $\mathcal{D}$ to be independent identically distributed (i.i.d) sample from a Bernoulli distribution, then the negative log likelihood is given by,

$$E_{\mathcal{D}} = \sum_{i=1}^{n} g\left\{-y_i f\left(x_i\right)\right\} \tag{5.10}$$

where $g\left\{\xi\right\} = \log\left\{1 + \exp\left(\xi\right)\right\}$. The first and second derivatives $(\frac{\partial E_{\mathcal{D}}}{\partial \alpha_j}, \frac{\partial^2 E_{\mathcal{D}}}{\partial \alpha_j^2})$ with respect to individual model parameters are continuous. A perfect model would be favored which selects smaller number of most significant features. Regarding this context a standard regularization strategy is added to negative log-likelihood corresponding to a Laplace prior over α. This yields a modified training criterion,

$$M = E_{\mathcal{D}} + \lambda E_{\alpha} \tag{5.11}$$

where $E_{\alpha} = \sum_{i=1}^{d} |\alpha_i|$ and λ is a regularization parameter which controls the bias-variance trade-off. At minima of M, the partial derivatives of M with respect to the model parameter will be uniformly zero and gives

$$\left|\frac{\partial E_{\mathcal{D}}}{\partial \alpha_j}\right| = \lambda \;\; if\, |\alpha_i| > 0 \;\; and \;\; \left|\frac{\partial^2 E_{\mathcal{D}}}{\partial \alpha_j^2}\right| < \lambda \;\; if\, |\alpha_i| = 0 \tag{5.12}$$

This infers that the sensitivity of the negative log-likelihood with respect to a model parameter α_i less than λ, then the value of that parameter is set to zero and the corresponding input feature will be clipped from the model.

The key weakness of this methodology is that no optimization problem with continuous derivatives is included. The optimization problem determines a appropriate value for the regularization parameter λ. This weakness can be excreted by the Bayesian regularization where the λ is integrated out analytically. At each iteration, a model parameter with the gradient of largest magnitude is chosen for the optimization. The active parameters (with non-zero values) are considered for the optimization to enhance the speed of convergence. If no active parameters are available, then only inactive parameters are considered for optimization. For the most part, the value of gradient is not lessened exactly to zero. Thus, just the parameters are considered for optimization if they have the gradient surpassing a predefined tolerance parameter (τ). When no such parameter (τ) is found, then the algorithm ends. The decreasing value of tolerance parameter increases the quality of approximation to the correct posterior.

The posterior distribution for α, the parameters of the model given in (5.8) and (5.9), can be expressed in a Bayesian regularization of the minimization of equation (5.11) as

$$prob\,(\alpha\,|\mathcal{D},\lambda\,) \propto prob\,(\mathcal{D}\,|\alpha\,)\,prob\,(\alpha\,|\lambda\,)$$

Then, the prior over model parameters, α is given by a separable Laplace distributon and written as

$$prob\,(\alpha\,|\lambda\,) = \left(\frac{\lambda}{2}\right)^N \exp\left\{-\lambda E_\alpha\right\} = \prod_{i=1}^{N} \frac{\lambda}{2} \exp\left\{-\lambda\,|\alpha_i|\right\} \tag{5.13}$$

where N is the number of active parameters. In a Bayesian regularization, the suitable value of λ is estimated by integrated out of it analytically. Here, the prior distribution over model parameter is marginalized over λ and given by

$$prob\,(\alpha) = \int prob\,(\alpha\,|\lambda)prob\,(\lambda)\,d\lambda \tag{5.14}$$

Now by replacing (5.13) in (5.14), it is reduced to

$$prob\,(\alpha) = \frac{1}{2^N} \int_0^\infty \lambda^{N-1} \exp\left\{-\lambda E_\alpha\right\} d\lambda \tag{5.15}$$

Here λ is strictly positive. Further, a Gamma integral, $\int_0^\infty x^{\nu-1}e^{-\mu x}dx = \frac{\Gamma(\nu)}{\mu^\nu}$ is used to obtain

$$prob\,(\alpha) = \frac{1}{2^N} \frac{\Gamma\,(N)}{E_\alpha^N} \Rightarrow -\log prob\,(\alpha) \propto N \log E_\alpha \tag{5.16}$$

Thus, an optimization criteria for sparse logistic regression with Bayesian regularization is represented as,

$$Q = E_\mathcal{D} + N \log E_\alpha \tag{5.17}$$

where λ has been eliminated. Now, differentiating the original and modified training criteria ((5.11) and (5.17)) we get,

$$\Delta M = \Delta E_\mathcal{D} + \lambda \Delta E_\alpha, \tag{5.18}$$

$$\Delta Q = \Delta E_\mathcal{D} + \widetilde{\lambda}\,\Delta E_\alpha. \tag{5.19}$$

where

$$\frac{1}{\widetilde{\lambda}} = \frac{1}{N} \sum_{i=1}^{N} |\alpha_i| \tag{5.20}$$

From a gradient descent perspective, minimizing Q becomes equivalent to minimizing M where the λ is continuously updated according to (5.20) for every change in the vector of model parameter α. This requires only a very minor modification of the code implementing the sparse logistic regression algorithm, whilst eliminating the only training parameter and hence the need for a model selection procedure in fitting the model.

5.4 Feature Extraction and Selection

Let ROI be an $N \times N$ matrix of the picture element (pixels) intensity values of a mammographic ROI. The ROI is divided into many square regions called as blocks using the predefined block size bs. Every block in the division of ROI is known as cropped image (CI) that is utilized to produce a coefficient matrix with the help of Two-Dimensional Slantlet Transform applied on that image block. A SLT matrix S having dimension $P \times P$ is computed from the l-scale SLT filter banks. The scale l is specified by $l = \log_2 bs$, where bs is the predefined block size that is used for ROI division. A Two-Dimensional Slantlet Transform is performed by the sequential row and column transformations on each block, producing the block coefficient matrix as,

$$CICF = S \times CI \times S^T \tag{5.21}$$

where S^T is the transposed SLT matrix. Subsequently, each coefficient matrix of every individual block is concatenated to yield the coefficient matrix (CF) of the whole ROI. Further, all the generated coefficient matrices of the K number of ROIs are used to build the feature matrix (FM) that has been described in the **Algorithm 8**.

A Bayesian Logistic Regression (BLogR) method has been employed to select the most significant features from the obtained feature matrix, FM. The feature selection algorithm utilizes a tolerance parameter (τ) and the target class vector ($target$) that contains the labels of the instances, for generation of relevant feature set. The algorithm calculates the weight of each feature value present in the feature matrix and finds a list of relevant features of non-zero weights. The list of significant features is arranged in a descending order of weights. The value of tolerance parameter (τ) is to be chosen in the range of 0 to a maximum value after which no more features will be selected by the selection algorithm. That value of the tolerance parameter is considered as the stopping criteria for the execution of the algorithm. The detail procedure of the generation of significant feature matrix (SFM) has been explained in the **Algorithms 9** and **10**.

Algorithm 8 Generation of feature matrix using 2D-SLT.

Require: K number of ROIs, block size (bs) and scale (l).

Ensure: $FM[K][M]$: Feature matrix. Function $resize()$ sets the dimension of each ROI and $sltmatrix()$ computes the SLT matrix (S).

Initialize N and bs, $row \leftarrow 0$, $col \leftarrow 0$, $c \leftarrow 1$
Create and set a matrix $CF[N][N] \leftarrow \phi$ and a vector V
$M \leftarrow N \times N$, $P \leftarrow bs$ { } M: Total number features
for $k \leftarrow 1$ to k **do**
  Read ROI_k
  $ROI_k \leftarrow resize(ROI_k, N)$
  Compute $l \leftarrow \log_2 bs$ and $S[P][P] \leftarrow sltmatrix(l)$
  for $ii \leftarrow 1$ to $N - bs + 1$ step bs **do**
    for $jj \leftarrow 1$ to $N - bs + 1$ step bs **do**
      for $i \leftarrow ii$ to $ii + bs - 1$ **do**
        $row \leftarrow row + 1$
        for $j \leftarrow jj$ to $jj + bs - 1$ **do**
          $col \leftarrow col + 1$
          $CI[row][col] \leftarrow A[i][j]$; $CICF[P][P] \leftarrow S \times CI \times S^T$
          $CF_k \leftarrow CF_k \cup CICF$
        end for
        $col \leftarrow 0$
      end for
      $row \leftarrow 0$
    end for
  end for
  for $p \leftarrow 1$ to N **do**
    for $q \leftarrow 1$ to N **do**
      $V_k[c][1] \leftarrow CF_k[p][q]$
      $c \leftarrow c + 1$
    end for
  end for
  Reset $c \leftarrow 1$
  for $m \leftarrow 1$ to M **do**
    $FM[m][k] \leftarrow V_k[m][1]$
  end for
end for
$FM[K][M] \leftarrow (FM[M][K])^T$

Algorithm 9 Generation of significant feature matrix.

Require: $FM[K][M]$, $target[K][1]$ and τ

Ensure: $SFM[K][R]$: Significant feature matrix. R: Total number of significant features to be selected. Function $size()$ sets the dimension of the array. Algorithm Feature Selection selects the most significant features.

```
Initialize τ with required value
SFlist ← Feature Selection(FM, target, τ)
R ← size(SFlist)
for i ← 1 to K do
    for j ← 1 to R do
        f ← SFlist[j]
        SF[i][j] ← FM[i][f]
    end for
end for
```

Algorithm 10 Feature selection using BLogR method.

Ensure: $SFlist1$: An array of significant features. Function $bayes\_log\_reg()$ computes the list of most significant features and $sqrt()$ computes the square root.

```
for k ← 1 to K do
    A[K][1] ← 1   {Creating unit vector A}
end for
sum ← 0
for m ← 1 to M do
    for k ← 1 to K do
        sum ← sum + FM[k][m]
    end for
    mean ← sum/k
    MEAN[1][m] ← mean   {Generating the mean vector}
    Reset sum ← 0
end for
for k ← 1 to K do
    for m ← 1 to M do
        FM1[k][m] ← FM[k][m] − A[k][1] × M[1][m]
    end for
end for
```

18: **for** $k \leftarrow 1$ to K **do**
19: **for** $m \leftarrow 1$ to M **do**
20: $SFM1[k][m] \leftarrow FM1[k][m] \times FM1[k][m]$
21: **end for**
22: **end for**
23: $sum \leftarrow 0$
24: **for** $m \leftarrow 1$ to M **do**
25: **for** $k \leftarrow 1$ to K **do**
26: $sum \leftarrow sum + SFM1[k][m]$
27: **end for**
28: $NF[1][m] \leftarrow sqrt(sum)$ {Normalizing features}
29: Reset $sum \leftarrow 0$
30: **end for**
31: **for** $m \leftarrow 1$ to M **do**
32: $RNF[1][m] \leftarrow 1/NF[1][m]$ {Reciprocal of NF}
33: **end for**
34: Compute diagonal matrix $DM[M][M]$ from $RNF[1][M]$
35: $FM1[K][M] \leftarrow FM1[K][M] \times DM[M][M]$
36: $SFlist1 \leftarrow bayes\_log\_reg(FM1, target, \tau)$

5.5 Balancing the Selected Feature Set

It is a general assumption that most machine learning algorithms assume the probabilities of target classes occurrence are same. On the contrary in real world applications like the detection of breast cancer, such assumptions are not true. We have noticed that the proportion of instances among classes are unequal. This type of condition is known as a class imbalance problem. The performance measure of the classifier may not be well-suited for evaluating the scheme using the imbalanced feature sets. To overcome this class imbalance problem, Gaussian distribution based balancing algorithm in association with sampling of instances has been used in the present case [85, 86]. The sampling based balancing approach is of two types namely, over-sampling and under-sampling. In an over-sampling technique, the new instances from minority class are replicated and added to the dataset until the dataset has an equal number of positive and negative instances. In the under-sampling technique, the instances belonging the majority class are selected until the balance is achieved between positive ad negative instances.

According to the central limit theorem, regardless of the actual sampling distribution, the sampling distribution of the mean will always approach a normal distribution [87]. Based on this criteria, the synthetic instances of minority class are generated instead of knowing the actual real sampling distribution. A new dataset is created that almost comply the actual dataset. After the new instances are put together with the original minority ones, the original sampling distribution is kept almost intact. So, the following assumptions are made about independence of the attributes, (i) every attribute of the dataset is taken to be random, and (ii) all attributes are considered to be independent of each other.

Now, assuming the derived feature dataset has K number of instances in which x_i, $i = 1, 2..., K$ is associated with class label, $C_i \rightarrow y_i \in \{-1, +1\}$, we can take K number of random variables for all the instances. In this method, the expected value of each variable is calculated using the data of the minority classes of the training set. Let us denote the standard deviation and mean of x_i as σ_i and μ_i respectively for all $i = 1, ..., K$. Consider $\mu_i^{'}$ as the mean and $\sigma_i^{'}$ as the standard deviation of the unknown random variable x_i. For the minority class instances, we assume that all the values of the attribute x_i are independent and random variables that are similarly distributed. The reason for such assumption is that they are results of different experiments, and each of them follows the same distribution function. So, according to the central limit theorem, as the sample size ($sample\_size$) becomes very large the underlying distribution tends towards a standard normal distribution i.e.,

$$\begin{aligned} &\frac{\mu_i - \mu_i^{'}}{\sigma_i \Big/ \sqrt{n}} \rightarrow N(0,1) \\ &\text{as,} \;\; sample\_size \rightarrow \infty \end{aligned} \tag{5.22}$$

where n is the number of minority class instances. Further an equation is to be induced for a given random number r_i that obeys the standard normal distribution expressed as,

$$\mu'_i = ((\mu_i - r_i)\, \sigma'_i) \big/ \left(\sqrt{n}\right) \tag{5.23}$$

where μ_i and $\mu_i^{'}$ are the means of x_i for the original and unknown minority class feature dataset. Thus for any given instance x_i, it is easy to synthesize the value for that attribute by the following equation.

$$x'_i = ((x_i - r_i)\, \sigma'_i) \big/ \left(\sqrt{n}\right), i \in 1, 2, ..., K \tag{5.24}$$

As the value of $\sigma_i^{'}$ is not known, its approximation is computed by using σ_i. To

generate the normal variates, equation (5.24) can be expressed as,

$$x'_i = ((x_i - r_i)\,\sigma_i)/\left(\sqrt{n}\right), i \in 1, 2, ..., K \tag{5.25}$$

The steps for balancing the significant feature dataset are given in **Algorithm 11**, which uses resampling strategy using Gaussian distribution.

Algorithm 11 Balancing of significant features.

Require: $SFM[K][R]$, $target[K][1]$ and C: total number of classes. K is the total number of instances in the dataset.

Ensure: $BSF[N][R]$: Balanced significant feature matrix. T: Total number of instances to be re-sampled per class label in the dataset, R: Total number of significant features.

Initialize T and $BSF \leftarrow \phi$
Create two empty matrices $M[C][K]$ and $S[C][K]$
for $c \leftarrow 1$ to C **do**
  for $k \leftarrow 1$ to K **do**
    Compute μ_{ck} and σ_{ck} $\{\mu_{ck}$: mean, σ_{ck}: standard deviation of instance k in class $c\}$
    $M[c][k] \leftarrow \mu_{ck}$
    $S[c][k] \leftarrow \sigma_{ck}$
  end for
end for
for $p \leftarrow 1$ to C **do**
  for $q \leftarrow 1$ to T **do**
    $BSF \leftarrow BSF \cup$ instance of $N(\mu_p, \sigma_p)$ $\{N(\mu_p, \sigma_p)$: normal distribution$\}$
    $target[q][1] \leftarrow p$
  end for
end for

5.6 Classification and Performance Evaluation

To validate the efficacy of the proposed scheme, several classifiers namely, Support Vector Machine (SVM), Back-Propagation Neural Network (BPNN), K-Nearest Neighbor (K-NN), Naive Bayes (NB), Logistic Model Tree (LMT), AdaBoost and Random Forest (AdaBoost-RF), and LogitBoost and Random Forest (LogitBoost-RF) are used for classification by utilizing the relevant Slantlet features for both the databases on the similar platform. The performance of the proposed scheme has

been assessed with the assistance of various metrics of confusion matrix (TP, FP, TN, and FN), classification accuracy rate (ACC), AUC value of Receiver Operating Characteristic (ROC) curve, F1 score (F_{score}), Matthews correlation coefficient (MCC), and Kappa statistics (κ). Another useful metric has been used to evaluate the performance of the classification task is a root-mean-square error (E_{rms}). The E_{rms} measures the difference between the number of predicted instances belonging to a class and actual class observed that is known as prediction errors. The classifier is having the smaller value of E_{rms} has a better performance.

5.7 Experimental Results and Discussion

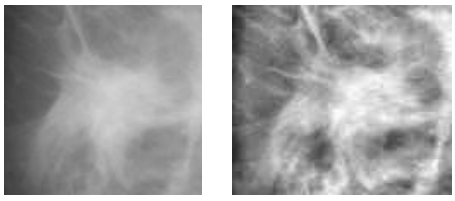
For the simulation experiments, mammographic images are taken from MIAS and DDSM databases. Two image class sets are built and used in the experiment namely, abnormal–normal and malignant–benign for the evaluation of the performance. The abnormal and malignant type of ROIs are considered as positive class in the abnormal–normal and malignant–benign image class set respectively. Each mammographic ROI has been taken of size 128 × 128 pixels used in the feature extraction phase to find the feature elements. The overall simulation is divided into three different experiments and discussed below in detail.

Experiment 1: Generation of Feature Matrix

In this scheme, a Two-Dimensional Slantlet Transform (2D-SLT) has been employed to generate the coefficients from the mammographic ROIs. Several SLT matrices are computed by using the different l-scale filter banks that have been described in Section 5.2. The value of scale, l is specified by the predefined block size, bs that is used in the division of ROI. In this experiment, suitable values of bs have been taken as 4, 8, 16, 32, 64 and 128 for the division of a 128 × 128 ROI to yield a number of cropped blocks or images. While, the value of $bs = 128$, then only one block is obtained from the ROI, which is same as the original ROI. Subsequently, various values of scale l are estimated by $l = \log_2 bs$. The corresponding SLT matrices (S) are generated using the obtained value of scale l.

Next, A Two-Dimensional Slantlet Transform is performed by the sequential row and column transformations on the individual block (CI), that produces the coefficient matrix $CICF$ that has been explained in Section 5.4. Then, all the obtained $CICF$ are concatenated to form the coefficient matrix CF of the whole ROI. The dimension of matrix CF is same as the size of ROI. The 2D-SLT of

the mammographic ROI is presented in Figure 5.5 using different values of block size, bs = 16, 32 and 64. Tables 5.1, 5.2, 5.3 and 5.4 show the different coefficient matrices generated from the individual cropped blocks using the values of block size, bs = 4, 8 for mammographic ROIs of MIAS and DDSM databases respectively. In the feature extraction phase (**Algorithm 8**), the M ($M = 128 \times 128 = 16384$) number of coefficients are obtained from a mammographic ROI of size 128×128 pixels. A feature matrix FM of size $K \times M$ is constructed by keeping all the ROI indices present in the database in rows and corresponding coefficients of each ROI in the column.



(a) Original ROI (b) Enhanced ROI

(c) SLT, $bs = 16$ (d) SLT, $bs = 32$ (e) SLT, $bs = 64$

Figure 5.5: 2D-SLT of the enhanced ROI using block sizes, bs = 16, 32, and 64.

Table 5.1: The $CICF_{4\times4}$ generated from individual blocks of different ROIs (128×128) with $bs = 4$ using SLT matrix $S_{4\times4}$ for MIAS database.

ROI type	Coefficient matrix ($CICF_{4\times4}$)			
Normal	675.7500	0.3354	0.0977	0.2558
	1.4534	-1.3500	-0.4680	0.3099
	-0.2558	0.2018	0.0764	-0.1618
	-0.0977	0.2725	0.0618	0.5236
Benign	579.7500	-1.9007	0.0977	0.2558
	-7.2672	0.7500	-0.4139	-0.0604
	-0.2185	2.8230	-0.3618	0.2236
	0.5721	1.7623	-0.2236	-0.1382
Malignant	615.7500	1.9007	-0.4513	-0.6094
	-6.1492	0.1500	-0.1144	-0.0437
	-1.2792	0.7135	0.1618	-0.2472
	-0.4886	0.0771	0.6472	-0.0618

Table 5.2: The $CICF_{8\times8}$ generated from individual blocks of different ROIs (128×128) with $bs = 8$ using SLT matrix $S_{8\times8}$ for MIAS database.

ROI type	Coefficient matrix ($CICF_{8\times8}$)							
Normal	1351.6250	-5.8101	-0.2753	-0.1200	-0.5854	0.0854	0.3191	0.4309
	11.4292	5.9583	-4.1517	-5.2849	0.3646	0.0718	-0.1371	0.0825
	0.6965	0.4372	-0.4108	0.0139	0.0631	0.1094	0.3950	0.2734
	-0.4593	-0.6960	0.3194	-0.0725	-0.0631	-0.1094	-0.4760	0.2819
	0.0264	-0.3588	-0.6984	-0.2336	0.0382	-0.0382	-0.1236	-0.1000
	0.4736	0.0315	-0.4027	-0.2465	-0.2618	0.2618	-0.1000	0.3236
	0.4736	0.0315	-0.4027	-0.2465	-0.2618	0.2618	-0.1000	0.3236
	-0.6545	-0.1615	-0.1544	0.2855	0.4000	0.1000	0.0764	-0.1618
Benign	1159.8750	-0.3546	-3.7710	-3.5813	0.4472	-0.4472	0.0691	0.1809
	5.7555	5.7321	0.9017	0.4612	-0.2484	-0.2972	-0.0266	-0.2462
	-9.7290	0.0808	-8.5698	-8.7758	0.8789	-0.5972	-0.5293	0.1168
	-10.7468	-0.2016	0.4330	0.5626	0.0030	0.3477	-0.3963	0.0183
	0.5163	0.5935	-1.6531	-1.0000	0.0618	0.3000	-0.0854	0.1382
	-0.2663	-0.2116	-2.4996	-2.9112	0.3000	-0.1618	0.3618	0.5854
	-0.8090	-0.1809	1.9570	2.9484	-0.7708	-0.2618	-0.3618	0.2236
	0.3090	-0.4737	1.6910	1.9417	-0.0382	0.5708	-0.2236	-0.1382
Malignant	1231.7500	-0.6001	0.0345	-0.0345	0.4309	0.3191	-0.0691	-0.1809
	-12.8749	-11.7143	-0.4778	0.7193	-0.4746	-0.8893	0.3166	0.8290
	3.3224	4.4116	0.0804	-0.0985	0.0262	0.2828	0.5707	-0.4456
	-3.1643	5.7669	-0.0658	-0.7161	0.4751	0.0064	-0.2776	-0.3219
	1.1219	-0.7558	0.1953	-0.4679	0.0854	0.2236	-0.0618	-0.3854
	2.1281	-0.3898	-0.0702	-0.1316	-0.2236	-0.5854	0.2854	0.1618
	-1.8517	0.3065	-0.2342	0.6649	0.1000	-0.1000	0.1618	-0.2472
	-0.3983	0.1845	0.3708	-0.0108	-0.1000	0.1000	0.6472	-0.0618

Table 5.3: The $CICF_{4\times4}$ generated from individual blocks of different ROIs (128×128) with $bs = 4$ using SLT matrix $S_{4\times4}$ for DDSM database.

ROI type	Coefficient matrix ($CICF_{4\times4}$)			
Normal	683.0000	4.6957	0.4743	-0.4743
	-2.9069	-3.2000	0.0501	0.8986
	0.1581	-0.8986	0.1000	-0.3236
	-0.1581	-0.0501	0.1236	0.1000
Benign	536.5000	-3.3541	-1.8512	-0.2701
	-4.9193	-0.2000	-0.7405	0.1080
	-0.1581	-0.0707	-0.2618	0.0382
	0.1581	0.0707	0.2618	-0.0382
Malignant	648.0000	-3.8013	-0.1581	0.1581
	-3.8013	0.0000	-0.4743	-0.4743
	-0.1581	-0.4743	0.2236	0.0000
	0.1581	-0.4743	0.0000	-0.2236

Table 5.4: The $CICF_{8\times8}$ generated from individual blocks of different ROIs (128×128) with $bs = 8$ using SLT matrix $S_{8\times8}$ for DDSM database.

ROI type	Coefficient matrix ($CICF_{8\times8}$)							
Normal	1360.0000	-5.7828	2.5067	3.8178	-0.7236	-0.2764	0.5854	-0.0854
	-6.2192	-6.4762	-5.5965	-4.2714	0.0258	-0.4622	-0.0702	1.0521
	-0.9409	4.6917	0.1326	-0.2609	-0.0939	-0.2367	0.0087	-0.3681
	-8.7041	-5.2092	0.2847	0.7436	-0.2763	0.2907	0.0787	0.5969
	2.0427	1.4968	0.1352	-0.0542	0.2000	0.3854	0.0000	0.2236
	1.7073	2.5947	-0.1136	-0.4418	-0.2854	0.2000	-0.2236	0.0000
	-2.3517	-2.8099	-0.2619	0.0935	-0.2236	-0.3618	0.1000	-0.3236
	-0.8983	-0.7361	0.1167	0.2098	-0.1382	0.2236	0.1236	0.1000
Benign	1063.8750	-7.1194	4.6744	-4.2792	1.8719	2.8781	-2.2826	-0.7174
	-12.0838	-5.9940	-0.3130	0.5717	-0.5921	-0.2262	0.0129	-0.2311
	-5.3364	0.1585	0.0941	0.0973	0.0031	-0.2920	0.1349	-0.1751
	-4.3876	0.4108	-0.0664	0.2250	0.0780	-0.2634	-0.8047	0.2124
	-0.3882	0.0129	0.1978	-0.0937	-0.0618	0.0618	0.6000	-0.1000
	-0.6118	-0.2311	0.3945	0.1339	-0.1618	0.1618	-0.1000	0.6000
	-0.0163	0.2750	-0.2544	-0.1596	0.1000	0.2618	-0.2618	0.0382
	0.7663	0.5433	0.0690	-0.1294	0.0382	0.1000	0.2618	-0.0382
Malignant	1296.2500	-0.3273	-3.1435	-3.9716	0.7663	-0.0163	-0.2927	0.0427
	5.7828	6.0833	0.3646	-0.8476	0.4156	0.2936	-0.3424	-0.3668
	-4.7664	0.3661	-0.7367	-0.0397	0.0209	0.0898	0.2674	0.0302
	-5.3529	-0.0556	-0.1270	-0.4966	0.1475	-0.4163	-0.3985	-0.3734
	0.4472	0.2785	0.1990	-0.1450	0.0618	0.0236	-0.1854	0.5472
	-0.4472	0.3761	-0.4606	0.0904	-0.4236	-0.1618	-0.3472	0.4854
	-0.3354	-0.1637	-0.3336	-0.4236	0.1000	-0.1000	0.2236	0.0000
	0.3354	-0.1637	0.2733	-0.4648	-0.1000	0.1000	0.0000	-0.2236

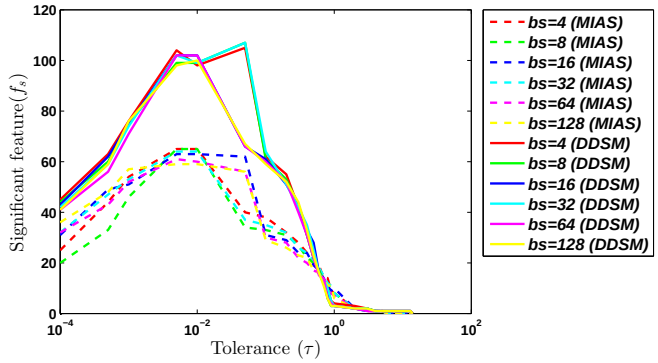
Experiment 2: Selection of Significant Features

The selection of relevant features from feature matrix is accomplished through the **Algorithms 9** and **10**. The detail description of the feature selection phase has been presented in the Section 5.4. A tolerance parameter (τ) has been utilized in the feature selection algorithm to select different number of relevant features. A value of 0.0001 has been taken in the experiment for initializing the τ. The final value of the τ is to be specified empirically, after which no further features are selected. Different values of τ are used in the feature selection algorithm in the range between initial and final values. However, the maximum and minimum number of relevant features are selected at only a particular value of τ. These specific values of τ vary for various extracted feature matrices based on the use of different block sizes (bs). Table 5.5 shows several maximum and minimum number of selected features (R_{max}, R_{min}) and the corresponding values of tolerance parameters ($\tau_{R_{max}}$, $\tau_{R_{min}}$) at which the values of earlier parameters are obtained.

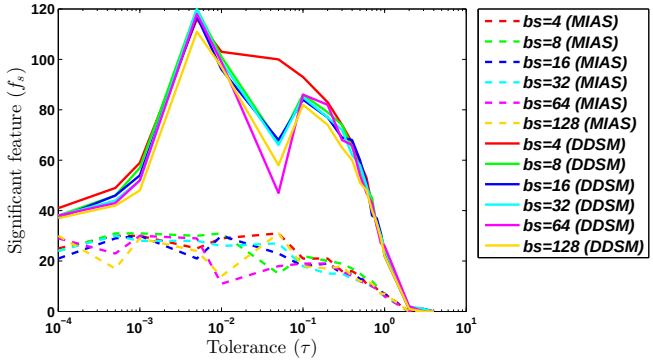
Table 5.5: Various numbers of maximum and minimum selected features.

Database	Class set	*bs*	R_{max}	$\tau_{R_{max}}$	R_{min}	$\tau_{R_{min}}$
MIAS	Abnormal–normal	4	65	0.005, 0.01	1	5
		8	65	0.005, 0.01	1	5
		16	63	0.005, 0.01	1	5
		32	64	0.005, 0.01	1	4
		64	61	0.005	1	5
		128	59	0.005, 0.01	1	5
	Malignant–benign	4	31	0.05	7	1
		8	31	0.0005, 0.001, 0.01	6	1
		16	30	0.001, 0.01	7	1
		32	30	0.0005	6	1
		64	30	0.001	6	1
		128	31	0.05	7	1
DDSM	Abnormal–normal	4	105	0.05	1	13
		8	107	0.05	1	13
		16	102	0.005, 0.01	1	13
		32	107	0.05	1	13
		64	102	0.005, 0.01	1	13
		128	1000	0.01	1	13
	Malignant–benign	4	116	0.005	1	3
		8	120	0.005	1	3
		16	117	0.005	1	3
		32	120	0.005	1	3
		64	118	0.005	2	2
		128	111	0.005	1	2

Here, R denotes the total number of selected features present in the significant feature list. It might be noted that the same number of maximum selected features are obtained for different values of τ. As shown in the Table 5.5, this condition is valid for all values of *bs* except $bs = 64$ in abnormal–normal class set of MIAS database. Similarly, for the malignant–benign class set of the MIAS, the same number of maximum features have been selected using $bs = 8$, and 16. In case of DDSM database, the values of *bs* are 16 and 64 in the abnormal–normal class set for which the above condition is valid. But, there is no such *bs* value in the malignant–benign class set of DDSM for which the same number of maximum features selected at different values of τ. The variation of the number of selected features (R) for different block sizes (*bs*) with respect to the various values of tolerance parameter (τ) has been indicated in Figure 5.6. The Figure shows that the number of selected features (R) goes to peak at some values of τ and gradually decreases to zero when the higher values of τ are used.



(a) Abnormal-normal class set



(b) Malignant-benign class set

Figure 5.6: The selection of significant features for various values of bs.

Experiment 3: Classification and Evaluation

As explained earlier in the Section 5.5, the classification performance is highly influenced by the balanced, relevant feature set that is accomplished by the Gaussian distribution based balancing procedure, than an imbalanced relevant feature set. In this work, the same number of instances are sampled for each class (both positive and negative classes) from the learnt Gaussian distribution to generate the balanced feature dataset. The number of instances belonging to each class in a set of abnormal–normal and malignant–benign of both the databases is re-sampled as given in Table 5.6. After getting several sets of balanced significant features, a classification experiment has been conducted utilizing different classifiers for both MIAS and DDSM databases. In this experiment, a 10-fold cross-validation procedure has been employed to partition the whole dataset into a train and test set for 10 number of rounds. In this process, the entire dataset is partitioned into 10 number folds out of which, nine folds are combined to form one dataset, and the remaining one fold is considered as another set. In this way, two disjoint sets are obtained containing 90% and 10% of

data that are used separately for training and validation purposes respectively. The cross-validation process is repeatedly executed for ten number of rounds with the random selection of training and testing dataset by classifiers. It has been observed that, LogitBoost-RF classifier gives a better performance using the proposed Slantlet features among all classifiers.

Table 5.6: Balancing of selected feature dataset. The abnormal and malignant types of ROIs are considered as positive.

Database	Class set	Number of instances to be re-sampled per each class	
		$y_i = +1$	$y_i = -1$
MIAS	Abnormal–normal	115	207
	Malignant–benign	51	64
DDSM	Abnormal–normal	500	500
	Malignant–benign	264	236

Several Random Forests having 10, 20, 40 and 100 number of trees with a maximum depth of two have been taken empirically in the classification experiment that are used as base learner in the LogitBoost algorithm. The optimum performance has been achieved by the Random Forest base learner with 20 number of trees. Different of performance measures namely kappa-statistics (κ), accuracy (ACC), and the root-mean-square error (E_{rms}) are estimated using that optimal structure of the classifier. The detailed computed values of various performance measures using balanced and imbalanced relevant feature set are given in Table 5.7. It might be noted in Table 5.7 that the obtained values of performance measures using the balanced feature set outperform the use of imbalanced feature set. Another vital factor has been observed that the same value of maximum accuracy is obtained at the multiple values of the block size (bs). However, we have considered the optimum block size (bs) at which the best accuracy (acc_{max}) with respect to the minimum E_{rms} value is obtained. This optimum block size $bs = 16$ for our proposed scheme. The best accuracy values (ACC_{max}) of 99.69% and 99.13% with a minimum E_{rms} have been achieved using the balanced feature set at $bs = 16$ for abnormal–abnormal and malignant–benign class sets of MIAS database respectively. The similar measures of 99.80% and 99.40% with the minimum E_{rms} are achieved using the balanced feature set at the same value $bs = 16$ for abnormal–abnormal and malignant–benign class sets of DDSM database respectively. The optimal fold-wise confusion matrices for both the databases are computed at various values of bs with tolerance parameter, $\tau = 0.01$ and are given in Table 5.8.

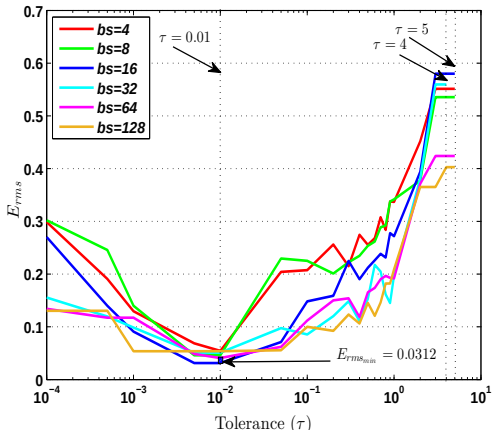
Table 5.7: Different values performance parameters such as κ, ACC E_{rms} for various values (bs) at optimal $\tau = 0.01$.

Feature set	Database	class set	Parameter	Block size (bs)						ACC_{max} (%)	bs with ACC_{max}
				bs=4	bs=8	bs=16	bs=32	bs=64	bs=128		
Imbalanced	MIAS	Normal–abnormal	ACC (%) =	91.93	93.17	94.41	93.48	93.48	93.48	99.41	16
			κ =	0.8171	0.8471	0.8759	0.8560	0.8566	0.8560		
			E_{rms}	0.2573	0.2366	0.2122	0.2224	0.2245	0.2382		
		Benign–malignant	ACC (%) =	93.91	92.17	95.65	92.17	94.78	95.65	95.65	16, 128
			κ =	0.8759	0.8399	0.9121	0.8411	0.8935	0.9114		
			E_{rms} =	0.2028	0.2561	0.1899	0.2370	0.2377	0.2142		
	DDSM	Normal–abnormal	ACC =	95.00	95.00	95.10	95.00	95.00	95.00	95.10	16
			κ =	0.9000	0.9000	0.9020	0.9000	0.9000	0.9000		
			E_{rms} =	0.2211	0.2231	0.2209	0.2235	0.2230	0.2230		
		Benign–malignant	ACC (%) =	80.4	82.6	83.40	82.60	83.40	82.60	83.40	16, 64
			κ =	0.6053	0.6494	0.6655	0.6494	0.6651	0.6481		
			E_{rms} =	0.3846	0.3496	0.3393	0.3490	0.3482	0.3546		
Balanced	MIAS	Normal–abnormal	ACC (%) =	99.69	99.69	99.69	99.38	99.69	99.67	99.69	4,8, 16,64, 128
			κ =	0.9932	0.9932	0.9932	0.9864	0.9932	0.9932		
			E_{rms} =	0.0544	0.0460	0.0312	0.0511	0.0409	0.0537		
		Benign–malignant	ACC (%) =	96.52	99.13	99.13	98.26	99.13	98.26	99.13	8,16 64
			κ =	0.9295	0.9823	0.9824	0.9646	0.9823	0.9648		
			E_{rms} =	0.1711	0.0996	0.0985	0.1048	0.0996	0.1035		
	DDSM	Normal–abnormal	ACC =	99.70	99.60	99.80	99.70	99.8	99.80	99.80	16,64, 128
			κ =	0.9940	0.9920	0.9960	0.9940	0.9960	0.9960		
			E_{rms} =	0.0512	0.0535	0.0340	0.0512	0.0360	0.0423		
		Benign–malignant	ACC (%) =	95.40	97.00	99.40	96.00	99.40	99.40	99.40	16,64, 128
			κ =	0.9077	0.9399	0.9880	0.9197	0.9880	0.9880		
			E_{rms} =	0.1892	0.1602	0.0663	0.1823	0.0663	0.0663		

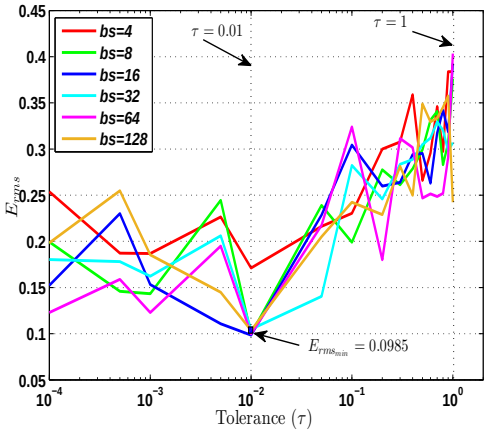
Table 5.8: Optimal confusion matrices of different databases (fold-wise) at block size, $bs = 16$ with tolerance, $\tau = 0.01$.

Database	Folds	Class set											
		Normal-abnormal						Benign-malignant					
		Training instances	Testing instances	TP	FP	TN	FN	Training instances	Testing instances	TP	FP	TN	FN
MIAS	Fold 1	289	33	12	0	21	0	103	12	6	0	6	0
	Fold 2	289	33	12	0	21	0	103	12	4	0	7	1
	Fold 3	290	32	12	0	20	0	103	12	5	0	7	0
	Fold 4	290	32	10	0	22	0	103	12	5	0	7	0
	Fold 5	290	32	11	0	21	0	103	12	5	0	7	0
	Fold 6	290	32	11	0	21	0	104	11	5	0	6	0
	Fold 7	290	32	11	0	21	0	104	11	5	0	6	0
	Fold 8	290	32	12	0	20	0	104	11	5	0	6	0
	Fold 9	290	32	11	0	20	1	104	11	5	0	6	0
	Fold 10	290	32	12	0	20	0	104	11	5	0	6	0
DDSM	Fold 1	900	100	50	0	50	0	450	50	27	0	23	0
	Fold 2	900	100	50	0	50	0	450	50	27	0	23	0
	Fold 3	900	100	50	0	50	0	450	50	26	0	23	1
	Fold 4	900	100	50	0	50	0	450	50	26	0	23	1
	Fold 5	900	100	50	0	50	0	450	50	26	1	23	0
	Fold 6	900	100	50	0	50	0	450	50	26	0	24	0
	Fold 7	900	100	50	0	50	0	450	50	26	0	24	0
	Fold 8	900	100	50	2	48	0	450	50	26	0	24	0
	Fold 9	900	100	50	0	50	0	450	50	26	0	24	0
	Fold 10	900	100	50	0	50	0	450	50	26	0	24	0

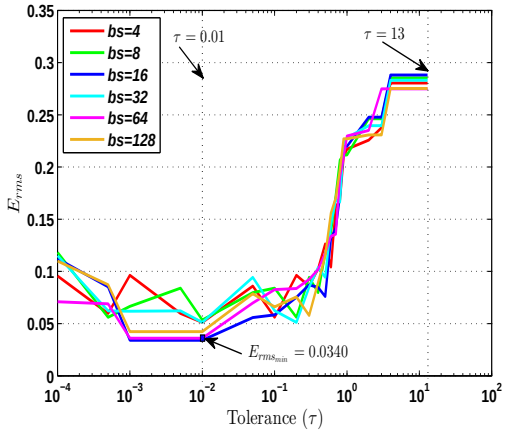
Further, a comparison among various values of E_{rms} has been made by using the different values of bs that are given in Figure 5.7 for both the class sets of MIAS and DDSM databases. All the minimum E_{rms} values are obtained at $\tau = 0.01$ using $bs = 16$. The minimum E_{rms} values of 0.0312 and 0.0985 are obtained for abnormal–normal and malignant–benign class sets of MIAS database respectively. The values of similar parameter are 0.0340 and 0.0663 for abnormal–normal and malignant–benign class sets of DDSM database respectively. The comparison of various accomplished performance measures namely ACC, S_n, κ, E_{rms}, MCC, and AUC of the LogitBoost-RF has been made with that of other classifiers using balanced feature set and are presented in Table 5.9.



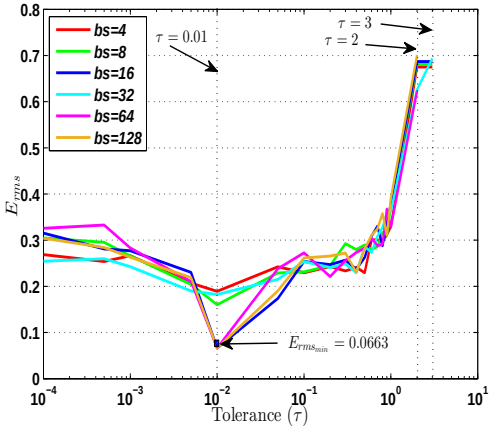
(a) Abnormal-normal class set (MIAS)



(b) Malignant-benign class set (MIAS)



(c) Abnormal-normal class set (DDSM)



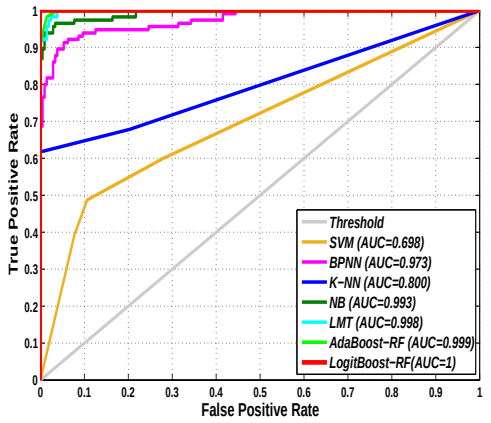
(d) Malignant-benign class set (DDSM)

Figure 5.7: Comparison of values of E_{rms} using the various values of *bs* and τ.

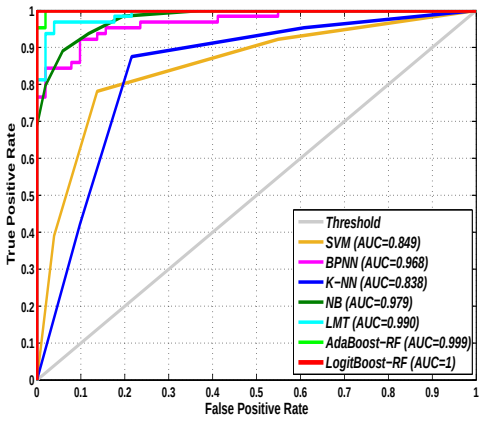
Table 5.9: Comparison of optimal classifier with other classifiers with respect to performance at $\tau = 0.01$, $bs = 16$.

Database	Classifier	Measures of performance													
		Abnormal-normal							Malignant-benign						
		ACC (%)	S_n (%)	κ	E_{rms}	F_{score}	MCC	AUC	ACC (%)	S_n (%)	κ	E_{rms}	F_{score}	MCC	AUC
MIAS	NB	95.96	93.9	0.9119	0.1662	0.943	0.912	0.993	91.30	98	0.8231	0.2685	0.900	0.824	0.979
	K-NN	86.34	61.7	0.6748	0.3684	0.763	0.714	0.800	83.48	78.4	0.6633	0.4027	0.808	0.664	0.838
	BPNN	92.86	85.2	0.8411	0.2550	0.895	0.844	0.973	90.45	88.2	0.8058	0.2883	0.891	0.806	0.968
	SVM	74.85	48.7	0.4100	0.4998	0.580	0.4260	0.698	81.74	86.3	0.6351	0.4234	0.807	0.640	0.849
	Adaboost-RF	98.45	98.3	0.9662	0.1767	0.978	0.966	0.999	97.39	88.2	0.9473	0.1317	0.971	0.947	0.999
	LogitBoost-RF	**99.69**	**99.1**	**0.9932**	**0.0312**	**0.996**	**0.993**	**1**	**99.13**	**98.0**	**0.9824**	**0.0985**	**0.990**	**0.983**	**1**
	LMT	96.27	98.6	0.9179	0.1400	0.971	0.919	0.998	94.78	92.2	0.8638	0.1990	0.940	0.894	0.990
DDSM	NB	95.80	95.6	0.9160	0.1813	0.958	0.916	0.989	92.00	96.2	0.8394	0.2711	0.925	0.839	0.973
	K-NN	85.9	72.6	0.7180	0.3751	0.837	0.745	0.859	75.00	61	0.5074	0.4989	0.720	0.535	0.758
	BPNN	96.1	94.8	0.9220	0.1850	0.960	0.922	0.988	91.20	92.8	0.8232	0.2609	0.918	0.823	0.970
	SVM	84.6	69.2	0.6920	0.3920	0.818	0.7270	0.846	67.20	54.2	0.3533	0.5714	0.636	0.371	0.680
	Adaboost-RF	98.00	99.2	0.9600	0.1430	0.980	0.960	0.998	97.20	93.2	0.9439	0.1539	0.973	0.944	0.997
	LogitBoost-RF	**99.80**	**100**	**0.9960**	**0.0340**	**0.998**	**0.996**	**1**	**99.40**	**99.2**	**0.9880**	**0.0663**	**0.994**	**0.988**	**1**
	LMT	93.7	94.2	0.8740	0.2478	0.937	0.874	0.974	90.4	90.9	0.8074	0.2659	0.909	0.807	0.967

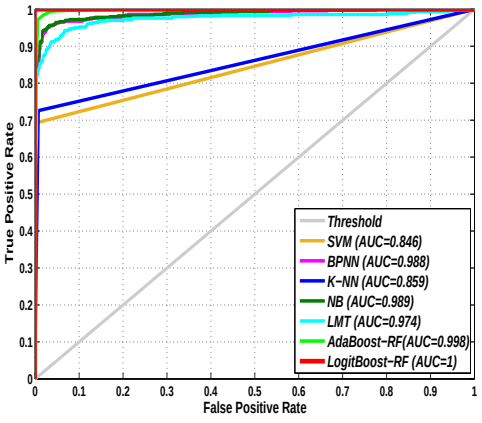
The classification performance of the proposed scheme has also been evaluated by the analysis of Receiver Operating Characteristic (ROC) curve. The index value AUC of area under the ROC curve achieved by the LogitBoost-RF classifier with that of other classifiers has been compared and given in Figure 5.8. It may be noted that the AUC values obtained by LogitBoost-RF is larger than that of other classifiers. The optimum best AUC value is 1 for all the class sets of both MIAS and DDSM database which has been accomplished by using $\tau = 0.01$ with $bs = 16$. Finally, a comparative analysis of different performance measures achieved by the proposed work with existing approaches is summarized in Table 5.10. It is clearly observed that the suggested scheme outperforms its competent approaches.



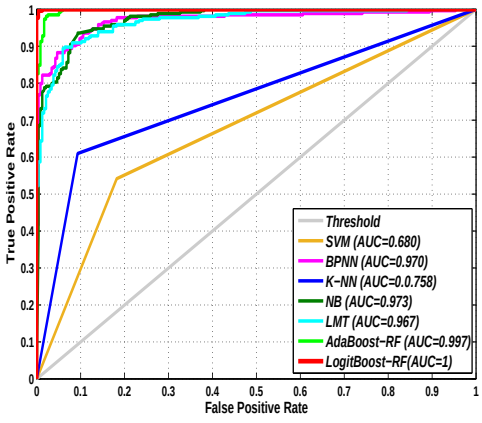
(a) ROC for abnormal ROIs (MIAS)



(b) ROC for malignant ROIs (MIAS)



(c) ROC for abnormal ROIs (DDSM)



(d) ROC for malignant ROIs (DDSM)

Figure 5.8: Comparison of ROC curves obtained by LogitBoost-RF classifier with that of other classifiers at $\tau = 0.01$ and $bs = 16$.

Table 5.10: Comparison of performances between proposed and existing schemes.

Scheme	**Technique**	**Database**	**Measurement**
Biswas *et al.* (2011) [35]	Multiscale filter bank, mixture of Gaussian distribution, EM algorithm	MIAS DDSM	$ACC = 82.5\%$, $AUC = 0.83$ $ACC = 88.3\%$, $AUC = 0.87$ (Architectural distortion)
Ramos *et al.* (2012) [38]	Db3 wavelet, GA, Random Forest	DDSM	$AUC = 0.90$ (abnormal–normal)
Eltoukhy *et al.* (2012) [39]	Curvelet transform, statistical t-test, SVM	MIAS	$ACC = 95.98\%$ (abnormal–normal) $ACC = 97.30\%$ (malignant–benign)
Nascimento *et al.* (2013) [43]	Biorthogonal 3.7 wavelet, Polynomial classifier	DDSM	$AUC = 0.98$ (malignant–normal) $AUC = 0.95$ (benign–normal) $AUC = 0.96$ (malignant–benign)
Li *et al.* (2015) [52]	Texton features, multiple subsampling strategies, K-NN	DDSM	$ACC = 85.96\%$ (malignant–benign)
Görgel *et al.* (2015) [51]	SWT, SVM	MIAS I.U. database	$ACC = 90.1\%$ (malignant–benign) $ACC = 91.4\%$ (malignant–benign)
Proposed scheme	**SLT+BLogR +LogitBoost-RF**	**MIAS**	$\boldsymbol{ACC = 99.69\%}$, $\boldsymbol{AUC = 1}$ **(abnormal–normal)** $\boldsymbol{ACC = 99.13\%}$, $\boldsymbol{AUC = 1}$ **(malignant–benign)**
		DDSM	$\boldsymbol{ACC = 99.80\%}$, $\boldsymbol{AUC = 1}$ **(abnormal–normal)** $\boldsymbol{ACC = 99.40\%}$, $\boldsymbol{AUC = 1}$ **(malignant–benign)**

5.8 Summary

In this chapter, a proficient mammogram classification scheme has been suggested to help radiologists for the interpretation of suspicious mammographic tissues in the early detection of breast cancer. The scheme applies a Two-Dimensional Slantlet Transform (2D-SLT) for the extraction of features from mammographic ROIs. The most discriminatory feature elements are selected by the use of the Bayesian Logistic Regression (BLogR) method. The relevant features are balanced by the Gaussian distribution based balancing method to attain the improved classification performance. Several classifiers such as Naive Bayes (NB), K-Nearest Neighbor (K-NN), Back-Propagation Neural Network (BPNN), Support Vector Machine (SVM), AdaBoost and Random Forest (AdaBoost-RF), and Logistic Model Tree (LMT) are used along with LogitBoost and Random Forest (LogitBoost-RF) for the classification of mammograms using selected Slantlet features for both MIAS and DDSM databases. The experimental results show an optimal classification performance has been achieved by LogitBoost-RF classifier. The proposed approach achieves the optimal accuracy results of 99.69% and 99.13% for abnormal–normal and the malignant–benign class set on MIAS database respectively. The similar parameters of 99.80% and 99.40% are accomplished for DDSM database. The optimal AUC of value 1 with respect to ROC curve is achieved for all the class sets of both the databases. Comparative analysis regarding various performance measures demonstrates that the suggested scheme outperforms its competent approaches.

Chapter 6

Mammogram Classification using Radial Symmetric Features followed by t-SNE Feature Selection

---

Across the globe, the breast cancer is frequently encountered in women, and it is the second most cause of deaths after lung cancer. Currently, the death rates have been declined sharply due to the early detection and effective treatments [88]. Recent developments in digital mammography imaging systems have aimed to better diagnosis of abnormalities in the barest. Digital mammograms are computerized scanned X-ray images of breasts. The early detection and diagnosis of breast cancer can be achieved through the mammography screening programs assisted by computer technologies [7]. In this regard, the Computer-Aided Diagnosis (CAD) is a very popular and efficient method that includes the sets of automatic tools using image processing and pattern recognition techniques to help the radiologists in the detection and classification of tissue abnormalities. Generally a CAD system consists of following three important stages, (a) extraction of features from ROI, (b) selection of useful features, and (c) classification of the breast tissues.

In this chapter, Fast Radial Symmetry Transform (FRST) has been utilized to extract the features from the mammographic ROI. A feature selection algorithm based on t-distributed Stochastic Neighbor Embedding (t-SNE) method has been used to select most significant features from a large number of radial

symmetric features. Different classifiers namely, Support Vector Machine (SVM), Back-Propagation Neural Network (BPNN), K-Nearest Neighbor (K-NN), Naive Bayes (NB), AdaBoost and Random Forest (AdaBoost-RF), LogitBoost and Random Forest (LogitBoost-RF), and Logistic Model Tree (LMT) are employed to characterize the mammograms as malignant–normal, malignant–benign, and benign–normal class sets using relevant features. It has been observed that LMT classifier shows a better performance among all the classifiers. In this chapter, detail explanation of classification performed by LMT has been given. Thus, the proposed scheme consists of three principal phases: feature extraction, selection, and classification. The overall block diagram of the proposed scheme is shown in Figure 6.1.

The chapter is organized as follows: The extraction of features using FRST method is described in Section 6.1. Section 6.2 outlines the selection of significant features. The classification and evaluation of the performance is explained in Section 6.3. Section 6.4 describes the experimental results obtained on the standard database MIAS. Section 6.5 summarizes the overall work proposed in this chapter.

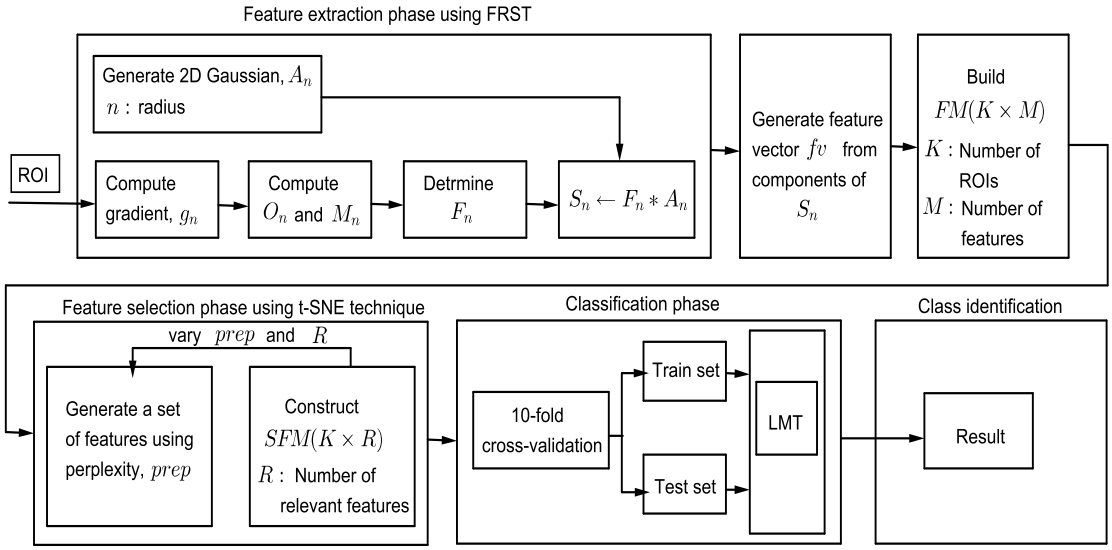


Figure 6.1: Block diagram of proposed scheme using FRST and t-SNE.

6.1 Extraction of Features using FRST

The Fast Radial Symmetry Transform (FRST) is proposed by Loy *et al.* that uses the local radial symmetry to derive the point of interest in an image [89]. The methodology decides the contribution of each pixel to the symmetry of pixels around it, instead of considering the contribution of a local neighborhood to a central pixel. Unlike

the circular Hough transform it does not require the gradient to be quantized into angular bins, the contribution of every orientation is calculated in a single pass over the image. The approach is extremely efficient regarding its cost of computation. The computational cost is of order $O(P \times L)$ while considering local radial symmetry in $L \times L$ neighborhoods over an image of P pixels. The FRST is exhibited in the following steps to detect the radially symmetric features by using one or more radii $n \in N$, where N is the set of radii of those features to be detected.

1. The gradient (g) is computed with the help of 3×3 sobel operator.

2. The value of the transform at radius n indicates the contribution to radial symmetry of the gradients with a distance n away from each point p.

3. Two images, orientation projection image (O_n) and magnitude projection image (M_n) are generated at each radius n, by considering the gradient (g) at each point p.

4. Two types of pixel, positive-affected pixel ($P_+(p)$) and negative-affected pixel ($P_-(p)$) are determined from the corresponding point (p) at which the gradient is computed as shown in Figure 6.2. The $P_+(p)$ is the pixel that the gradient of it, $g(p)$ is pointing to at a distance n away from p. Similarly, $P_-(p)$ is the pixel that the gradient is away from that pixel at a distance n.

5. The coordinates of pixels $P_+(p)$ and $P_-(p)$ are given by,

$$\begin{aligned} P_+(p) &= p + \text{ round } \left(\frac{g(p)}{\|g(p)\|} n\right), \text{ and} \\ P_-(p) &= p - \text{ round } \left(\frac{g(p)}{\|g(p)\|} n\right). \end{aligned} \tag{6.1}$$

6. Initially, $O_n = 0$ and $M_n = 0$. Next, for each pair of affected pixels, the corresponding $P_+(p)$ and $P_-(p)$ are updated as,

$$\begin{aligned} O_n(P_+(p)) &= O_n(P_+(p)) + 1, \\ O_n(P_-(p)) &= O_n(P_-(p)) - 1, \\ M_n(P_+(p)) &= M_n(P_+(p)) + \|g(p)\|, \\ M_n(P_-(p)) &= M_n(P_-(p)) - \|g(p)\|. \end{aligned} \tag{6.2}$$

7. Finally, the radial symmetry contribution at radius n is given as convolution and expressed by,

$$S_n = F_n * A_n \tag{6.3}$$

where

$$F_n(p) = \frac{M_n(p)}{k_n}\left(\frac{\left|\widetilde{O_n(p)}\right|^{\alpha}}{k_n}\right), \text{ and}$$
$$\widetilde{O_n}(p) = \begin{cases} O_n(p) & \text{if } O_n(p) < k_n \\ k_n & \text{otherwise.} \end{cases} \tag{6.4}$$

Here, A_n: 2D Gaussian, α: radial strictness parameter, k_n: scaling factor that normalizes O_n and M_n over different radii.

The Gaussian kernel A_n is required to spread the influence of the pixels, $P_+(p)$ and $P_-(p)$ as a function of n. In this approach, a rotational invariant 2D Gaussian (A_n) has been chosen since it has a consistent effect over all the $g(p)$. The A_n is also separable for which the convolution performed by it can be resolved efficiently. The strictness parameter, α determines how strictly the FRST produce a valid feature value. A larger value of the α eliminates non-radially symmetric features. The normalizing scale factor normalizes the O_n and M_n over the radius, n for representing them on a similar scale. In this scheme, the O_n and M_n are normalized through the division by their corresponding maximum values. The FRST is performed on the mammographic ROIs based on the bright symmetry, for which only the positive-affected pixels, $P_+(p)$ are considered to determine O_n and M_n. Thus, the mathematical components of the computed radial symmetry contribution (S_n) of ROIs are kept in feature vectors (fvs) and further these K number of vectors are used to construct the radial symmetric feature matrix (FM) as described in **Algorithm 12**.

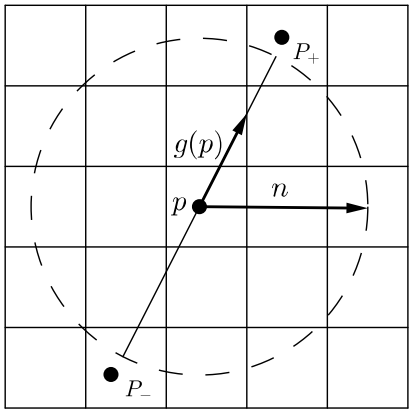


Figure 6.2: The locations of pixels $P_+(p)$ and $P_-(p)$ affected by the gradient $g(p)$ at a point p for a range of radius, n.

Algorithm 12 Feature matrix generation using FRST.

Require: K number of ROIs, n: radius, α: strictness parameter, and sf: standard factor

Ensure: $FM[K][M]$: Feature matrix. Here, M: Total number features. Function $gradient()$ computes the gradient of ROI using 3×3 Sobel operator. Function $round()$ rounds the value of the parameter to the nearest integer. Functions $abs()$, and $max()$ compute the absolute and maximum values, respectively.

Create empty feature vector fv
Initialize n and α, sf with required values, and $c \leftarrow 1$
$M \leftarrow N \times N$ {N: Size of the ROI}
Generate a 2D Gaussian filter, A_n using n and sf
for $k \leftarrow 1$ to K **do**
  Read ROI_k
  $g \leftarrow gradient(ROI_k)$
  for $i \leftarrow 1$ to N **do**
    for $j \leftarrow 1$ to N **do**
      Determine coordinate of point: $p \leftarrow (i, j)$
      compute gradient of point p: $g_p \leftarrow g(p)$
      Normalize the g_p: $g_{p_{norm}} \leftarrow \sqrt{g_p \times g'_p}$
      if $g_{p_{norm}} > 0$ **then**
        $P_+(p) \leftarrow p + round(g_p / g_{p_{norm}}) \times n$
        $O_n(P_+(p)) \leftarrow O_n(P_+(p)) + 1$
        $M_n(P_+(p)) \leftarrow M_n(P_+(p)) + g_{p_{norm}}$
      end if
    end for
  end for
  $O_n \leftarrow abs(O_n(P_+(p)))$, $M_n \leftarrow abs(M_n(P_+(p)))$
  $O_{n_{norm}} \leftarrow O_n / \max(O_n)$ {Normalize O_n and M_n}
  $M_{n_{norm}} \leftarrow M_n / \max(O_n)$
  $F_n \leftarrow M_{n_{norm}} \times (O_{n_{norm}})^{\alpha}$
  $S_{n_k} \leftarrow F_n * A_n$ {S_{n_k}: radial symmetry contribution}
  for $p \leftarrow 1$ to N **do**
    for $q \leftarrow 1$ to N **do**
      $fv_k[c][1] \leftarrow S_{n_k}[p][q]$
      $c \leftarrow c + 1$
    end for
  end for

31: Reset $c \leftarrow 1$
32: **for** $m \leftarrow 1$ to M **do**
33: $FM[m][k] \leftarrow fv_k[m][1]$
34: **end for**
35: **end for**
36: $FM[K][M] \leftarrow (\text{FM[M][K]})'$

6.2 Selection of Features using t-SNE method

In feature selection phase, an optimal set of relevant features are selected from the extracted feature matrix. The high-dimensional dataset is reduced to low-dimensional space by the use of t-distributed Stochastic Neighbor Embedding (t-SNE) method [90]. The SNE is a probabilistic approach to the task of placing feature matrix $FM = \{fv_1, fv_1, ..., fv_K\}$, represented by high-dimensional dataset or by pairwise similarities, in a low-dimensional significant dataset $SFM = \{sfv_1, sfv_1, ..., sfv_K\}$, in such a way that it preserves neighbor identities. Under the Gaussian centered of each object, the similarities of data points fv_j to fv_i in FM, and sfv_j to sfv_i in SFM are the affinities of data points, and can be represented as conditional probabilities $p_{j|i}$ and $q_{j|i}$, respectively, and expressed as,

$$p_{j|i} = \frac{\exp\left(-\|fv_i - fv_j\|^2 / 2\sigma_i^2\right)}{\sum_{m \neq i} \exp\left(-\|fv_i - fv_m\|^2 / 2\sigma_i^2\right)}, \tag{6.5}$$

$$q_{j|i} = \frac{\exp\left(-\|sfv_i - sfv_j\|^2\right)}{\sum_{m \neq i} \exp\left(-\|sfv_i - sfv_m\|^2\right)}, \tag{6.6}$$

where σ_i is the variance of the Gaussian centered on data point fv_i for computation of $p_{j|i}$. For computation of $q_{j|i}$ the value of σ_i is set to $1/\sqrt{2}$.

A natural cost function that measures the faithfulness of $p_{j|i}$ and $q_{j|i}$, is a sum of Kullback-Leibler (KL) divergences over all data points and given as,

$$C = \sum_i KL\left(P_i \| Q_i\right) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}} \tag{6.7}$$

where P_i is the conditional probability distribution over all other data points given data point fv_i, and Q_i represents the conditional probability distribution over all other map points given map point sfv_i. The value of σ_i is determined by the binary

search performed by SNE that produces a P_i with a fixed perplexity ($prep$) parameter that is determined by manually. The perplexity parameter is given as,

$$prep\left(P_i\right) = 2^{H(P_i)}, \tag{6.8}$$

where $H(P_i)$ is the Shannon entropy of P_i and expressed as,

$$H\left(P_i\right) = -\sum_j p_{j|i} \log_2 p_{j|i}\,. \tag{6.9}$$

The perplexity determines the smooth measure of the effective number of neighbors. A gradient descent method is used to minimize the cost function given in (6.7). The form of gradient descent method is expressed as,

$$\frac{\partial C}{\partial sfv_i} = 2\sum_j \left(p_{j|i} - q_{j|i} + p_{i|j} - q_{i|j}\right)\left(sfv_i - sfv_j\right). \tag{6.10}$$

The gradient descent is initialized by sampling map points randomly from an isotropic Gaussian with a small variance that is centered around the origin. A relatively high momentum term is added to the gradient to speed up the optimization as well as avoid the poor local minima. That is, the current gradient is added to an exponentially decaying sum of previous gradients to determine the changes in the coordinates of the map points at each iteration of the gradient search. The gradient update with a momentum term is defined as,

$$\begin{aligned} SFM^{(n)} \leftarrow SFM^{(n-1)} + \eta\tfrac{\partial C}{\partial SFM} + \\ \alpha\left(n\right)\left(SFM^{(n-1)} - SFM^{(n-2)}\right) \end{aligned} \tag{6.11}$$

where $SFM^{(n)}$ is the solution at iteration n, η is the learning rate, and $\alpha(n)$ is the momentum at iteration n.

In t-SNE, a Student t-distribution with one degree of freedom as the heavy-tailed distribution is used in the low-dimensional map. Utilizing this distribution, the joint probabilities q_{ij} is given as,

$$q_{ij} \leftarrow \frac{\left(1 + \left\|sfv_i - sfv_j\right\|^2\right)^{-1}}{\sum_{m \neq l}\left(1 + \left\|sfv_i - sfv_j\right\|^2\right)^{-1}}. \tag{6.12}$$

Now, the gradient of KL-divergence between joint probability distribution, P in high-dimensional space and the joint probability distribution, Q in low-dimensional space is given as,

$$\begin{aligned} \tfrac{\partial C}{\partial SFM} \leftarrow 4\sum_j \left(p_{ij} - q_{ij}\right)\left(sfv_i - sfv_j\right) \\ \left(\left(1 + \left\|sfv_i - sfv_j\right\|^2\right)^{-1}\right). \end{aligned} \tag{6.13}$$

The mere description of the selection of subsets of significant features using the t-SNE technique is given in **Algorithm 13**.

Algorithm 13 Significant feature selection using t-SNE.

Require: $FM = \{fv_1, fv_1, ..., fv_K\}$: feature matrix containing K number of feature vectors, *prep*: perplexity parameter, N: number of iteration, η: learning rate, and $\alpha(n)$: momentum

Ensure: $SFM^{(n)} = \{sfv_1, sfv_1, ..., sfv_K\}$: reduced set of significant features. Here, sfv is the vector contains reduced number (R) of significant features

1: Compute pairwise affinities with *prep*:

$$p_{j|i} \leftarrow \frac{\exp\left(-\|fv_i - fv_j\|^2 / 2\sigma_i^2\right)}{\sum_{m \neq i} \exp\left(-\|fv_i - fv_m\|^2 / 2\sigma_i^2\right)}$$

2: $p_{ij} \leftarrow \frac{p_{j|i} + p_{i|j}}{2K}$

3: Sample initial solution:

$$SFM^{(0)} \leftarrow \{sfv_1, sfv_1, ..., sfv_K\} \text{ from } \mathcal{N}(0, 10^{-4}I)$$

4: **for** $n \leftarrow 1$ to N **do**

5: Compute low-dimensional affinities:

$$q_{ij} \leftarrow \frac{\left(1 + \|sfv_i - sfv_j\|^2\right)^{-1}}{\sum_{m \neq l} \left(1 + \|sfv_i - sfv_j\|^2\right)^{-1}}$$

6: Compute gradient:

$$\frac{\partial C}{\partial SFM} \leftarrow 4 \sum_j \left(p_{ij} - q_{ij}\right)\left(sfv_i - sfv_j\right) \left(\left(1 + \|sfv_i - sfv_j\|^2\right)^{-1}\right)$$

7: Compute the gradient update:

$$SFM^{(n)} \leftarrow SFM^{(n-1)} + \eta \frac{\partial C}{\partial SFM} + \alpha(n)\left(SFM^{(n-1)} - SFM^{(n-2)}\right)$$

8: **end for**

6.3 Classification and Performance Evaluation

In order to validate the efficacy of proposed scheme, several classifiers such as Support Vector Machine (SVM), Back-Propagation Neural Network (BPNN), K-Nearest

Neighbor (K-NN), Naive Bayes (NB), AdaBoost and Random Forest (AdaBoost-RF), LogitBoost and Random Forest (LogitBoost-RF), and Logistic Model Tree (LMT) are used utilizing the relevant feature sets. Various measurements namely, true positive rate (TPR) or sensitivity (S_n), false positive rate (FPR), precision (p), recall (r), accuracy (ACC), F-measure (F_{score}), and Matthews correlation coefficient (MCC) are used to evaluate the performance of classifiers. The Kappa-statistic (κ), root-mean-square error (E_{rms}), and area under curve (AUC) value of Receiver Operating Characteristic (ROC) are also important performance evaluation metrics that are assessed in this chapter. The root-mean-square error measures the difference between the number of predicted samples belonging to a class and actual class observed that is known as prediction errors. The classifier is having the smaller value of E_{rms} determines a better performance. In this chapter, LMT classifier shows the better performance than all other classifiers.

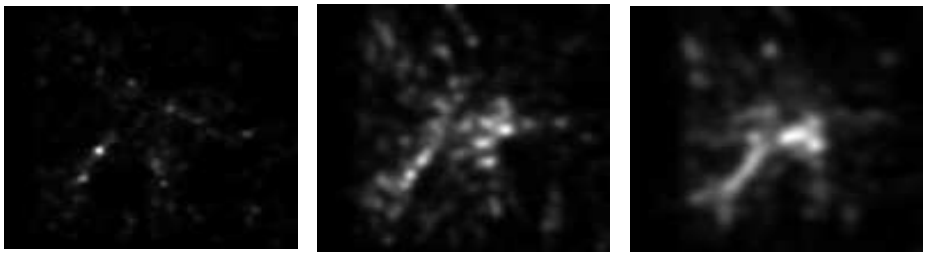
6.4 Experimental Results and Analysis

For simulation, mammographic images are taken from MIAS and DDSM databases. Each mammographic ROI has been taken of size 128×128 pixels used in the feature extraction phase to find the feature elements. The overall simulation is divided into three different experiments and discussed below in detail.

Results for Feature Extraction:

Using FRST transform, the bright radial symmetric contribution (S_n) is computed at radius n for each of the K number of mammographic ROIs as described in **Algorithm 12**. A standard factor value of 0.1 has been used for the calculation of the Gaussian kernel A_n. For generation of high performance feature, the α of value 2 is used in the experiment. A total M ($M = 128 \times 128 = 16384$) number of mathematical components are generated in each contribution, S_n. All the mathematical components of each S_n are stored in the respective feature vectors (fv). A feature matrix FM of size $K \times M$ is constructed by keeping all K number of fvs in row wise fashion. In this work, the values of radius, $n = 1,\ 2,\ \ldots,\ 25$ are taken empirically to compute S_n and it has been found that at $n = 7$, the useful feature sets are generated. The FRSTs of ROI at radii, $n = 1$, 7, 13, 19, and 25 have been computed, and shown in Figure 6.3. The FRST at $n = 7$ highlights the bright interest points excluding more or less unwanted feature components.

(a) Original ROI (b) $n = 1$ (c) $n = 7$



(d) $n = 13$ (e) $n = 19$ (f) $n = 25$

Figure 6.3: Fast radial symmetry transform (FRST) of the mammographic ROI, (a) original malignant ROI (mdb117 of MIAS database), and (b), (c), (d), (e), and (f) show the transformed ROIs that are computed at radii, $n = 1$, 7, 13, 19, and 25.

Results for Feature Selection:

The selection of significant features (SFM) from feature matrix is carried out through the **Algorithm 13**. In the t-SNE method, the reduction of features is based on the different values of perplexity (*prep*) parameter, and the dimension (R) of the SFM. In this experiment, we have taken the typical values of the *prep* parameter between 5 and 50. It has been observed that, at the value of $prep = 30$, the reduced feature subsets are very useful with respect to the classification performance. Similarly, the values of dimension parameter (R) are taken as 30, 50, 70, 90, 110, . . . , 490, 510 to build the relevant feature matrix (SFM). Deciding the dimension of SFM is a heuristic based on its usefulness. It has been noticed that the SFM of higher dimension that starts from 450 gives the saturated performance in the classification phase, which is presented in Figure 6.4. Therefore, the setting of the higher value of the dimension has been stopped at $R = 510$ for reduction of the FM to SFM.

Results for Classification:

For classification, three image class sets are formed and used in the experiment namely, malignant-normal, malignant–benign, and benign–normal. The malignant

type of ROIs is considered as the positive class in both malignant–normal, and malignant–benign class sets. Similarly, the benign type is considered as the positive class in the benign–normal class set. The number of samples per each class set for both MIAS and DDSM databases used in the classification experiment is given in Table 6.1.

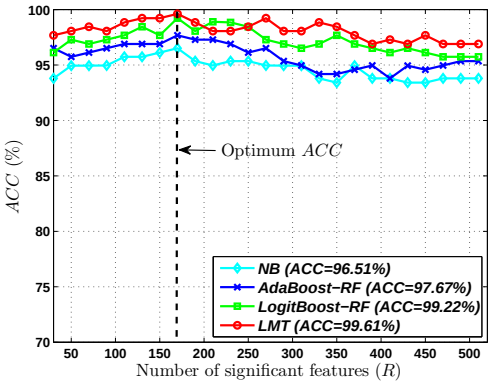
Table 6.1: The number of samples per each class set used in the classification.

Class set	Number of samples per each class	
	MIAS	DDSM
Malignant–normal	258	764
Malignant–benign	115	500
Benign–normal	271	736

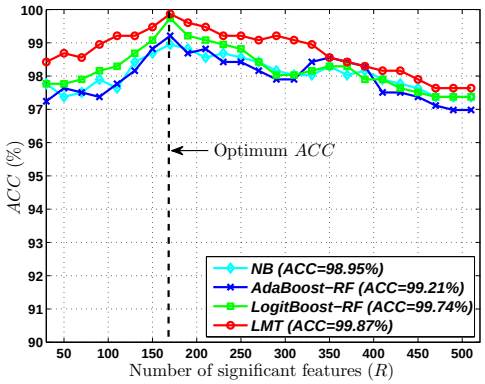
In this experiment, a 10-fold cross validation procedure has been employed to partition the whole dataset into a train and test set for ten rounds. In this process, the entire dataset is partitioned into ten folds out of which, nine folds are combined to form one dataset, and the remaining one fold is considered as another set. In this way, two disjoint sets are obtained containing 90% and 10% of data that are used separately for training and validation purposes respectively. The cross-validation process is repeatedly executed for ten rounds with the random selection of training and testing dataset by the classifier. Several classifiers namely, Support Vector Machine (SVM), Back-Propagation Neural Network (BPNN), K-Nearest Neighbor (K-NN), Naive Bayes (NB), AdaBoost and Random Forest (AdaBoost-RF), and LogitBoost and Random Forest (LogitBoost-RF) other than Logistic Model Tree (LMT) are used to compare the classification performance.

Figure 6.4 summarizes the classification accuracies (ACC) achieved by four different classifiers such as NB, AdaBoost-RF, LogitBoost-RF, and LMT using different numbers of significant feature (R). It has been observed that the best values of accuracies are obtained at $R = 170$ for all four classifiers on both MIAS and DDSM database. Thus, $R = 170$ is considered as the optimal dimension of relevant feature set for classification.

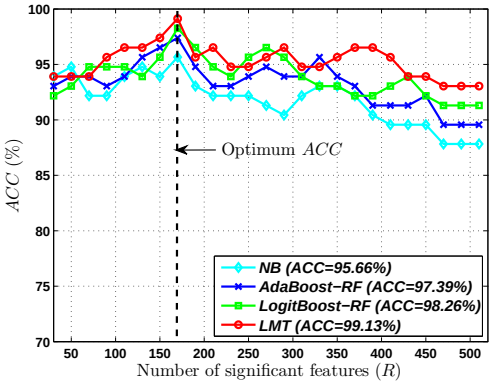
The classification performance of the proposed scheme has also been evaluated by the analysis of Receiver Operating Characteristic (ROC) curve. The AUC value of ROC curve achieved by the LMT classifier with that of other classifiers has been



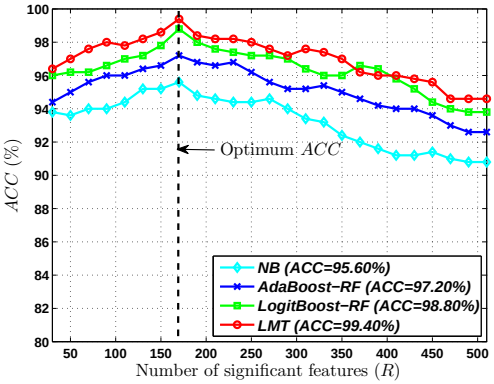
(a) Malignant–normal (MIAS)



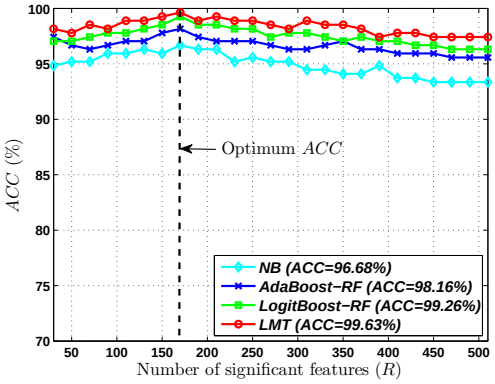
(b) Malignant–normal (DDSM)



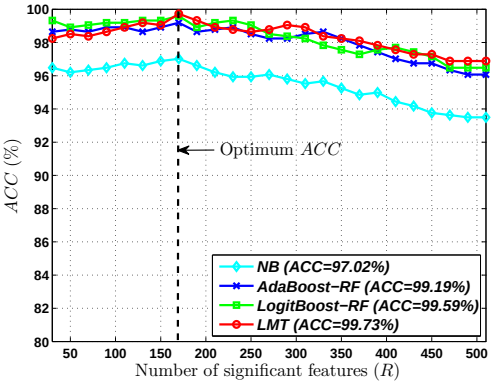
(c) Malignant–benign class set (MIAS)



(d) Malignant–benign (DDSM)



(e) Benign–normal (MIAS)



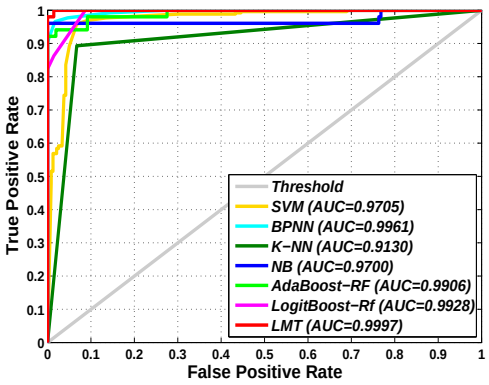
(f) Benign–normal class (DDSM)

Figure 6.4: Comparison of values of classification accuracy (ACC) obtained by various classifiers at different numbers of significant feature (R). The optimum values of accuracy are obtained at $R = 170$ for all classifiers on both MIAS and DDSM database.

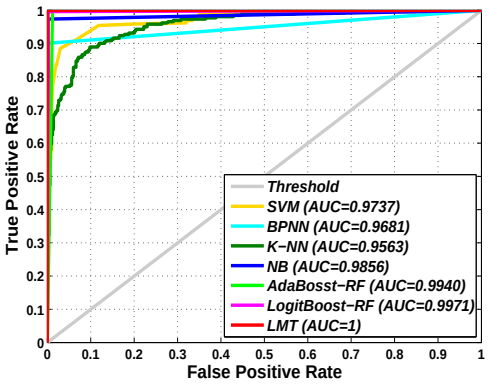
compared and given in Figure 6.5. It may be noted that the AUC values obtained by LMT classifier are larger than that of other classifiers for all the image class sets on MIAS and DDSM databases. The best values of AUC are 0.9997, 1, and 0.9998 for malignant–normal, malignant–benign, and benign–normal class sets, respectively using MIAS database. Similarly, the parameters are 1, 1, and 0.9968 for DDSM database. The optimal fold-wise confusion matrices of both the databases computed by the LMT classifier using an optimal subset of relevant features is given in Table 6.2.

The performance of mammogram classification is also assessed by different measures obtained through various classifiers using the optimal number of significant features ($R = 170$) is presented in Table 6.3. Similarly, various performance metrics namely, kappa-statistics (κ), accuracy (ACC), root-mean-square error (E_{rms}), and AUC value of ROC curve are estimated using the optimal structure of the classifier. The detailed computed values of different performance metrics using relevant feature set (SFM) having size, $R = 170$ are given in Table 6.4. The proposed approach achieves the optimal accuracy (ACC) values of 99.61%, 99.13%, and 99.63% for malignant–normal, malignant–benign, and benign–normal class sets, respectively using MIAS database . For DDSM database, the similar parameters are of 99.87%, 99.40%, and 99.73%.

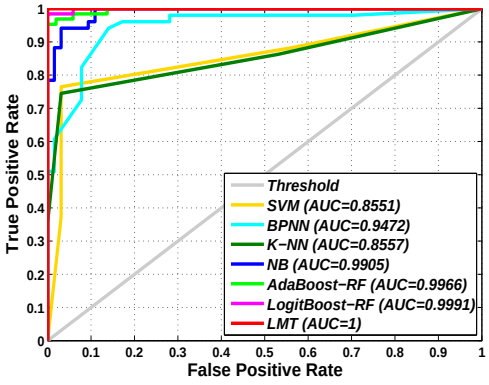
Lastly, a comparative analysis of different performance measures achieved by the proposed work with existing approaches are summarized in Table 6.5. It is clearly observed that the suggested scheme outperforms its competent approaches.



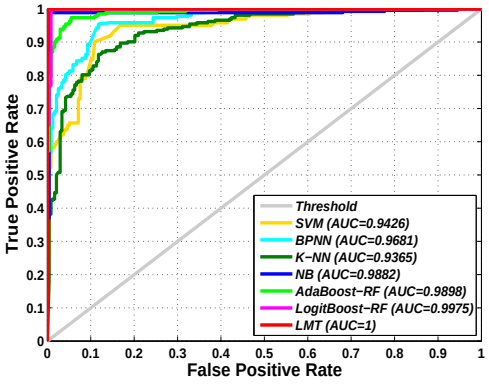
(a) Malignant–normal (MIAS)



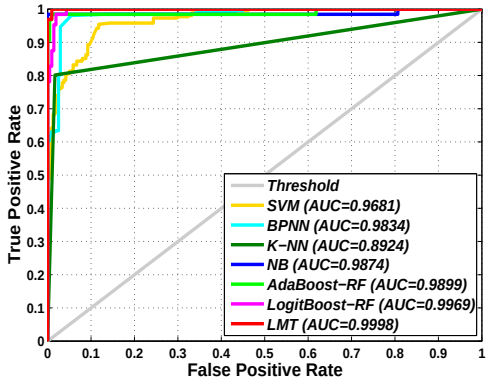
(b) Malignant–normal (DDSM)



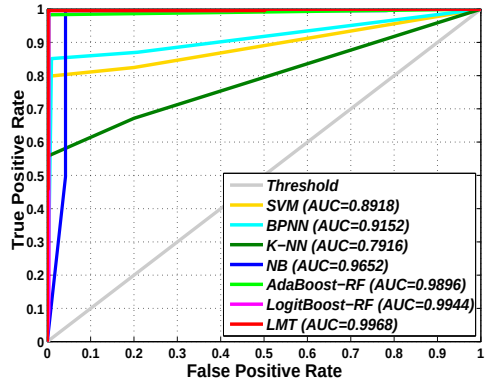
(c) Malignant–benign (MIAS)



(d) Malignant–benign (DDSM)



(e) Benign–normal (MIAS)



(f) Benign–normal class set (DDSM)

Figure 6.5: Comparison of ROC curves obtained by LMT classifier with that of other classifiers at optimal number of relvant features ($R = 170$).

Table 6.2: fold-wise optimal confusion matrices for different databases computed by the LMT classifier.

Class set	Folds	Database											
		MIAS						DDSM					
		Training instances	Testing instances	TP	FP	TN	FN	Training instances	Testing instances	TP	FP	TN	FN
Malignant–normal	Fold 1	232	26	5	0	21	0	687	77	27	0	50	0
	Fold 2	232	26	5	0	21	0	687	77	27	0	50	0
	Fold 3	232	26	5	0	21	0	688	76	26	0	50	0
	Fold 4	232	26	4	0	21	1	688	76	26	0	50	0
	Fold 5	232	26	5	0	21	0	688	76	26	0	50	0
	Fold 6	232	26	5	0	21	0	688	76	26	0	50	0
	Fold 7	232	26	5	0	21	0	688	76	26	0	50	0
	Fold 8	232	26	6	0	20	0	688	76	26	0	50	0
	Fold 9	233	25	5	0	20	0	688	76	26	0	50	0
	Fold 10	233	25	5	0	20	0	688	76	26	1	49	0
Malignant–benign	Fold 1	103	12	6	0	6	0	450	50	27	0	23	0
	Fold 2	103	12	4	0	7	1	450	50	27	0	23	0
	Fold 3	103	12	5	0	7	0	450	50	25	0	24	1
	Fold 4	103	12	5	0	7	0	450	50	25	1	23	1
	Fold 5	103	12	5	0	7	0	450	50	26	1	23	0
	Fold 6	104	11	5	0	6	0	450	50	26	0	24	0
	Fold 7	104	11	5	0	6	0	450	50	26	0	24	0
	Fold 8	104	11	5	0	6	0	450	50	26	0	24	0
	Fold 9	104	11	5	0	6	0	450	50	26	0	24	0
	Fold 10	104	11	5	0	6	0	450	50	26	1	23	0
Benign–normal	Fold 1	243	28	7	0	21	0	662	74	24	0	50	0
	Fold 2	244	27	7	0	20	0	662	74	23	1	50	0
	Fold 3	244	27	7	0	20	0	662	74	24	0	50	0
	Fold 4	244	27	6	0	20	1	662	74	24	0	50	0
	Fold 5	244	27	6	0	21	0	662	74	24	0	50	0
	Fold 6	244	27	6	0	21	0	662	74	24	0	50	0
	Fold 7	244	27	6	0	21	0	662	74	24	0	49	1
	Fold 8	244	27	6	0	21	0	662	74	24	0	50	0
	Fold 9	244	27	6	0	21	0	663	73	23	0	50	0
	Fold 10	244	27	6	0	21	0	663	73	23	0	50	0

Table 6.3: Different performance measures obtained by various classifiers.

Class set	Classifier	Performance measures											
		MIAS						DDSM					
		TPR (%)	FPR (%)	p (%)	r (%)	F_{score}	MCC	TPR (%)	FPR (%)	p (%)	r (%)	F_{score}	MCC
Malignant–normal	NB	82.4	0	100	82.4	0.903	0.888	97.3	2.0	99.6	97.3	0.985	0.977
	K-NN	93.3	10.7	88.8	93.3	0.910	0.825	77.1	4.4	90.2	77.1	0.831	0.758
	BPNN	97.7	4.6	95.9	97.7	0.968	0.932	90.1	0	100	90.1	0.948	0.925
	SVM	96.9	9.2	92.0	96.9	0.944	0.881	88.5	3.0	93.9	88.5	0.912	0.868
	AdaBoost-RF	88.2	0	100	88.2	0.938	0.926	100	12.0	97.8	100	0.989	0.983
	LogitBoost-RF	96.1	0	100	96.1	0.980	0.975	99.6	2.0	99.6	99.6	0.996	0.994
	LMT	**98.0**	**0**	**100**	**98.0**	**0.990**	**0.988**	**100**	**2.0**	**99.6**	**100**	**0.998**	**0.997**
Malignant–benign	NB	94.1	3.1	96.0	94.1	0.950	0.912	97.3	6.3	94.4	97.3	0.959	0.912
	K-NN	74.5	3.1	95.0	74.5	0.835	0.745	90.1	1.93	83.7	90.1	0.868	0.713
	BPNN	94.1	14.1	84.2	94.1	0.889	0.795	95.4	1.18	89.9	95.4	0.926	0.841
	SVM	76.5	3.1	95.1	76.5	0.848	0.761	85.3	8.0	90.6	85.3	0.879	0.776
	AdaBoost-RF	98.0	3.1	96.2	98.0	0.971	0.947	98.5	4.2	96.3	98.5	0.974	0.944
	LogitBoost-RF	98.0	1.6	98.0	98.0	0.980	0.965	98.9	1.3	98.9	98.9	0.989	0.976
	LMT	**98.0**	**0**	**100**	**98.0**	**0.990**	**0.983**	**99.2**	**1.3**	**98.8**	**99.2**	**0.994**	**0.980**
Benign–normal	NB	90.6	1.4	95.1	90.6	0.928	0.907	99.6	4.2	91.9	99.6	0.956	0.935
	K-NN	80.2	1.7	98.1	80.2	0.882	0.792	55.7	0	100	55.7	0.716	0.673
	BPNN	98.1	5.5	95.2	98.1	0.966	0.928	99.0	1.5	92.7	99.0	0.957	0.873
	SVM	95.4	1.18	89.9	95.4	0.926	0.841	99.8	2.02	90.4	99.8	0.949	0.846
	AdaBoost-RF	93.8	0.5	98.4	93.8	0.960	0.948	98.3	0.4	99.2	98.3	0.987	0.981
	LogitBoost-RF	96.9	0	100	96.9	0.984	0.980	99.6	0.4	99.2	99.6	0.994	0.991
	LMT	**98.4**	**0**	**100**	**98.4**	**0.992**	**0.990**	**99.6**	**0.2**	**99.6**	**99.6**	**0.996**	**0.994**

Table 6.4: Values of various evaluation metrics achieved by various classifiers.

Class set	Classifier	Evaluation metric							
		MIAS				DDSM			
		κ	E_{rms}	ACC (%)	AUC	κ	E_{rms}	ACC (%)	AUC
Malignant–normal	NB	0.8822	0.2323	96.51	0.9700	0.9766	0.1025	98.95	0.9856
	K-NN	0.8240	0.2960	91.20	0.9130	0.7530	0.3147	89.23	0.9563
	BPNN	0.9318	0.2121	96.60	0.9661	0.9226	0.1847	96.59	0.9504
	SVM	0.8794	0.2420	94.00	0.9705	0.8673	0.2386	94.09	0.9737
	AdaBoost-RF	0.9233	0.1528	97.67	0.9906	0.9826	0.0887	99.21	0.9940
	LogitBoost-RF	0.9752	0.0880	99.22	0.9928	0.9942	0.0512	99.74	0.9971
	LMT	**0.9877**	**0.0620**	**99.61**	**0.9997**	**0.9971**	**0.1491**	**99.87**	**1**
Malignant–benign	NB	0.9117	0.2070	95.66	0.9905	0.9117	0.1935	95.60	0.9882
	K-NN	0.7298	0.3578	86.97	0.8557	0.7102	0.3595	85.6	0.9365
	BPNN	0.7911	0.2934	89.57	0.9472	0.8391	0.2823	92.00	0.9681
	SVM	0.7484	0.3457	87.83	0.8551	0.7749	0.3101	88.80	0.9426
	AdaBoost-RF	0.9473	0.1523	97.39	0.9966	0.9438	0.1326	97.20	0.9898
	LogitBoost-RF	0.9648	0.1107	98.26	0.9991	0.9759	0.0991	98.80	0.9975
	LMT	**0.9824**	**0.0985**	**99.13**	**1**	**0.9836**	**0.0663**	**99.40**	**1**
Benign–normal	NB	0.9064	0.1451	96.68	0.9874	0.9333	0.1727	97.02	0.9652
	K-NN	0.7775	0.3339	88.80	0.8924	0.6229	0.3896	84.78	0.7916
	BPNN	0.9277	0.2278	96.40	0.9834	0.8680	0.2400	94.23	0.9152
	SVM	0.8391	0.2823	92.00	0.9681	0.8352	0.2658	92.91	0.8918
	AdaBoost-RF	0.9480	0.1368	98.16	0.9899	0.9814	0.0902	99.19	0.9896
	LogitBoost-RF	0.9793	0.0847	99.26	0.9969	0.9907	0.0640	99.59	0.9944
	LMT	**0.9897**	**0.0607**	**99.63**	**0.9998**	**0.9938**	**0.1532**	**99.73**	**0.9968**

Table 6.5: Performance comparison of the proposed work with existing approaches.

Approach	**Technique**	**Database**	**Measures of performance**
Cao *et al.* (2010) [30]	age, intensity, shape, texture, SVM	MIAS	$AUC = 0.948$ (Malignant–benign)
Tahmasbi *et al.* (2011) [34]	Zernike moments, MLP	MIAS	$AUC = 0.976$ (Malignant–benign)
Nanni *et al.* (2012) [41]	LTP, NPE, SVM	DDSM	$AUC = 0.97$ (Malignant–benign)
Nascimento *et al.* (2013) [43]	DWT, polynomial classifier	DDSM	$AUC = 0.98$ (Malignant-normal) $AUC = 0.95$ (benign-normal) $AUC = 0.96$ (Malignant–benign)
Diaz *et al.* (2014) [49]	Spatial, texture, spectral feature, SVM	MIAS	$AUC = 0.976\%$ (Abnormal–normal)
Kim *et al.* (2014) [50]	Stellate feature, AdaBoost, SVM	DDSM	$AUC = 0.956$ (Malignant–benign)
Rouhi *et al.* (2015) [53]	Thresholding, GLCM Zernike moments, GA, ANN	MIAS	$ACC = 96.47\%$ (Malignant–benign)
		DDSM	$ACC = 90.6\%$ (Malignant–benign)
Korkmaz *et al.* (2015) [54]	Texture, mRMR, KL-classifier	DDSM	$ACC = 98.3\%$ (Malignant–benign-normal)
Jiang *et al.* (2015) [55]	SIFT, weighted majority vote	DDSM	$ACC = 90.8\%$ (Abnormal–normal)
Oliveira *et al.* (2015) [59]	Taxonomic indexes, SVM	DDSM	$ACC = 98.88\%$ (Abnormal–normal)
Dhahbi *et al.* (2015) [56]	Curvelet transform, moment theory, t-test, K-NN	MIAS	$ACC = 91.27\%$ (Abnormal–normal) $ACC = 81.35\%$) (Malignant–benign
		DDSM	$ACC = 86.46\%$ (Abnormal–normal) $ACC = 60.43\%$ (Malignant–benign)
Karabatak (2015) [57]	Nine features of each record	Wisconsin database	$ACC = 98.25\%$ (Malignant–benign)
Xie *et al.* (2015) [58]	Gray level & texture features, SVM, ELM	MIAS	$ACC = 96.02\%$, $AUC = 0.9659$ (Malignant–benign)
Proposed scheme	**FRST+t-SNE +LMT**	**MIAS**	$\boldsymbol{ACC = 99.61\%}$, $\boldsymbol{AUC = 0.9997}$ **(Malignant–normal)** $\boldsymbol{ACC = 99.13\%}$, $\boldsymbol{AUC = 1}$ **(Malignant–benign)** $\boldsymbol{ACC = 99.63\%}$, $\boldsymbol{AUC = 0.9998}$ **(Benign–normal)**
		DDSM	$\boldsymbol{ACC = 99.87\%}$, $\boldsymbol{AUC = 1}$ **(Malignant–normal)** $\boldsymbol{ACC = 99.40\%}$, $\boldsymbol{AUC = 1}$ **(Malignant–benign)** $\boldsymbol{ACC = 99.73\%}$, $\boldsymbol{AUC = 0.9968}$ **(Benign–normal)**

6.5 Summary

In this chapter, an efficient mammogram classification scheme has been proposed to help radiologists for the identification of suspicious mammographic tissues in early diagnosis of breast cancer. The scheme utilizes the radial symmetric features produced by the Fast Radial Symmetry Transform (FRST) of mammographic ROIs. The subset of relevant features is chosen by the use of the t-distributed Stochastic Neighbor Embedding (t-SNE) method. Various classifier like Naive Bayes (NB), K-Nearest Neighbor (K-NN), Back-Propagation Neural Network (BPNN), Support Vector Machine (SVM), AdaBoost and Random Forest (AdaBoost-RF), LogitBoost and Random Forest (LogitBoost-RF), and Logistic Model Tree (LMT) have been employed to classify the mammographic ROIs into malignant–normal, malignant–benign, and benign–normal classes. The proposed LMT classifier achieves the optimal accuracies (ACCs) of 99.61%, 99.13%, and 99.63% for malignant–normal, malignant–benign, and benign–normal class sets respectively using MIAS database . The similar parameters are 99.87%, 99.40%, and 99.73% for DDSM database. In general, it has been observed that the suggested FRST features with LMT classifier shows superior performance as compared to others.

Chapter 7

Conclusions and Future Work

In this thesis we suggest five new features for the characterization of mammograms in the diagnosis of breast cancer. In addition to the development of feature, efficient feature reduction schemes along with improved classifiers are implemented to increase classification accuracy. The suggested contributions are distributed in five chapters with the corresponding simulation results.

In Chapter 2, a mammogram classification scheme has been proposed to support the decision of radiologists. The scheme utilizes 2D-DWT and GLCM in succession to derive feature matrix form mammograms. To select relevant features from the feature matrix, both *t-test* and *F-test* have been applied independently along with random forest method. Chapter 3 suggests a scheme that utilizes SFTA method to extract the features from the digital mammograms. An effective feature selection technique FCBF has been used to select the most significant feature set from the extracted features. Chapter 4 deals with the proposed mammogram classification in which, 2D-DOST is utilized to extract features from the mammographic images. A null-hypothesis technique using the *two-sample t-test* is used to select the most discriminant features from high dimensional feature space. In Chapter 5, a proficient mammogram classification scheme has been suggested that applies 2D-SLT for the extraction of features from mammographic ROIs. The most discriminatory feature elements are selected by the use of the Bayesian Logistic Regression (BLogR) method. The relevant features are balanced by the Gaussian distribution based balancing method to achieve an improved classification performance. Another mammogram classification scheme has been proposed in Chapter 6 based on radial symmetric features produced by Fast Radial Symmetry Transform (FRST) of mammograms. The

useful subset of feature elements is chosen by the use of the t-distributed Stochastic Neighbor Embedding (t-SNE) method.

To validate the efficacy of the suggested features and feature reduction techniques, the relevant features so obtained are applied to several classifiers namely, Naive Bayes (NB), K-Nearest Neighbor (K-NN), Back-Propagation Neural Network (BPNN), Support Vector Machine (SVM), AdaBoost and Random Forest (AdaBoost-RF), LogitBoost and Random Forest (LogitBoost-RF), and Logistic Model Tree (LMT).

The training and testing samples are kept similar in all classifiers. Further, other existing features are also used in all classifiers under the similar scenario. The suggested features are compared with existing features in each case and an overall inference is drawn with respect to classifier's accuracy performance. For each feature, the corresponding classifier which outperforms others is considered to be winner and elaborated in detail. Finally, an overall comparison has been made among our suggested schemes and is shown in Table 7.1.

Table 7.1: Classification Performance comparison between the proposed schemes and existing approaches.

Approach	**Technique**	**Database**	**Classification accuracy (ACC) (%)**
Proposed scheme 1 [Chapter 2]	DWT +GLCM +*t-test* +BPNN	MIAS	$ACC = 98.13\%$, (Abnormal–normal) $ACC = 94.20\%$ (Malignant–benign)
		DDSM	$ACC = 98.80\%$ (Abnormal–normal) $ACC = 97.40\%$ (Malignant–benign)
Proposed scheme 2 [Chapter 3]	SFTA +FCBF +SVM	MIAS	$ACC = 98.76\%$, (abnormal–normal) $ACC = 95.65\%$, (Malignant–benign)
		DDSM	$ACC = 99.20\%$, (Abnormal–normal) $ACC = 98.00\%$ (Malignant–benign)
Proposed scheme 3 [Chapter 4]	DOST +null-hypothesis +AdaBoost-RF	MIAS	$ACC = 98.75\%$, (Abnormal–normal) $ACC = 98.26\%$ (Malignant–benign)
		DDSM	$ACC = 99.30\%$ (Abnormal–normal) $ACC = 98.80\%$ (Malignant–benign)
Proposed scheme 4 [Chapter 5]	SLT +BLogR +LogitBoost-RF	MIAS	$ACC = 99.69\%$ (Abnormal–normal) $ACC = 99.13\%$ (Malignant–benign)
		DDSM	$ACC = 99.80\%$ (Abnormal–normal) $ACC = 99.40\%$ (Malignant–benign)
Proposed scheme 5 [Chapter 6]	FRST +t-SNE +LMT	MIAS	$ACC = 99.61\%$ (Malignant–normal) $ACC = 99.13\%$ (Malignant–benign) $ACC = 99.63\%$ (Benign–normal)
		DDSM	$ACC = 99.87\%$ (Malignant–normal) $ACC = 99.40\%$ (Malignant–benign) $ACC = 99.73\%$ (Benign–normal)

It may be observed that, the performances achieved by all the schemes are very close to each other. The proposed schemes 4 and 5 perform the same best result on malignant-benign class set. However, the proposed scheme 5 (FRST + t-SNE + LMT) achieves optimal results on the different class sets.

The objective of each scheme suggested in this thesis is to classify the types of suspicious breast tumor as cancerous or non-cancerous. The thesis does not cover on prediction of different types of cancer if the suspicious tissue is found cancerous. Depending on the morphology, there are different types of masses namely, round, oval, lobulated, stellate etc. present in the breast. The common types of microcalcification such as ring shaped, punctuate, linear, needle shaped, coarsely granular are seen on the mammographic images. The detection of thesis masses and its type also an important task in the CAD of mammograms which are limitations of this thesis. In future work, we have planned to carry out this diagnosis by designing the suitable image segmentation methods. The next thrust will be given to investigate more feature extraction, selection, and classification techniques. The existing scheme will be validated on the real time scenario in the nearby hospital, where mammogram facilities are available. To initiate the process an understanding has been made in Ispat General Hospital (IGH), Rourkela. Further, at present the ROIs are extracted manually which needs to be automated using suitable techniques. We are looking for a suitable enhancement scheme to improve the mammogram quality so that the features extracted are more accurate. Our final thrust will be to generate a large database by collecting mammograms with their classification labels from different health centers so that validation of schemes will be trustworthy and be accepted by the practitioners.

Bibliography

[1] T. M. Deserno, "Fundamentals of Biomedical Image Processing," in *Proc. Biomedical Image Processing*. Springer, 2010, pp. 1 – 51.

[2] H. Ostensen, "Diagnostic imaging: What is it? When and how to use it where resources are limited?" *World Health Organization (WHO)*, pp. 1 – 31, 2001.

[3] "Latest world cancer statistics, International Agency for Research on Cancer (IARC) press releases 2013, World Health Organisation (WHO)," http://www.iarc.fr/en/media-centre/pr/2013/index.php.

[4] "Globocan project 2012, International Agency for Research on Cancer (IARC), World Health Organisation: cancer fact sheets," http://globocan.iarc.fr.

[5] "Statistics of breast cancer in India," http://www.breastcancerindia.net/statistics/stat_global.html.

[6] S. V. Engeland, *Detection of mass lesions in mammograms by using multiple views*. [Sl: sn], 2006.

[7] L. Tabar and P. Dean, "Mammography and breast cancer: the new era," *International Journal of Gynecology & Obstetrics*, vol. 82, no. 3, pp. 319 – 326, 2003.

[8] H. Cheng, X. Shi, R. Min, L. Hu, X. Cai, and H. Du, "Approaches for automated detection and classification of masses in mammograms," *Pattern Recognition*, vol. 39, no. 4, pp. 646 – 668, 2006.

[9] R. Kohavi and F. Provost, "Glossary of terms," *Machine Learning*, vol. 30, no. 2 - 3, pp. 271 – 274, 1998.

[10] T. Pang-Ning, M. Steinbach, and V. Kumar, "Introduction to data mining," in *Proc. Library of Congress*, 2006, p. 74.

[11] "The mini-MIAS database of mammograms," http://peipa.essex.ac.uk/info/mias.html.

[12] T. M. Deserno, M. Soiron, and J. E. de Oliveira, "Texture patterns extracted from digitizes mammograms of different BI-RADS classes, Image Retrieval in Medical Applications project, release: V1.0, 2012," http://ganymed.imib.rwth-aachen.de/irma/datasets_en.php.

[13] A. P. Dhawan, Y. Chitre, C. Kaiser-Bonasso, and M. Moskowitz, "Analysis of mammographic microcalcifications using gray-level image structure features," *IEEE Transactions on Medical Imaging*, vol. 15, no. 3, pp. 246 – 259, 1996.

[14] D. Wei, H.-P. Chan, N. Petrick, B. Sahiner, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "False-positive reduction technique for detection of masses on digital mammograms: Global and local multiresolution texture analysis," *Medical Physics*, vol. 24, no. 6, pp. 903 – 914, 1997.

[15] S. Liu, C. F. Babbs, and E. J. Delp, "Multiresolution detection of spiculated lesions in digital mammograms," *IEEE Transactions on Image Processing*, vol. 10, no. 6, pp. 874 – 884, 2001.

[16] R. J. Ferrari, R. M. Rangayyan, J. L. Desautels, and A. F. Frère, "Analysis of asymmetry in mammograms via directional filtering with gabor wavelets," *IEEE Transactions on Medical Imaging*, vol. 20, no. 9, pp. 953 – 964, 2001.

[17] L. Zhen and A. K. Chan, "An artificial intelligent algorithm for tumor detection in screening mammogram," *IEEE Transactions on Medical Imaging*, vol. 20, no. 7, pp. 559 – 567, 2001.

[18] M. Masotti, "A ranklet-based image representation for mass classification in digital mammograms," *Medical Physics*, vol. 33, no. 10, pp. 3951 – 3961, 2006.

[19] M. E. Mavroforakis, H. V. Georgiou, N. Dimitropoulos, D. Cavouras, and S. Theodoridis, "Mammographic masses characterization based on localized texture and dataset fractal analysis using linear, neural and support vector machine classifiers," *Artificial Intelligence in Medicine*, vol. 37, no. 2, pp. 145 – 162, 2006.

[20] L. d. O. Martins, A. M. d. Santos, A. C. Silva, and A. C. Paiva, "Classification of normal, benign and malignant tissues using co-occurrence matrix and bayesian neural network in mammographic images," in *Proc. Ninth Brazilian Symposium on Neural Networks*. IEEE, 2006, pp. 24 – 29.

[21] F. Sakellaropoulos, S. Skiadopoulos, A. Karahaliou, G. Panayiotakis, and L. Costaridou, "Wavelet-based feature analysis for classification of breast masses from normal dense tissue," *Artificial Intelligence Applications and Innovations*, vol. 204, pp. 722 – 729, 2006.

[22] E. A. Rashed, I. A. Ismail, and S. I. Zaki, "Multiresolution mammogram analysis in multilevel decomposition," *Pattern Recognition Letters*, vol. 28, no. 2, pp. 286 – 292, 2007.

[23] R. R. Pereira Jr, P. M. A. Marques, M. O. Honda, S. K. Kinoshita, R. Engelmann, C. Muramatsu, and K. Doi, "Usefulness of texture analysis for computerized classification of breast lesions on mammograms," *Journal of Digital Imaging*, vol. 20, no. 3, pp. 248 – 255, 2007.

[24] A. Khademi, F. Sahba, A. Venetsanopoulos, and S. Krishnan, "Region, lesion and border-based multiresolution analysis of mammogram lesions," in *Image Analysis and Recognition*. Springer, 2009, pp. 802 – 813.

[25] A. Dong and B. Wang, "Feature selection and analysis on mammogram classification," in *Proc. IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*. IEEE, 2009, pp. 731 – 735.

[26] S. Dua, H. Singh, and H. W. Thompson, "Associative classification of mammograms using weighted rules," *Expert Systems with Applications*, vol. 36, no. 5, pp. 9250 – 9259, 2009.

[27] B. Prathibha and V. Sadasivam, "Breast tissue characterization using variants of nearest neighbour classifier in multi texture domain," *IE (I) Journal*, vol. 91, pp. 7 – 13, 2010.

[28] B. Verma, P. McLeod, and A. Klevansky, "Classification of benign and malignant patterns in digital mammograms for the diagnosis of breast cancer," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3344 – 3351, 2010.

[29] F. Moayedi, Z. Azimifar, R. Boostani, and S. Katebi, "Contourlet-based mammography mass classification using the svm family," *Computers in Biology and Medicine*, vol. 40, no. 4, pp. 373 – 383, 2010.

[30] Y. Cao, X. Hao, X. Zhu, and S. Xia, "An adaptive region growing algorithm for breast masses in mammograms," *Frontiers of Electrical and Electronic Engineering in China*, vol. 5, no. 2, pp. 128 – 136, 2010.

[31] I. Buciu and A. Gacsadi, "Directional features for automatic tumor classification of mammogram images," *Biomedical Signal Processing and Control*, vol. 6, no. 4, pp. 370 – 378, 2011.

[32] M. A. Al Mutaz, S. Dress, and N. Zaki, "Detection of masses in digital mammogram using second order statistics and artificial neural network," *International Journal of Computer Science & Information Technology*, vol. 3, no. 3, pp. 176 – 186, 2011.

[33] M. Fraschini, "Mammographic masses classification: novel and simple signal analysis method," *Electronics Letters*, vol. 47, no. 1, pp. 14 – 15, 2011.

[34] A. Tahmasbi, F. Saki, and S. B. Shokouhi, "Classification of benign and malignant masses based on zernike moments," *Computers in Biology and Medicine*, vol. 41, no. 8, pp. 726 – 735, 2011.

[35] S. K. Biswas and D. P. Mukherjee, "Recognizing architectural distortion in mammogram: a multiscale texture modeling approach with GMM," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 7, pp. 2023 – 2030, 2011.

[36] N.-C. Tsai, H.-W. Chen, and S.-L. Hsu, "Computer-aided diagnosis for early-stage breast cancer by using wavelet transform," *Computerized Medical Imaging and Graphics*, vol. 35, no. 1, pp. 1 – 8, 2011.

[37] J. Jona and N. Nagaveni, "A hybrid swarm optimization approach for feature set reduction in digital mammograms," *WSEAS Transactions on Information Science and Applications*, vol. 9, pp. 340 – 349, 2012.

[38] R. P. Ramos, M. Z. do Nascimento, and D. C. Pereira, "Texture extraction: An evaluation of ridgelet, wavelet and co-occurrence based methods applied to mammograms," *Expert Systems with Applications*, vol. 39, no. 12, pp. 11 036 – 11 047, 2012.

[39] M. M. Eltoukhy, I. Faye, and B. B. Samir, "A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation," *Computers in Biology and Medicine*, vol. 42, no. 1, pp. 123 – 128, 2012.

[40] M. Muštra, M. Grgić, and K. Delač, "Breast density classification using multiple feature selection," *AUTOMATIKA: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije*, vol. 53, no. 4, pp. 362 – 372, 2012.

[41] L. Nanni, S. Brahnam, and A. Lumini, "A very high performing system to discriminate tissues in mammograms as benign and malignant," *Expert Systems with Applications*, vol. 39, no. 2, pp. 1968 – 1971, 2012.

[42] P. Görgel, A. Sertbas, and O. N. Ucan, "Mammographical mass detection and classification using local seed region growing–spherical wavelet transform (lsrg–swt) hybrid scheme," *Computers in Biology and Medicine*, vol. 43, no. 6, pp. 765 – 774, 2013.

[43] M. Z. do Nascimento, A. S. Martins, L. A. Neves, R. P. Ramos, E. L. Flores, and G. A. Carrijo, "Classification of masses in mammographic image using wavelet domain features and polynomial classifier," *Expert Systems with Applications*, vol. 40, no. 15, pp. 6213 – 6221, 2013.

[44] S. M. Kumar and G. Balakrishnan, "Multi resolution analysis for mass classification in digital mammogram using stochastic neighbor embedding," in *Proc. International Conference on Communications and Signal Processing.* IEEE, 2013, pp. 101 – 105.

[45] C. Oral and H. Sezgin, "Effects of dimension reduction in mammograms classification," in *Proc. 8th International Conference on Electrical and Electronics Engineering.* IEEE, 2013, pp. 630 – 633.

[46] X. Liu and J. Tang, "Mass classification in mammograms using selected geometry and texture features, and a new SVM-based feature selection method," *IEEE Systems Journal*, vol. 8, no. 3, pp. 910 – 920, 2014.

[47] K. Ganesan, U. R. Acharya, C. K. Chua, C. M. Lim, and K. T. Abraham, "One-class classification of mammograms using trace transform functionals," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 2, pp. 304 – 311, 2014.

[48] Y. A. Reyad, M. A. Berbar, and M. Hussain, "Comparison of Statistical, LBP, and Multi-Resolution Analysis Features for Breast Mass Classification," *Journal of Medical Systems*, vol. 38, no. 9, pp. 1 – 15, 2014.

[49] C. Diaz-Huerta, E. Felipe-Riveron, and L. Montaño-Zetina, "Quantitative analysis of morphological techniques for automatic classification of micro-calcifications in digitized mammograms," *Expert Systems with Applications*, vol. 41, no. 16, pp. 7361 – 7369, 2014.

[50] D. H. Kim, J. Y. Choi, and Y. M. Ro, "Region based stellate features combined with variable selection using AdaBoost learning in mammographic computer-aided detection," *Computers in Biology and Medicine*, vol. 63, pp. 238 – 250, 2014.

[51] P. Görgel, A. Sertbas, and O. N. Uçan, "Computer-aided classification of breast masses in mammogram images based on spherical wavelet transform and support vector machines," *Expert Systems*, vol. 32, no. 1, pp. 155 – 164, 2015.

[52] Y. Li, H. Chen, G. K. Rohde, C. Yao, and L. Cheng, "Texton analysis for mass classification in mammograms," *Pattern Recognition Letters*, vol. 52, pp. 87 – 93, 2015.

[53] R. Rouhi, M. Jafari, S. Kasaei, and P. Keshavarzian, "Benign and malignant breast tumors classification based on region growing and CNN segmentation," *Expert Systems with Applications*, vol. 42, no. 3, pp. 990 – 1002, 2015.

[54] S. A. Korkmaz and M. F. Korkmaz, "A new method based cancer detection in mammogram textures by finding feature weights and using kullback–leibler measure with kernel estimation," *Optik-International Journal for Light and Electron Optics*, vol. 126, no. 20, pp. 2576 – 2583, 2015.

[55] M. Jiang, S. Zhang, H. Li, and D. N. Metaxas, "Computer-aided diagnosis of mammographic masses using scalable image retrieval," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, pp. 783 – 792, 2015.

[56] S. Dhahbi, W. Barhoumi, and E. Zagrouba, "Breast cancer diagnosis in digitized mammograms using curvelet moments," *Computers in Biology and Medicine*, vol. 64, pp. 79 – 90, 2015.

[57] M. Karabatak, "A new classifier for breast cancer detection based on naïve bayesian," *Measurement*, vol. 72, pp. 32 – 36, 2015.

[58] W. Xie, Y. Li, and Y. Ma, "Breast mass classification in digital mammography based on extreme learning machine," *Neurocomputing*, vol. 173, pp. 930 – 941, 2015.

[59] F. S. S. de Oliveira, A. O. de Carvalho Filho, A. C. Silva, A. C. de Paiva, and M. Gattass, "Classification of breast regions as mass and non-mass based on digital mammograms using taxonomic indexes and svm," *Computers in Biology and Medicine*, vol. 57, pp. 42 – 53, 2015.

[60] Y.-D. Zhang, S.-H. Wang, G. Liu, and J. Yang, "Computer-aided diagnosis of abnormal breasts in mammogram images by weighted-type fractional Fourier transform," *Advances in Mechanical Engineering*, vol. 8, no. 2, pp. 1 – 11, 2016.

[61] "SVM Fundamentals," http://www.vlfeat.org/api/svm-fundamentals.html#svm-feature-maps.

[62] Y.-D. Zhang, S. Chen, S.-H. Wang, J.-F. Yang, and P. Phillips, "Magnetic resonance brain image classification based on weighted-type fractional Fourier transform and nonparallel support vector machine," *International Journal of Imaging Systems and Technology*, vol. 25, no. 4, pp. 317 – 327, 2015.

[63] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors)," *The Annals of Statistics*, vol. 28, no. 2, pp. 337 – 407, 2000.

[64] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. International Conference on Machine Learning (ICML)*, vol. 96, 1996, pp. 148 – 156.

[65] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5 – 32, 2001.

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

[67] N. Landwehr, M. Hall, and E. Frank, "Logistic Model Trees," *Machine Learning*, vol. 59, no. 1-2, pp. 161 – 205, 2005.

[68] P. Doetsch *et al.*, "Logistic Model Trees with AUC Split Criterion for the KDD Cup 2009 Small Challenge," in *Proc. Conference on Knowledge Discovery and Data Mining (KDD) Cup*, 2009, pp. 77 – 88.

[69] J. R. Quinlan, *C4.5: Programs for Machine Learning.* Elsevier, 2014.

[70] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees.* CRC press, 1984.

[71] C. Gonzalez and E. Woods, *Digital Image Processing.*

[72] S. Mallat, *A wavelet tour of signal processing.* Access Online via Elsevier, 1999.

[73] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, no. 6, pp. 610 – 621, 1973.

[74] F. Albregtsen *et al.*, "Statistical texture measures computed from gray level coocurrence matrices," *Image Processing Laboratory, Department of Informatics, University of Oslo*, pp. 1 – 14, 2008.

[75] A. F. Costa, G. Humpire-Mamani, and A. J. M. Traina, "An efficient algorithm for fractal analysis of textures," in *Proc. 25th Conference on Graphics, Patterns and Images (SIBGRAPI).* IEEE, 2012, pp. 39 – 46.

[76] A. R. Backes and O. M. Bruno, "A new approach to estimate fractal dimension of texture images," in *Image and Signal Processing.* Springer, 2008, pp. 136 – 143.

[77] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in *Proc. International Conference on Machine Learning (ICML)*, vol. 3, 2003, pp. 856 – 863.

[78] S. Drabycz, R. G. Stockwell, and J. R. Mitchell, "Image texture characterization using the discrete orthonormal S-transform," *Journal of Digital Imaging*, vol. 22, no. 6, pp. 696 – 708, 2009.

[79] D. F. Groebner, P. W. Shannon, P. C. Fry, and K. D. Smith, *Business Statistics: A Decision Making Approach.* Prentice Hall/Pearson, 2011.

[80] S. M. Pizer *et al.*, "Contrast-limited adaptive histogram equalization: speed and effectiveness," in *Proc. First Conference on Visualization in Biomedical Computing*. IEEE, 1990, pp. 337 – 345.

[81] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization for real-time image enhancement," *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology*, vol. 38, no. 1, pp. 35 – 44, 2004.

[82] I. W. Selesnick, "The slantlet transform," *IEEE Transactions on Signal Processing*, vol. 47, no. 5, pp. 1304 – 1313, 1999.

[83] G. C. Cawley and N. L. Talbot, "Gene selection in cancer classification using sparse logistic regression with Bayesian regularization," *Bioinformatics*, vol. 22, no. 19, pp. 2348 – 2355, 2006.

[84] S. K. Shevade and S. S. Keerthi, "A simple and efficient algorithm for gene selection using sparse logistic regression," *Bioinformatics*, vol. 19, no. 17, pp. 2246 – 2253, 2003.

[85] P. Bermejo, J. A. Gamez, J. M. Puerta, and R. Uribe, "Improving kNN-based e-mail classification into folders generating class-balanced datasets," in *Proc. Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, 2008, pp. 529 – 536.

[86] P. Bermejo, J. A. Gámez, and J. M. Puerta, "Improving the performance of Naive Bayes multinomial in e-mail foldering by introducing distribution-based balance of datasets," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2072 – 2080, 2011.

[87] J. Rice, *Mathematical Statistics and Data Analysis*. Cengage Learning, 2006.

[88] C. DeSantis, R. Siegel, and A. Jemal, "Breast cancer facts and figures 2013-2014," *American Cancer Society*, pp. 1 – 38, 2013.

[89] G. Loy and A. Zelinsky, "Fast radial symmetry for detecting points of interest," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 959 – 973, 2003.

[90] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. 85, pp. 2579 – 2605, 2008.

Dissemination

Published

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

2. S. Beura, B. Majhi, R. Dash and S. Roy, "Classification of mammogram using two-dimensional discrete orthonormal S-transform for breast cancer detection", *Healthcare Technology Letters, IET*, vol. 2, no. 2, pp. 46 – 51, 2015.

3. S. Beura, B. Majhi and R. Dash, "Automatic Characterization of Mammograms using Fractal Texture Analysis and Fast Correlation Based Filter Method", *In Proc. Second International Conference on Perception and Machine Intelligence, ACM, C-DAC, Kolkata*, 2015, pp. 85 – 91.

4. S. Beura, B. Majhi and R. Dash, "Identification of Abnormality in Mammograms using Multiresolution Analysis and Artificial Neural Network", *In Proc. Eighth International Conference on Image and Signal Processing, Elsevier, Bangalore*, 2014, pp. 171 – 179.

Shradhananda Beura

Department of Computer Science and Engineering,
National Institute of Technology Rourkela,
Rourkela – 769 008, Odisha, India.
+91 90408 38168

beura.shradhananda@gmail.com

Qualification

- Ph.D. (CSE) (*Continuing*)
  National Institute of Technology Rourkela
- M.Tech (CSE)
  Biju Patnaik University of Technology Rourkela
- B. Tech (IT)
  Biju Patnaik University of Technology Rourkela

Publications

- Journals: 02
- Conferences: 02

Permanent Address

Oliha, PO - Deulapada,
Via - Sri Baladev Jew
Kendrapada 754 212, Odisha.

Date of Birth

10th June 1983