CLASSIFICATION OF BREAST LESIONS USING FEATURE EXTRACTION TECHNIQUES

Dissertation submitted in partial fulfillment of the requirements for the degree of

MASTER OF TECHNOLOGY
IN
ELECTRONICS & COMMUNICATION ENGINEERING

By

Sahil Bhusri

Enrollment No. 142003

Under the Supervision of

Dr. Shruti Jain

DEPARTMENT OF ELECTRONICS AND COMMUNICATIOENGINEERING
JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY
WAKNAGHAT, SOLAN - 173234, INDIA
MAY-2016

TABLE OF CONTENTS

DECLARATION BY CANDIDATE

I hereby declare that the work reported in the M-Tech thesis entitled titled **"Classification of Breast Lesions using Feature Extraction Techniques"** submitted at **Jaypee University of Information Technology, Waknaghat-India** is an authentic record of my work carried out under the supervision carried under the guidance of **Dr. Shruti Jain.** I have not submitted this work elsewhere for any other degree or diploma.

Sahil Bhusri

Department of ECE

Jaypee University of Information Technology, Waknaghat , India

Date : 25 May 2016

SUPERVISOR'S CERTIFICATE

This is to certify that the work reported in the M-Tech. thesis entitled "**Classification of Breast Lesions using Texture Feature Techniques"** submitted by **Sahil Bhusri** at **Jaypee University of Information Technology, Waknaghat**, **India** is a bonafide record of his original work carried out under my supervision. This work has not been submitted elsewhere for any other degree or diploma.

.

Dr. Shruti Jain

Assistant Professor

Department of ECE

Jaypee University of Information Technology, Waknaghat , India

 Date : 25 May 2016

ACKNOWLEDGMENT

Sahil Bhusri

Date : 25 May 2016

ABSTRACT

The most common form of cancer being diagnosed in women worldwide is breast cancer. A region endures from damage through any disease then region is known as lesion. Therefore, characterization of breast lesions is clinically significant. Therefore, there is a significant impetus among the research community to develop computer aided diagnostic (CAD) systems for differential diagnosis between different cases of breast lesions.

Thus, in order to provide the radiologists with a second opinion tool for validating their diagnosis, various CAD systems have been developed in the present work for two-class breast lesions classification.

For the design of this CAD system, ultrasound images are taken and for each ultrasound image, ROI is marked according to the shape of abnormality. The CAD system consists of input ultrasound images, ROI extraction module, feature extraction module and the classification module. In the feature extraction module, four methods for extracting the features are employed, (*a*) Morphological methods (*b*) Signal processing based method (*c*) Transform domain based methods and *(d)* Statistical based methods.

In the classification module, performance of four different classifiers namely k-NN, probabilistic neural network (PNN), SVM and smooth support vector machine (SSVM) is evaluated to obtain the class of the unknown testing instances.

The classification accuracy coming out for two class breast lesions is coming up of 97.1 % with Laws' Mask of length

LIST OF ABBREVIATIONS

ASM	Angular Second Moment
BI-RADS	Breast Imaging Reporting and Data System
CAD	Computer Aided Diagnosis
CM	Confusion Matrix
DWT	Discrete Wavelet Transform
FV	Feature Vector
FOS	First Order Statistics
FPS	Fourier Power Spectrum
GLDS	Gray Level Difference Statistics
GLCM	Gray Level Co-occurrence Matrix
GLRLM	Gray Level Run Length Matrix
GWT	Gabor Wavelet Transform
ICA	Individual Class Accuracy
IAOI	Inner Area of Interest
k-NN	k-Nearest Neighbor
NGTDM	Neighborhood Gray Tone Difference Matrix
OAOI	Outer Area of Interest
OCA	Overall Classification Accuracy
PC	Principal Component
PCA	Principal Component Analysis
PNN	Probabilistic Neural Network
ROIs	Regions of Interest
SFM	Statistical Feature Matrix
SSVM	Smooth Support Vector Machine
SVM	Support Vector Machine
TEM	Texture Energy Measures
TFV	Texture Feature Vector
WPT	Wavelet Packet Transform

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

1.1 Overview

Human body is made up of million numbers of cells which group together to make a tissue or organ. Different types of tissues are present in different parts of body. Either these cells are replaced or reproduced at a regular interval of time but whenever uncontrolled division of cells is present, it leads to cancer [1].The malignant tumor can travels to other parts of the body to form new tumors and destroy other healthy tissues in the body. Metastasis is the process of invasion and destruction of healthy tissues [2]. Breast cancer is the type of cancer that develops form breast cells. It is the most common form of cancer found in women [3]. For the women in United Kingdom, risk of being diagnosed with breast cancer is 1 in 8 [4]. According to the Survey figures by American Cancer Society, Surveillance and Health Services Research in United States, total number of breast cancer cases estimated to occur in 2015 are 2,92,130 and total no of deaths estimated are 40,290 [5].

It has been strongly advocated by many researchers in their study that increased breast density is strongly correlated to the risk of developing breast cancer [6-7]. Even though breast cancer is considered to be a fatal disease but the chances of survival are significantly improved if detected at an early stage. There are various imaging modalities like ultrasound, MRI, Computerized tomography, etc. that can be used for diagnosing the breast diseases but Mammography has limitations in detecting cancer in dense tissues of young patients. Mammography has high false report low conspicuity, and noisy nature of images. Ultrasound is having benefits of low cost, requiring no ionizing radiation, portable and very effective for women younger in age having dense breast tissue [8-10]

1.2 Ultrasonography

Ultrasound is a painless procedure that uses sound waves to make images of the inside of your body. The echoes of sound waves bouncing back from the surface of body are recorded and transfigured into photographs or videos. It's often used in addition to mammography to tell whether a lump in a breast is a cyst or a solid mass, which might be cancer. Ultrasound can also help pin point the position of a tumor. It is useful for doctor to insert a needle to an exact place during a biopsy [11].

There are two types of breast cancers: *a*) Benign *b*) Malignant

a) *Benign Type*

This is the case when the cancer is still inside its place and not broken out.

b) *Malignant Type*

The cancer cells break out from inside the lobules or ducts and invade near by tissues or eventually ,make their way to other organs say bones , lungs ,liver.

To describe the findings on the ultrasounds, The American College of Radiology gave a standard system called Breast Imaging Reporting and Data System (BI-RADS). The categories were [12,13]:

(*i*) *Category 0*: It may not be clearly noticeable and more tests are needed.

(*ii*) *Category 1*: Negative- No abnormalities found to report.

(*iii*) *Category 2*: Benign finding- The finding in the ultrasound is non-cancerous like lymph nodes.

(*iv*) *Category 3*: Probably benign finding- The finding is most probably non-cancerous but is expected to change over time so a follow up is required regularly.

(*v*) *Category 4*: Suspicious abnormality- The findings might or might not be cancerous so to find the exact nature of the finding, the patient should consider taking a biopsy test.

(*vi*) *Category 5*: Highly suggestive malignancy- The finding has more than 95 % chance of being cancerous and biopsy examination is highly recommended for the patient.

(*vii*) *Category 6*: Known biopsy-proven malignancy- The findings on the ultrasound have been shown to be cancerous by a previous biopsy.

Breast lesions are further classified into the following categories:
(*a*) *Benign (B) / Malignant (D) (Two-class classification)*

(*b*) *Primary Benign (PB) / Primary Malignant (PM) / Secondary Malignant (SM) (Three-class classification)*

The benign type cancer has wider than taller shape, intense and uniform hyper echogenicity, an oval shape with a thin, consistent capsule and have two to three gentle lobulations where Malignant type has a marked hypo echogenicity , acoustic shadowing, microlobulation , a duct extension, a 'taller than wide' shape and angular margins.

1.3   Need of Computer Aided Diagnostic Systems

With the advancement in computer technology, there has been an increase in the opportunities for researchers to investigate the capability of computer-aided diagnostic (CAD) systems for texture analysis and lesions characterization in ultrasound images. The overall aim of developing a computerized tissue characterization system is to provide additional diagnostic information about the lesions which cannot be captured by visual inspection of medical images. Fig1 shows the ultrasound cases of three class classification of breast lesions
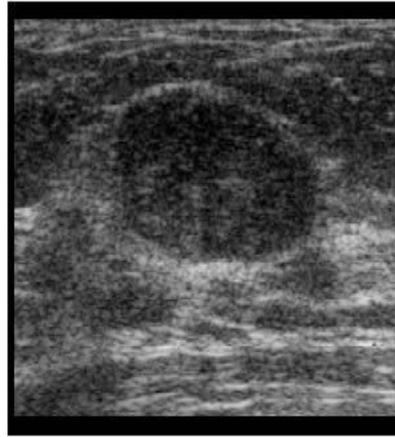
Figure 1.1 (a)   Figure 1.1 (b)   Figure 1.1 (c)
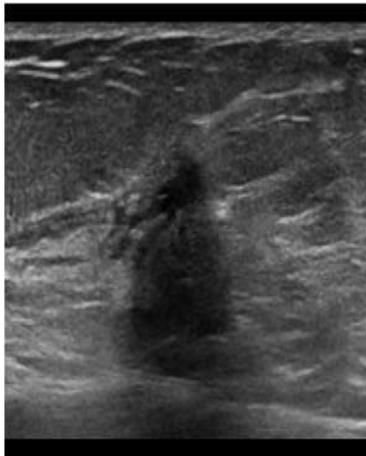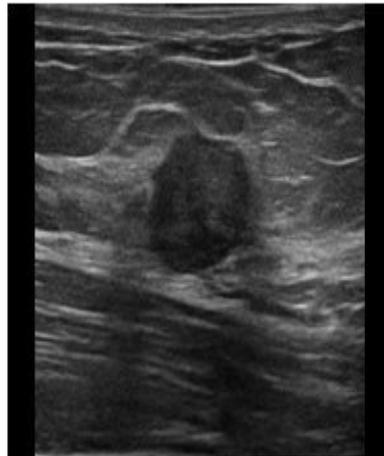
Figure 1.1 (d)   Figure 1.1 (e)   Figure 1.1 (f)

Figure 1.1: (a-b) : Primary Benign (c-d) : Primary Malignant (e-f) : Secondary Malignant

1.4   Objective of the Present Study

The main objective of the research work presented in this project report is to enhance the diagnostic potential of ultrasound images for identification of different breast lesions by developing efficient CAD system designs using a representative image database. The various research objectives formulated for the present work are described below:

(*i*) *The collection of a comprehensive and representative database*: To develop efficient CAD systems, it is necessary that the classifiers used in the classification module of the CAD system are trained with an image database that contains representative images from each subclass. Thus collection of a comprehensive database containing

4

representative images for different breast lesions (*a*) Primary benign (Fibroadenoma) (*b*) Primary malignant (Carcinoma) and (*c*) Secondary malignant (Metastasis) is considered as the first objective of the present research work.

(*ii*) The design and development of an efficient CAD system for two and three class breast lesions classification

Chapter 2 presents a brief review of the other related studies carried out for classification of breast lesions using ultrasound images.

Chapter 3 focuses on the research methodology that is followed for undertaking the present research work. This chapter gives a description of the image database used in the CAD system, the protocol followed to extract ROIs form each image. Various modules of the proposed CAD system.

Chapter 4 gives a detailed description of the proposed CAD system design for two class breast lesions classification using feature extraction methods.

Chapter 5 gives a detailed description of the proposed CAD system design for breast lesions classification using morphological features.

Chapter 6 gives a detailed description of the proposed CAD system design for breast lesions classification using difference vector of the statistical features.

LITERATURE REVIEW

2.1 Introduction

Characterization of breast lesions according to their nature is clinically significant because the invasive nature is associated with the risk of developing breast cancer. The radiologists after analyzing the ultrasounds predict the breast lesions but this visual analysis is highly subjective. For many cases, the differential diagnosis between breast lesions is difficult and it is considered to be daunting challenge even for experienced radiologists. Therefore there is a huge impetus among various researchers to develop CAD systems useful for differentiating between breast lesions.

2.2. Different CAD System Designs for Two-Class Breast Lesion Classification

Various researchers in the past have developed different CAD systems for classifying the breast lesion in to two classes namely benign and malignant. A brief description of these related studies is given in Table 2.1

Table 2.1: Summary of studies of comparison of two class breast lesions classification

Paper	Az	Accuracy(%)	Sensitivity(%)	Specificity(%)
Alveranga et al [17]	0.8	84	83	85
Wan et al [18]	Not Specified	86	78	91
Lio's et al [19]	Not Specified	87	75	96
Gomez et al [20]	0.97	90	90	91
Shi et al [21]	0.96	94	92	96

Alveranga et al [17] has investigated seven morphological features to distinguish between the malignant and the benign breast tumors on the US images. Accuracy achieved was 84 % and classification rate (A_z) was 0.8. The most relevant individual features were Normalized Residual Value (NRV) and the overlap ratio, both calculated from the convex polygon technique and the circularity (C) .When the Normalized Residual Value and circularity were taken together with the roughness calculated from the Normalized Radial Length (NRL).

Wan et al [18] formulated that by detecting and choosing essential features, the low rank matrix based on feature selection method can improve the classification outcomes. The dataset comprising the benign and malignant cases have training datasets of 92 benign cases and 172 malignant cases where as the testing datasets contains 21 benign cases and 36 malignant cases.

Lio group [19] established a new set of features for differentiating benign tumors from the malignant tumors. Sonograms of 321 pathologically proven breast lesions were analyzed and evaluated using SVM classifier. Gomez et al [20] proposed a CAD system for the breast ultrasound cases in which a differential evolution technique was used to optimize the structure of a radial basis function neural network.

Sci and his team [21] developed a CAD system, in which the classifier used is Fuzzy SVM classifier, the experimental results were achieved with the data set of 87 cases in which 36 malignant masses and 51 benign masses were included. Texture features extracted from ultrasound images are efficient features for classifying breast tumors [22]. Gomes et al. [23] extracted twenty two textural features from 436 ultrasound images and obtained classification accuracy with an area A_z of 0.87. An accuracy of 92.83% was achieved on 5500 images of prostate cancer by combining the histogram, GLCM, gray level run length matrix (GLRLM) to differentiate the different types [24].

Huang et al. [25] achieved an A_z of 0.909 using extracted nineteen morphological features using support vector machine (SVM) classifier from 118 breast ultrasound images. Wu et al. [26] combined auto-covariance texture features and morphological features to discriminate breast tumors in ultrasound images and achieves an accuracy of

92.86% using SVM classifier from 210 breast ultrasound images. In a later work [27] using the SVM genetic classifier on the same data base, an accuracy of 96.14% was achieved. Alvaranga et al. [28] achieved an accuracy of 85.37% on a database of 246 ultrasound images in distinguishing breast tumors using fisher linear discriminant analysis(FLDA) classification.

2.3   Concluding Remarks

From the above discussion, it can be concluded that most of the researchers have worked to characterize the breast lesion in two categories i.e. Benign and Malignant classes by using feature extraction techniques and classifiers. Feature extraction module extracts feature with the use of Morphological methods and Texture based methods

<div align="right">

CHAPTER 3

</div>

METHODOLOGY

3.1 Introduction

From the literature survey presented in the previous chapter, it can be concluded that most of the related studies carried out in the past are based on the pre-processing of ultrasound image to extract the segmented breast lesion. In the present work, taking into consideration the effect of ROI, different CAD system designs are proposed for the classification of different breast lesions based on their underlying feature characteristics.

3.2 Proposed CAD System Designs for Classification of Breast Lesions

In the present work, various CAD system designs have been proposed to classify the different breast lesions. A general framework of the different CAD schemes employed in the present work is shown in Figure 3.1. For the design of this CAD system, a dataset of 167 ultrasound images was taken. The CAD system consists of (*a*) ROI extraction module (*b*) feature extraction module (*c*) classification module. In the feature extraction module, for extracting the features two methods are employed (*1*) Morphological Features (*2*) Texture based features. The texture based features have the

Table 3.1 : Texture based features

Statistical Methods	• FOS • GLCM • GLRLM • NGTDM • SFM
Signal Processing Methods	• Laws' Features
Transform Domain Methods	• Gabor Wavelet Based • FPS features • Discrete Wavelet Transform

9

Each feature set is normalized by using min-max normalization. The normalized feature set is then bifurcated into training and testing datasets. In the classification module, performance of four different classifiers namely k-NN, probabilistic neural network (PNN), SVM and smooth support vector machine (SSVM) is evaluated to obtain the class of the unknown testing instances.
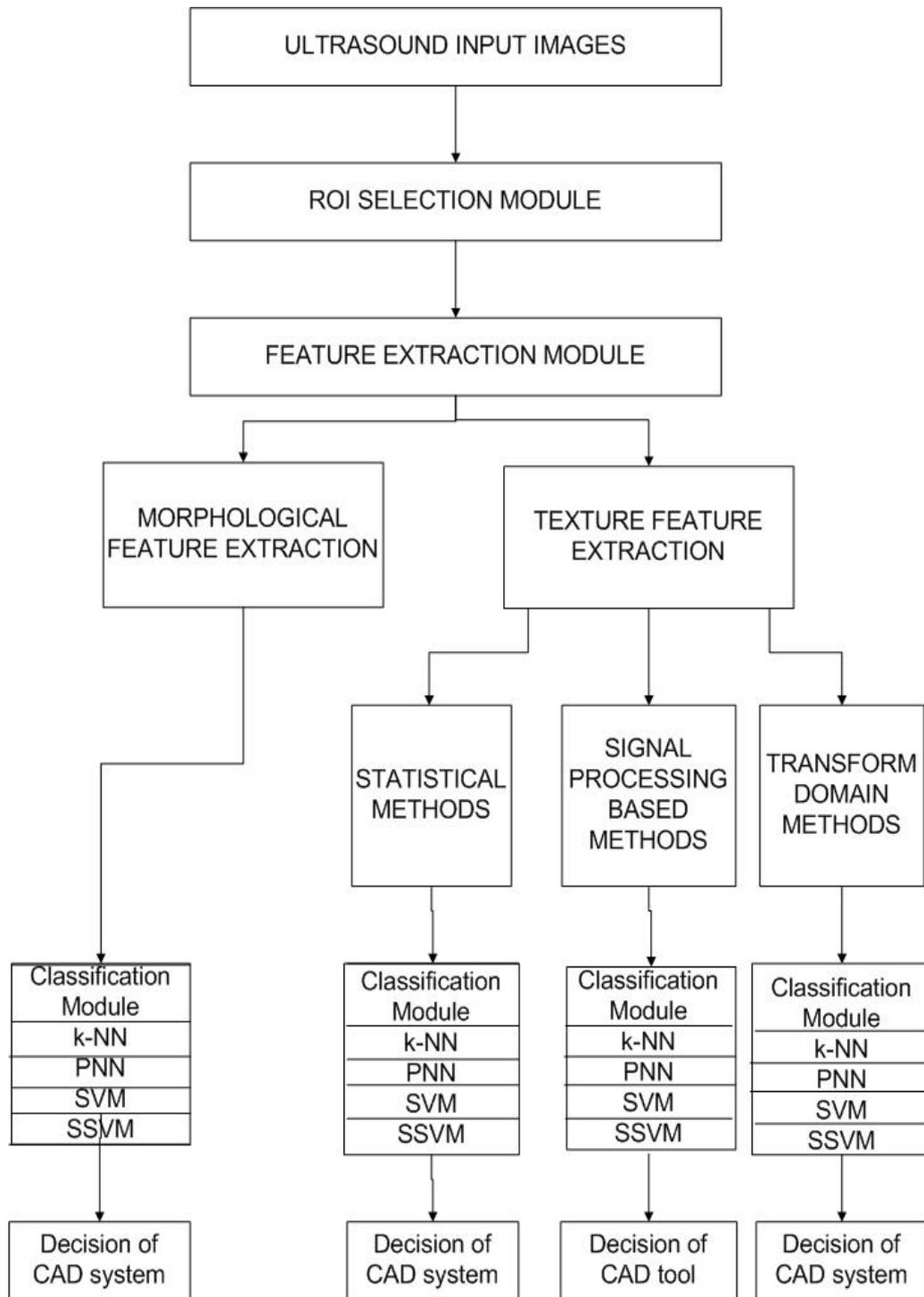
Figure 3. 1: General framework of the proposed CAD system design

10

3.2   Dataset Description

In order to test the proposed CAD system design, benchmarked database is taken online from http://www.ultrasoundcases.info/category.aspx?cat=67 [15]
This database consists of total of 167 ultrasound images of both the left and right breast. The images in the database are categorized into two categories as primary benign (51 images), primary malignant cases (74 images) and secondary malignant cases (42 images). The database doesn't include the dataset of biopsy cases and the cases having the Color Doppler Effect present. In the present work CAD system designs have been proposed for (*a*) two-class breast lesions classification i.e. (Benign and Malignant class) and (*b*) three-class breast lesions classification i.e. (Primary Benign, Primary Malignant and Secondary Malignant). For implementing CAD systems for two-class breast lesion classification, the dataset of primary benign is taken for benign case and the cases of Primary and secondary Malignant are combined for the cases of Malignant. The description of the dataset used for two-class is shown in Figure 3.2
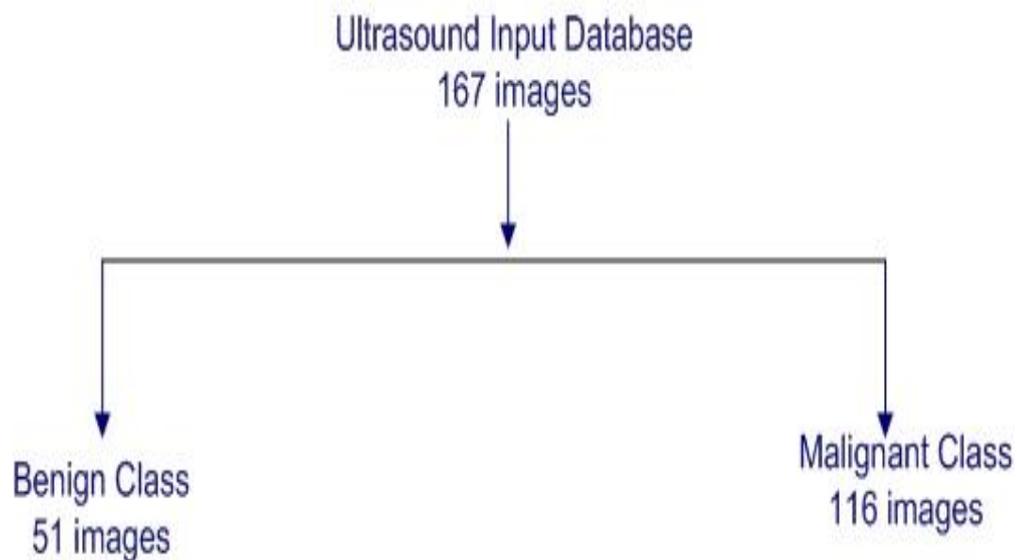
Figure 3.2 : Dataset description of two class breast lesions

and three-class dataset description is shown in Figure [3.3]
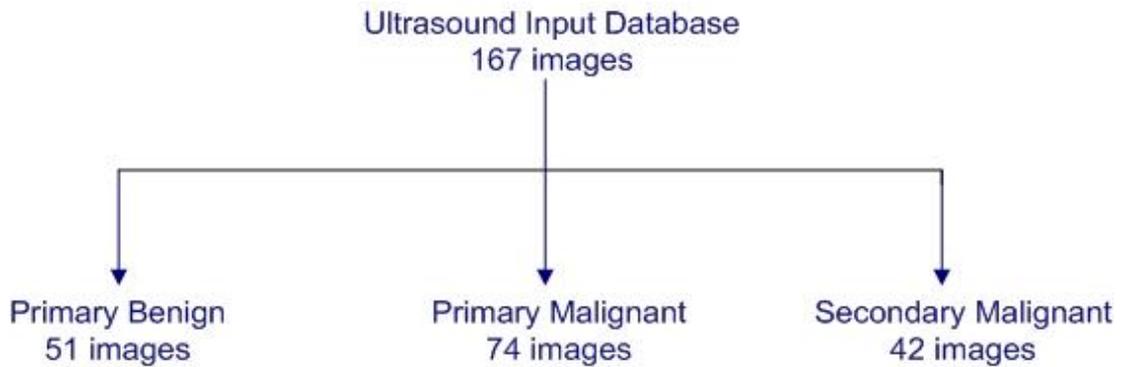
Figure 3. 3: Dataset description of three class Breast Lesions

3.3.1 Selection of Regions of Interest (ROIs)

The ROI size is selected carefully considering the fact that it should provide a good population of pixels for computing texture features. The abnormality in the ultrasound is identified and marked with the help of an experienced radiologist and segmented with the help of software *Image J* [16].This software helps to load the image, mark the infected area and segment it. Further the segmented region is enclosed into a rectangular bounding box adjoining the boundaries of abnormality.

Selection of ROI is represented in Figure 3.4

(a)   (b)
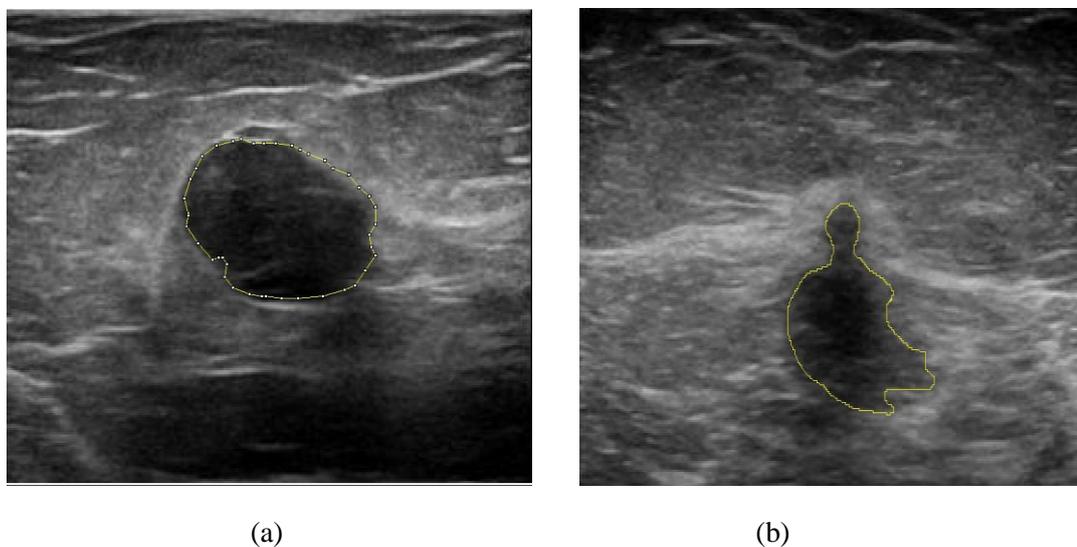
Figure 3. 4   **:** (a) Benign Marked (b) Malignant Marked

12

3.4 Feature Extraction Module

The feature extraction is the process used to transform the visually extractable and non-extractable features into mathematical descriptors. These descriptors are shape-based (morphological features) and the intensity distribution based (textural features). The intensity distribution methods include (*a*) Statistical Methods (*b*) Signal Processing Based Methods (*c*) Transform domain methods

3.4.1. Texture Features

3.4.1.1. Statistical Features

The statistical methods are used to extract the texture features from an image based on the gray level intensities of the pixels of that image. Based on the number of pixels used to compute the texture features, statistical methods can be classified into first-order statistics, second-order statistics and higher-order statistics.

3.4.1.1.1. First Order Statistics (FOS)

The first order statistics are derived from the gray level intensity histograms of the image. Six features namely average gray level, standard deviation, smoothness, kurtosis and entropy are computed for each ROI [30].

3.4.1.1.2. Second Order Statistics-GLCM Features

Second order statistics includes the computations with the GLCM (Gray level co-occurrence matrix). GLCM explains the frequent combinations of pixels pairs having different gray level occurring in an image having separation of different dimensions in different directions say 0°, 45°, 90°, 135°. Total of 13 GLCM features are taken for computation in this paper , these features includes contrast, entropy, sum entropy, difference entropy , correlation, inverse difference moment, sum average, variance, sum variance, difference variance, information measures of correlation-1 and 2, Angular second moment [31-33].

3.4.1.1.3. Higher Order Statistics-GLRLM Features

Higher order statistics are computed with the use of GLRLM (Gray Level Run Length Matrix).Texture features are computed using the different combinations of intensities at relative position of each other. Gray level run is made by the set of consecutive pixels of gray levels that are collinear to each other and run length denotes the no of times a run occurs. The 11 GLRLM features that are computed in this work are long run emphasis, short run emphasis, low gray level run emphasis, high gray level run emphasis, run length non-uniformity and run percentage, short run low gray level emphasis, long run low gray level emphasis, gray level non uniformity, short run high gray level emphasis, long run high gray level emphasis [34, 35].

3.4.1.1.4. Other Statistical Features

(a) Edge Features (Absolute Gradient): There is always more information present in the edges than the other parts. The spatial variation in an image is computed by calculating the gradient value. If there are abrupt changes present then gradient will be high else it will be low. Edge features computes two features: Absolute gradient mean and absolute gradient variance [36].

(b) Neighborhood Gray Tone Difference Matrix (NGTDM) Features: NGTDM (Neighborhood Gray Tone Difference Matrix) computes busyness, coarseness, complexity, contrast, and strength and considers a difference between the gray level between pixels [37].

(c) Statistical Feature Matrix (SFM): SFM (Statistical Feature Matrix) computes coarseness, contrast, periodicity and roughness of pixels at different distances within an image.

(d) Gray Level Difference Statistics (GLDS): GLDS (Gray Level Difference Statistics) computes contrast, energy, entropy, homogeneity, and mean on the basis of the co-occurrence of a pixel pair that have difference in gray levels separated by a particular distance [38, 39].

3.4.1.2   Signal Processing based Methods

Laws' Mask Texture Analysis: In this method small convolution masks are used as filters and ROIs are convolved with these filters so that the underlying texture characteristics are enhanced. These filters determine the properties of the texture by performing averaging, edge detection, spot detection, wave detection and ripple detection [40-45]. Laws' masks of lengths 3, 5, 7 and 9 are used to compute five statistical parameters i.e. mean, standard deviation, skewness, kurtosis and entropy from each ROI as explained in Appendix.

3.4.1.3   Transform Domain based Methods

Feature extraction can also be done in the transform domain over various scales by using different multi resolution schemes like wavelet packet transform (WPT) and Gabor Wavelet transform (GWT). It is logical to compute texture features in the transform domain as human visual system processes images in a multi scale way and scale is considered to be an important aspect for analysis of texture [46-48].

3.4.1.3.1. Two Dimensional Discrete Wavelet Transform

A two-dimensional DWT when applied to images can be seen as two one-dimensional transform functions applied to rows and columns of the image separately [49]. When this operation is applied to an ROI image, a decomposition is done up to second level. The choice of wavelet filter used for feature extraction is based on some properties which are significant for texture description [50, 51]. The properties that are considered for selecting an appropriate wavelet filter include: support width, orthogonality or biorthogonality, shift invariance and symmetry. Wavelet filters that provide compact support are desirable due to their ease of implementation. Orthogonality is required for energy conservation at each level of decomposition. Symmetry is desired to avoid any dephasing while processing images. The properties of these filters are summarized in Table 3.2

Table 3.2: Properties of wavelet filters

Wavelet Filter	Biorthogonal	Orthogonal	Symmetry	Asymmetry	Near Symmetry	Compact Support
Db	No	Yes	No	Yes	No	Yes
Haar	No	Yes	Yes	No	No	Yes
Bior	Yes	No	Yes	No	No	Yes
Coif	No	Yes	No	No	Yes	Yes
Sym	No	Yes	No	No	Yes	Yes

3.4.1.3.2. Two Dimensional Gabor Wavelet Transform

The application of 2D-GWT results in a set of frequency and orientation selective filters that capture energy at a specific frequency and orientation. The 2D-GWT, considering three scales (0,1 and 2) and seven angles (22.5°, 45°, 67.5°, 90°, 112.5°, 135° and 157.5°) result in a group of 21 wavelets (7 × 3). When this group of wavelets is convolved with the ROI image, a set of 21 feature images are obtained. Each of these filtered images represents image information at a certain scale and orientation [52, 53]. From these 21 feature images, mean and standard deviation are computed as texture features forming a texture feature vector (TFV) of length 42. The real part of the 21 wavelets resulting from a 13 × 13 convolution mask with 3 scales and 7 orientations are shown in Figure 3.6.
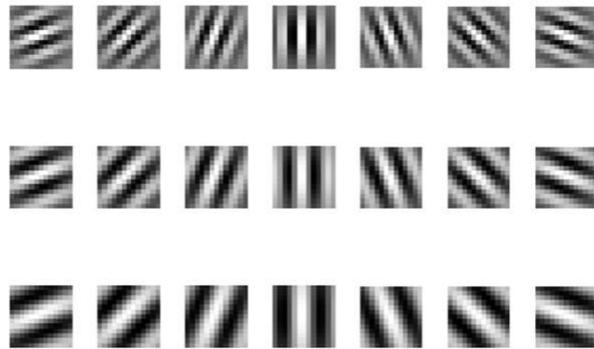
Figure 3. 5 : Real part of Gabor filter family of 21 wavelets.

3.4.1.3.3. Fourier Power Spectrum Features

Two spectral features namely radial sum and angular sum are computed from each ROI using discrete Fourier transform.

3.4.2   Morphological Features

Morphological methods include the shape based properties which includes Area, Perimeter, Convexity, Eccentricity, Extent, Hole Area Ratio (HAR) and Solidity [17] are calculated over the entire class of Benign and Malignant.

(*a*). Area: It calculates the area of the lesion.

(*b*). Perimeter: It calculates the perimeter of the lesion.

(*c*).Convexity: It is the ratio of the perimeter of the convex hull to the overall contour.

(*d*). Diameter: It is the diameter of the circle which have an equivalent area   as the region.

(*e*). Major Axis and Minor Axis: These are the diameters of the ellipse, where major axis is the longest diameter and the minor axis is the smallest diameter.

(*f*).Eccentricity: It is the ratio of the minor axis to the major axis. Its value always lies between the 0 and 1.

(*g*).Extent: It is the ratio of the pixels in the bounding box that are also present in the region.

$$\text{Extent} = \text{Area} / \text{Bounding Area} \qquad\qquad (1)$$

(*h*). Euler No: It defines the relationship between the no of contiguous part to the no of holes in the shape.

(*i*). Solidity: It gives the extent to which the given shape is convex or concave.

$$\text{Solidity} = \text{Area} / \text{Convex Area} \qquad\qquad (2)$$

3.5   Classification Module

Classification is the process of grouping the testing samples into the corresponding classes. Classification is characterized into two types viz. supervised classification and the unsupervised classification. Classification is supervised if the classes are already defined for the training sets otherwise it is unsupervised classification. In this module different classifiers like

- k-NN
- PNN
- SVM
- SSVM

are employed to classify the unknown testing instances of different ultrasound classes based on the training instances. Min-Max normalization procedure is used to normalize the extracted features in the range [0,1] in order to avoid any bias by unbalanced features.

3.5.1 k-Nearest Neighbour (k-NN) Classifier

The k-NN classifier is based on the idea of estimating the class of an unknown instance from its neighbors. It works on the assumption that the feature vector lying close to each other belongs to the same class therefore it tries to group the instances of feature vector into same classes lying close to each other. The class of an unknown instance is selected by looking among k-nearest in the training dataset. The advantage of k-NN is its ability to deal with multiple class problems and as it averages the k- nearest neighbors so it is robust to the noisy data problems [54-57]. Various distance metrics can be used to calculate say Euclidean distance, Cosine distance, City Block, Chebychev, Minkowski and Correlation. In this module Euclidean distance is used as a distance metric. The classification performance of k-NN classifier depends on the value of k. The example depicting the classification of an unknown instance is shown in Figure 3.7. In the example the test sample ($\times$) should be either classified to the class of cross (+) or to the class of dash (-). When $k = 3$, the algorithm looks for three nearest neighbors.
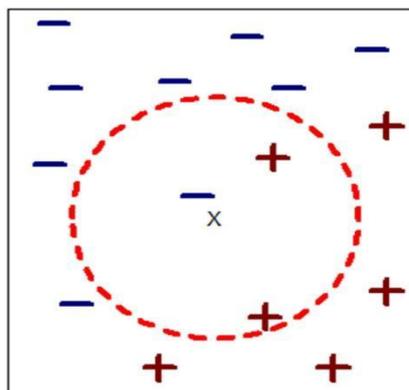
Figure 3.6 : Example of k-NN classification for $k = 3$.

Note: $\times$: unknown instance, +: Instance of class 1, −: Instance of class 2.

18

3.5.2 Probabilistic Neural Network (PNN) Classifier

It is supervised Bayesian based feed-forward neural network used for estimating the class of unknown instances [58-61]. The PNN classifier comprises of four layers: input layer, pattern layer, summation layer and output layer. The values from the testing dataset are passed to the 'n' neurons in the input unit. The values from the input layer are further forwarded to the pattern units in the pattern layer where responses for each unit are calculated on basis of probability density function.

The values of the pattern unit are forwarded to the summation layer where the responses are summed get response in each category. The maximum response obtained from all categories is taken into the decision layer to get the class of the unknown instance. The choice of spread parameter (SP) or the kernel width parameter is crucial for the PNN classifier. The optimal values used for SP to design a PNN classifier are determined by performing repeated experiments for values of SP.

3.5.3 Support Vector Machine (SVM) Classifier

Support Vector Machine (SVM) classifier is included into class of supervised learning machine and works on basis of statistical theory. SVM classifier supports both linear and non linear classification. Support vector machines creates a hyperplane between the classes with the help of the training data available and good separation is achieved intuitively but the sets that are available to discriminate are not linearly separable in the space. Therefore in non linear classification problems, the input data is mapped in to the kernel functions in which the data is mapped from input space to the higher dimensional feature space. For the classification task, Gaussian radial basis kernel's function has been used. Recent algorithms include the sub gradient and coordinate descent methods that have a big plus of having large and sparse datasets. LibSVM library is included for the implementation of SVM classifier [62-64].

3.5.4 Smooth Support Vector Machine (SSVM) Classifier

SVM is associated with the traditional quadratic program, therefore to unconstraint

smoothing unconstrained optimization reformulation SSVM classifier is used. SSVM work on the concept of smoothing unconstrained optimization reformulation the problems related to SVM for the pattern classification [65, 66]. The SSVM toolbox has been developed by Laboratory of Data Science and Machine Intelligence, Taiwan for implementation of SSVM. Similar to SVM implementation in case of SSVM also, ten-fold cross validation is carried out on training data for each combination. The procedure of grid search in parameter space gives the optimum values of C and for which training accuracy is maximum.

3.6   Concluding Remarks

After carrying out extensive literature survey, it was concluded that various CAD system designs have proven useful to the radiologists in routine medical practice as second opinion tools for breast lesions classification of ultrasounds in cases where a clear discrimination cannot be made subjectively. In light of this fact, different CAD system designs employing the morphological features , the texture analysis techniques of feature extraction and feature classification have been proposed in the present work for two-class breast lesions classification of ultrasound.

CAD SYSTEM DESIGN FOR BREAST LESIONS CLASSIFICATION USING TEXTURE FEATURES

4.1 Introduction

The differential diagnosis between breast lesions from ultrasound images is a daunting challenge even for the experienced radiologists. Therefore a CAD system for the classification of the different breast lesions from ultrasound images is highly desirable. In light of this fact, a CAD system design is proposed in this chapter to evaluate the performance of different classifiers for two-class breast lesions classification for different texture features. The texture features include (*a*) Statistical features (*b*) Signal Processing methods (*c*) Transform Domain Features

4.2 Proposed CAD System Design

The block diagram of the proposed CAD system design for two-class breast lesions classification using statistical features is shown in Figure 4.1.

The approach is implemented on the 167 ultrasound database. From each ultrasound image, ROIs of are extracted and from each ROI image, different statistical features are calculated like FOS, GLCM features, GLRL features, features derived from GLDS, NGTDM features, SFM features and Edge features.
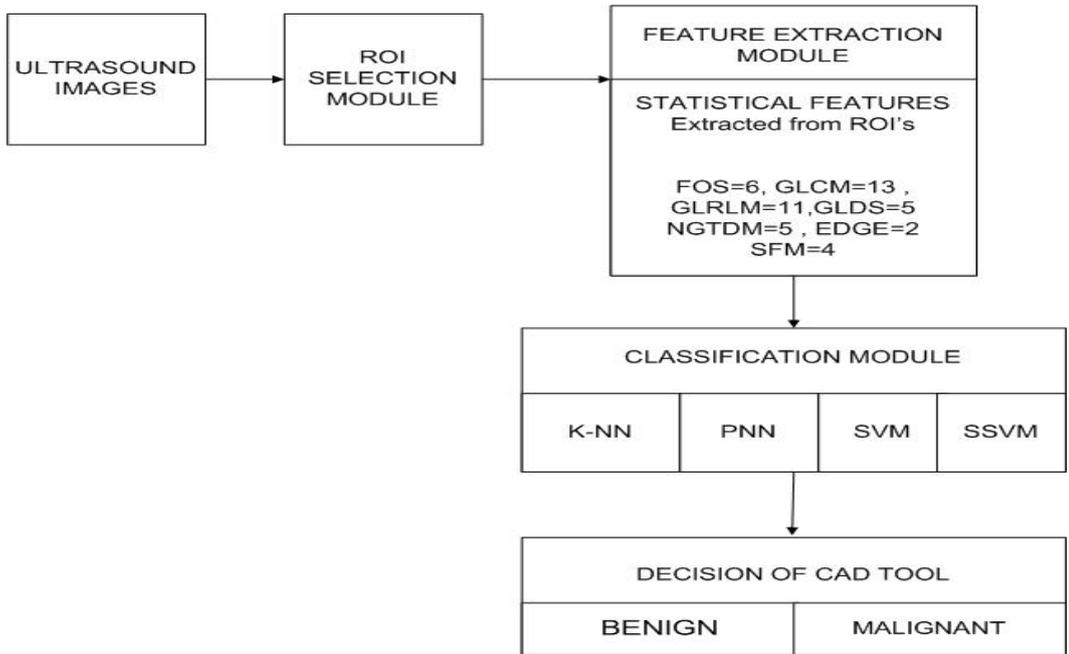
Figure 4.1 : Proposed CAD system design using statistical features for two-class breast lesions classification for statistical methods

The block diagram of the proposed CAD system design for two-class and three-class breast lesions classification using transform domain features is shown in Figure 4.2.
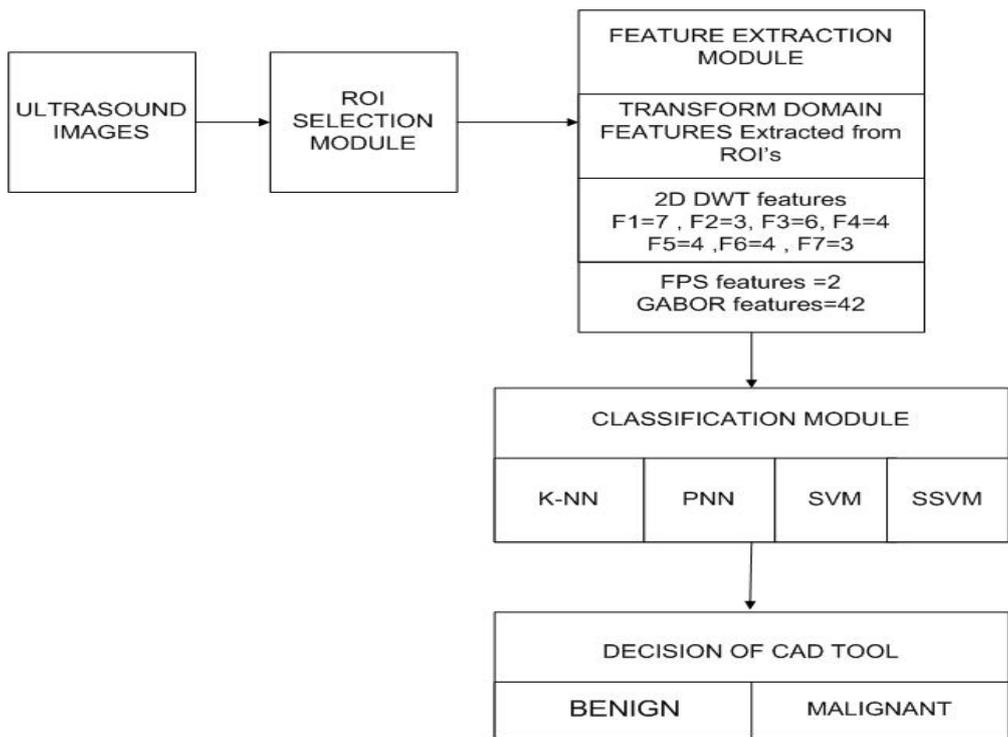
Figure 4.2: Proposed CAD system design for two-class breast lesions classification using transform domain methods

The transform domain methods include Gabor features, FPS features and wavelet energy based features.

The block diagram of the proposed CAD system design for two-class breast lesions classification using Laws' features is shown in Figure 4.3.
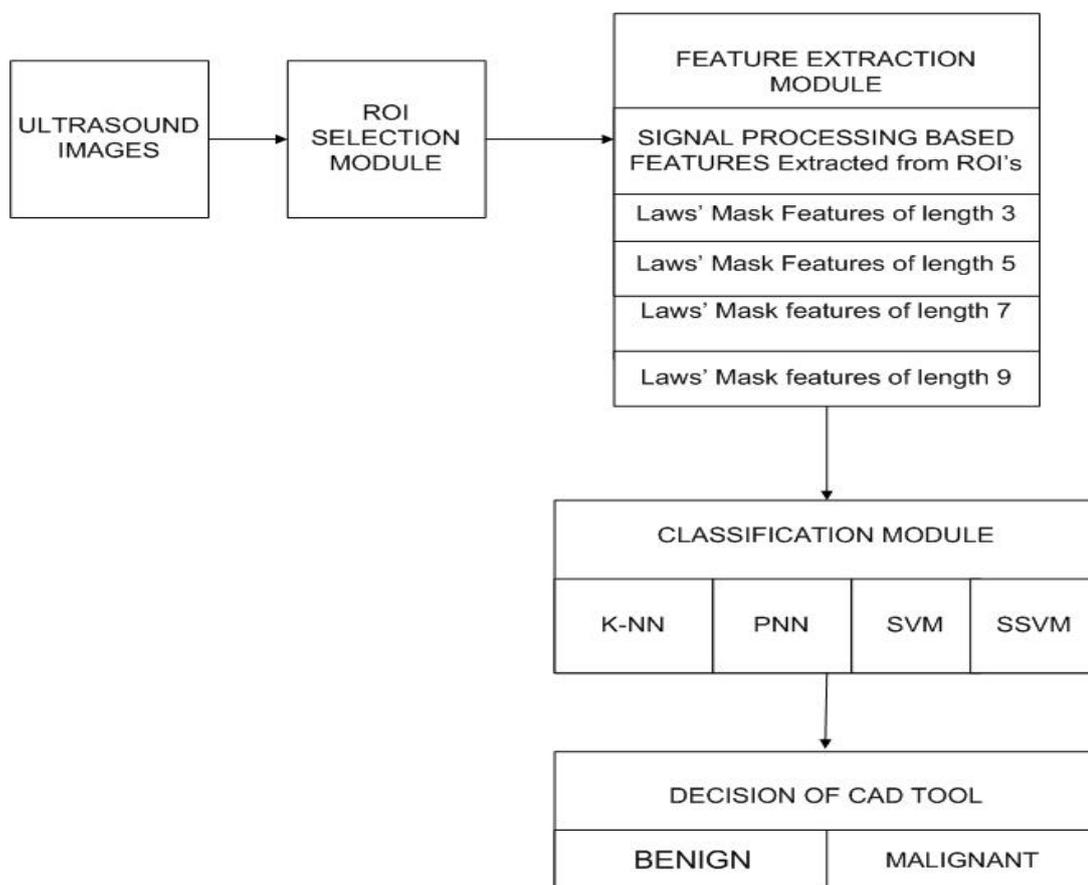
Figure 4.3: Proposed CAD system design using Laws' texture features for two-class breast lesion classification

4.3 Experimental Workflow and Results

For evaluating the performance of the proposed CAD system design, rigorous experimentation has been carried out for the characterization of the breast lesions. A brief description of experiments is given in Table 4.1 and Table 4.2, respectively for two-class

breast lesions classification for statistical features and Transform domain methods. Table 4.3 gives the description of texture feature vectors used in Laws' Mask based methods and Table 4.4 is for the description for experiments using the Laws' Mask.

Table 4.1: Description of experiments carried out breast lesion classification using Statistical Features

Experiment : To obtain the classification performance of statistical features for two-class breast lesions classification using k-NN, PNN, SVM and SSVM classifiers.

Table 4.2: Description of experiments carried out for breast lesions classification using Transform Domain Methods

Experiment : To obtain the classification performance of TFV for two-class breast lesions classification using different classifiers

Note: TFV: Texture feature vector.

Table 4.3: Description of experiments carried out for two-class breast lesions classification using Signal Processing Methods

Experiment : To obtain classification performance of different TFVs (derived from Laws' masks of length 3, 5, 7 and 9) using different classifiers

Note: TFV: Texture feature vector.

4.3.1. Experiment carried out for two-class breast lesions classification for Statistical features

In this experiment, classification performance of TFV containing different statistical features is evaluated for two-class breast lesions classification using different classifiers. The results of the experiment are shown in Table 4.3. It can be concluded from the table that for statistical features, the overall classification accuracy of 83.1 %, 80.5 %, 79.2 % and 76.6 % is achieved using k-NN, PNN, SVM and SSVM classifiers, respectively. It can also be concluded that the highest in individual class accuracy for benign case is 57.1 % with SVM classifier and highest individual class accuracy for malignant case is 100.0 %, using SSVM classifier.

Table 4.4: Classification performance of statistical features using kNN, PNN, SVM and SSVM classifiers for two-class breast lesions classification

Classifier	CM			OCA (%)	ICA_B(%)	ICA_M(%)
		B	M			
k-NN	B	11	10	83.1	52.3	94.6
	M	3	53			
PNN	B	11	10	80.5	52.3	91
	M	5	51			
SVM	B	12	9	79.2	57.1	87.5
	M	7	49			
SSVM	B	3	18	76.6	14.2	100
	M	0	56			

Note: CM: Confusion matrix, B:Benign class, M: Malignant class, OCA: Overall classification accuracy; ICA_B: Individual class accuracy for benign class. ICA_M: Individual class accuracy for malignant class.

4.3.2. Experiment carried out for two-class breast lesions classification for Transform Domain features

Table 4.5: Classification performance of Transform Domain Features using k-NN, PNN, SVM and SSVM classifiers for two-class breast lesions classification

Classifier	CM			OCA (%)	ICA_B (%)	ICA_M (%)
		B	M			
kNN	B	16	5	88.3	76.1	92.8
	M	4	52			
PNN	B	17	4	88.3	80.9	91.0
	M	5	51			
SVM	B	19	2	88.3	90.4	87.5
	M	7	49			
SSVM	B	6	15	75.3	28.5	92.8
	M	4	52			

Note: CM: Confusion matrix, B: Benign class, M: Malignant class, OCA: Overall classification accuracy; ICA_B: Individual class accuracy for benign class. ICA_M: Individual class accuracy for malignant class.

In this experiment, it can be concluded from the table that for TFV, the overall classification accuracy values of 88.3 %, 88.3 %, 88.3 % and 75.3 % are achieved using

k-NN, PNN, SVM and SSVM classifiers, respectively. It can also be concluded that the highest in individual class accuracy for benign class is 90.4 % with SVM classifier and highest individual class accuracy for malignant class is 92.8 %, using k-NN classifier.

4.3.3. Experiments carried out for two-class breast lesion classification using Law's Mask

Table 4.6 : Description of texture feature vectors of signal processing methods

TFV1:	TFV derived from Laws' mask of length 3 (5 statistical features computed from 6 TR images i.e. 5 $\times$ 6= 30 features)
TFV2:	TFV derived from Laws' mask of length 5 (5 statistical features computed from 15 TR images i.e. 5 $\times$ 15 =75 features)
TFV3:	TFV derived from Laws' mask of length 7 (5 statistical features computed from 6 TR images i.e. 5 $\times$ 6= 30 features)
TFV4:	TFV derived from Laws' mask of length 9 (5 statistical features computed from 15 TR images i.e. 5 $\times$ 15 =75 features)

Note: TFV: Texture feature vector, TR: Rotation invariant images.

4.3.3.1. Experiment 1: To obtain classification performance of different TFVs (derived from Laws' masks of length 3, 5, 7 and 9) using k-NN classifier.

In this experiment the classification performance of the feature set containing instances of TFV1, TFV2, TFV3 and TFV4 is evaluated using *k*-NN classifier. The results are shown in Table 4.7.

Table 4.7: Classification performance of different TFVs using *k*-NN classifier for two-class breast lesion classification

TFV(*l*)	CM			OCA (%)	ICA$_B$ (%)	ICA$_M$(%)
		B	M			
TFV1(30)	B	16	5	92.2	76.1	98.2
	M	1	55			
TFV2(75)	B	15	6	92.2	71.4	100
	M	0	56			
TFV3(30)	B	13	8	89.6	61.9	100
	M	0	56			
TFV4(75)	B	16	5	93.5	76.1	100
	M	0	56			

From the table it can be concluded that a classification accuracy of 92.2 %, 92.2 %, 89.6 % and 93.5 % is achieved using TFV1, TFV2, TFV3 and TFV4, respectively. The individual class accuracy of benign class is 76.1 %, 71.4 %, 61.9 % and 76.1 % for TFV1, TFV2, TFV3 and TFV4, respectively. For malignant class the individual class accuracy is 98.2 %, 100 %, 100 % and 100 % for TFV1, TFV2, TFV3 and TF4

4.3.3.2. Experiment 2: To obtain classification performance of different TFVs (derived from Laws' masks of length 3, 5, 7 and 9) using PNN classifier.

In this experiment the classification performance of the feature set containing instances of TFV1, TFV2, TFV3 and TFV4 is evaluated using PNN classifier. The results are shown in Table 4.10.

Table 4.8 : Classification performance different TFVs using PNN classifier for two-class breast lesions classification

TFV(*l*)	CM			OCA (%)	ICA_B (%)	ICA_M (%)
		B	M			
	B	12	9			
TFV1(30)				88.3	57.1	100
	M	0	56			
	B	13	8			
TFV2(75)				89.6	61.9	100
	M	0	56			
	B	10	11			
TFV3(30)				85.7	47.6	100
	M	0	56			
	B	14	7			
TFV4(75)				90.9	66.6	100
	M	0	56			

From the table it can be concluded that a classification accuracy of 88.3 %, 89.6 %, 85.7 % and 90.9 % is achieved using TFV1, TFV2, TFV3 and TFV4, respectively. The individual class accuracy for benign is 57.1 %, 61.9 %, 47.6 % and 66.1 % for TFV1,

TFV2, TFV3 and TFV4, respectively. For malignant class the individual class accuracy is 100 % for TFV1, TFV2, TFV3 and TFV4, respectively.

4.3.3.3. Experiment 3: To obtain classification performance of different TFVs (derived from Laws' masks of length 3, 5, 7 and 9) using SVM classifier

Table 4.9: Classification performance of different TFVs using SVM classifier for two-class breast lesions classification

TFV(*l*)	CM			OCA (%)	ICA$_B$ (%)	ICA$_M$ (%)
		B	M			
TFV1(30)	B	14	7	89.6	66.6	98.2
	M	1	55			
TFV2(75)	B	19	2	97.4	90.4	100
	M	0	56			
TFV3(30)	B	12	9	87.0	57.1	98.2
	M	1	55			
TFV4(75)	B	16	5	93.5	76.1	100
	M	0	56			

Note: TFV: Texture feature vector, *l*: Length of TFV, CM: Confusion Matrix, B: Benign, M: Malignant, OCA: Overall classification accuracy; ICA$_B$: Individual class accuracy for Benign class. ICA$_M$: Individual class accuracy for Malignant class.

In this experiment the classification performance of the feature set containing instances of TFV1, TFV2, TFV3 and TFV4 is evaluated using the SVM classifier. The results are shown in Table 4.11. From the table it can be concluded that a classification accuracy of 89.6 %, 97.4 %, 87 % and 93.5 %, is achieved using TFV1, TFV2, TFV3 and TFV4, respectively. The individual class accuracy for benign class is 66.6 %, 90.4 %, 57.1 % and 76.1 % for TFV1, TFV2, TFV3 and TFV4, respectively. For Malignant class the individual class accuracy is 98.2 %, 100 %, 98.2 % and 100 % for TFV1, TFV2, TFV3 and TFV4, respectively.

4.3.3.4. Experiment 3: To obtain classification performance of different TFVs (derived from Laws' masks of length 3, 5, 7 and 9) using SSVM classifier

Table 4.10: Classification performance of different TFVs using SSVM classifier for two-class breast lesions classification

TFV(l)	CM			OCA (%)	ICA$_B$ (%)	ICA$_M$ (%)
		B	M			
	B	18	3			
TFV1(30)				96.1	85.7	100
	M	0	56			
	B	15	6			
TFV2(75)				92.2	71.4	100
	M	0	56			
	B	10	11			
TFV3(30)				83.1	47.6	96.4
	M	2	54			
	B	7	14			
TFV4(75)				81.8	33.3	100
	M	0	56			

Note: TFV: Texture feature vector, l: Length of TFV, CM: Confusion Matrix, B: Benign, M: Malignant, OCA: Overall classification accuracy; ICA$_B$: Individual class accuracy for Benign class. ICA$_M$: Individual class accuracy for Malignant class.

From the table it can be concluded that a classification accuracy of 96.1 %, 92.2 %, 83.1 % and 81.8 % is achieved using TFV1, TFV2, TFV3 and TFV4, respectively. The highest individual class accuracy of benign class is 96.1 % for TF1, respectively. For malignant class the highest individual class accuracy is 100 %for TFV1, TFV2, TFV3 and TF4

4.4   Concluding Remarks

From the results obtained from the above experiments, it can be concluded that for two-class breast lesions, k-NN classifier achieves highest classification accuracy of 83.3%   for two-class breast lesion classification for statistical features, 88.3 % with SSVM classifier for transform domain methods classification and 97.4 % accuracy is achieved with Laws' mask of length 5.

CAD SYSTEM DESIGN FOR BREAST LESIONS CLASSIFICATION USING MORPHOLOGICAL FEATURES

5.1 Introduction

The differential diagnosis between breast lesions from ultrasound images is a daunting challenge even for the experienced radiologists. Therefore a CAD system for the classification of the different breast lesions from ultrasound images is highly desirable. In light of this fact, a CAD system design is proposed in this chapter to evaluate the performance of different classifiers for two-class breast lesions using morphological features and combination with statistical features.

5.2 Proposed CAD System Design

The block diagram of the proposed CAD system design for two-class breast lesions classification using morphological features is shown in Figure 5.1. The approach is implemented on the 167 ultrasound database. From each ultrasound image, ROIs of are extracted and from each ROI image, different morphological features are calculated like Area, Perimeter, Solidity, Convexity, Extent etc.
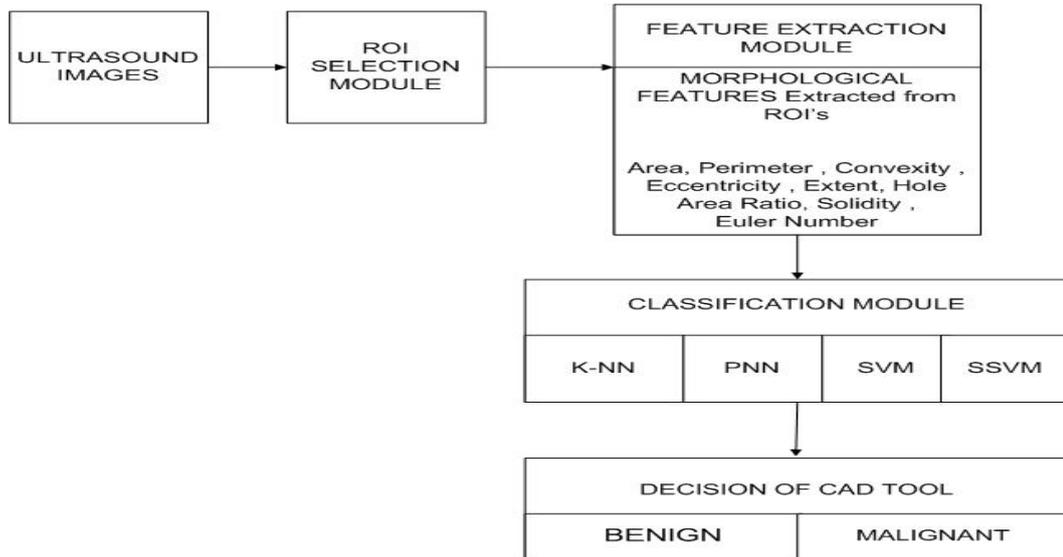
Figure 5.1: Proposed CAD system design for two-class breast lesions classification using morphological features

5.3 Experimental Workflow and Results

For evaluating the performance of the proposed CAD system design, rigorous experimentation has been carried out for the characterization of the breast lesions. A brief description of experiments is given in Table 5.1

Table 5.1: Description of experiments carried out for breast tissue lesions classification using morphological features

Experiment 1: To obtain the classification performance of Morphological features for two-class breast lesions classification using k-NN, PNN, SVM and SSVM classifiers.

Experiment 2: To obtain the classification performance of morphological features and individual statistical features for two-class breast lesions classification using SVM classifier.

Experiment 3 : To obtain the classification performance of morphological features and all statistical features for two-class breast lesions classification using SVM classifier.

5.3.1. Experiments carried out for two-class breast lesions classification using Morphological Feature

In this experiment, classification performance of TFV containing different morphological features is evaluated for two-class breast lesions classification using different classifiers. The results of the experiment are shown in Table 5.2. It can be concluded from the table that for morphological features, the overall classification

31

accuracy of 77.9 %, 75.3 %, 81.8 % and 72.7 % is achieved using k-NN, PNN, SVM and SSVM classifiers, respectively. It can also be concluded that the highest in individual class accuracy for benign class is 71.4 % with SVM classifier and highest individual class accuracy for malignant class is 100.0 %, using SSVM classifier.

Table 5.2: Classification performance of statistical features using k-NN, PNN, SVM and SSVM classifiers for two-class breast lesions classification

Classifier	CM			OCA (%)	ICA$_B$(%)	ICA$_M$ (%)
		B	M			
k-NN	B	14	7	77.9	66.6	82.1
	M	10	46			
PNN	B	14	7	75.3	66.6	78.5
	M	12	44			
SVM	B	15	6	81.8	71.4	85.7
	M	8	48			
SSVM	B	0	21	72.7	Nil	100
	M	0	56			

Note: CM: Confusion matrix, B: Benign , M: Malignant, OCA: Overall classification accuracy; ICA$_B$: Individual class accuracy for benign , ICA$_M$: Individual class accuracy for malignant class.

5.3.2. Experiments carried out for two-class breast lesions classification using combination of morphological and individual statistical features.

In this experiment, the classification performance containing combination of morphological and different individual statistical features are evaluated for two-class breast lesions classification using different classifiers. The results are shown in Table 5.3. It can be concluded from the table that the overall classification accuracy of 89.6 is achieved using combination of FOS with morphological features. The highest individual class accuracy is 89.3 % using SVM classifier, for benign class the highest individual class accuracy achieved is 76.1 % and for the malignant class is 98.2 %

Table 5.3 : Result of Classification of Combination of Morphological Features with various Indiv
Statistical features

Features	CM			OCA (%)	ICA$_B$(%)	ICA$_M$ (%)
		B	M			
Edge +MF	B	14	7	84.4	66.6	91
	M	5	51			
SFM +MF	B	16	5	85.7	76.1	90.9
	M	6	50			
NGTDM +MF	B	13	8	83.1	61.9	91
	M	5	51			
FOS +MF	B	15	6	89.6	71.4	94.6
	M	5	51			
GLCM +MF	B	16	5	87.0	76.1	91
	M	5	51			
GLRLM +MF	B	15	6	87.0	71.4	92.8
	M	4	52			
GLDS +MF	B	13	8	88.3	61.9	98.2
	M	1	55			

Note : CM :Confusion matrix , OCA : Over all classification accuracy , B: benign class , M: Malignant Class , ICA$_B$: Individual class accuracy of Benign class , ICA$_M$: Individual class accuracy of Malignant class , MF: Morphological Features, FOS : First order Statistics , GLCM : Gray Length Co-occurrence Matrix , GLRLM : Gray level Run Length Matrix , GLDS: Gray level difference statistics SFM : Statistical Feature Matrix , NGTDM : Neighbourhood Gray Tone Difference Matrix

5.3.3. Experiments carried out for two-class breast lesions classification using combination of morphological and statistical features.

In this experiment, the classification performance containing combination of morphological and statistical features are evaluated for two-class breast lesions classification using SVM classifier. The results are shown in Table 5.4. It can be concluded from the table that the overall classification accuracy of 83.1 is achieved using combination of morphological features and statistical features . The highest individual class accuracy for benign class is 66.6 % using SVM classifier and for the malignant class is 89.2 % .

Table 5.4: Classification performance of combination of morphological and statistical features using SVM classifier

FV	CM			OCA (%)	ICA$_B$(%)	ICA$_M$ (%)
		B	M			
	B	14	7			
MF+SF				83.1	66.6	89.2
	M	10	46			

Note: CM: Confusion matrix, B: Benign , M: Malignant, OCA: Overall classification accuracy; ICA$_B$: Individual class accuracy for benign , ICA$_{M:}$ Individual class accuracy for malignant class.

5.4   Concluding Remarks

The exhaustive experiments carried out in the present work indicate it can be concluded that the combined feature vector consisting of MFV + TFV consisting of FOS features yield the highest OCA of 89.6 % and the ICA values of 71.6 % and 94.6 % for benign and malignant class respectively.

CAD SYSTEM DESIGN FOR BREAST LESIONS CLASSIFICATION USING THE DIFFERENCE VECTOR OF STATISTICAL FEATURES

6.1 Introduction

The differential diagnosis between breast lesions from ultrasound images is a daunting challenge even for the experienced radiologists. In light of this fact, a CAD system design is proposed in this chapter to evaluate the performance of SVM classifier for two-class lesions classification using the difference vector of statistical features.

6.2 Proposed CAD System Design

The block diagram of the proposed CAD system design for two-class breast lesions classification using the difference vector of statistical features is shown in Figure 6.1. The approach is implemented on the 130 ultrasound database having 54 cases of benign and 76 cases of malignant class. A variable size rectangular area of interest is taken from inside and outside the lesion. The outside area of interest (*OAOI*) contains the lesion and some surrounding tissue where as the inside area of interest (*IAOI*) contains the area inside the lesion as shown in Fig[6.2-6.7]

Figure 6.2 (a): Sample of Benign case
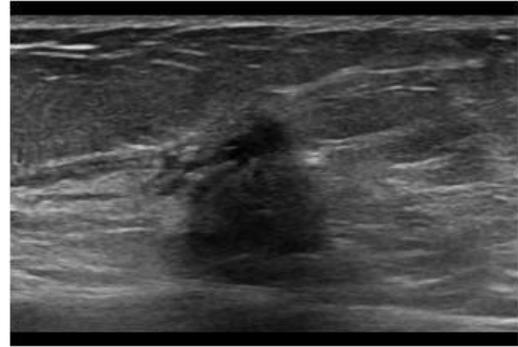
Figure 6.2 (b): Sample of Malignant case
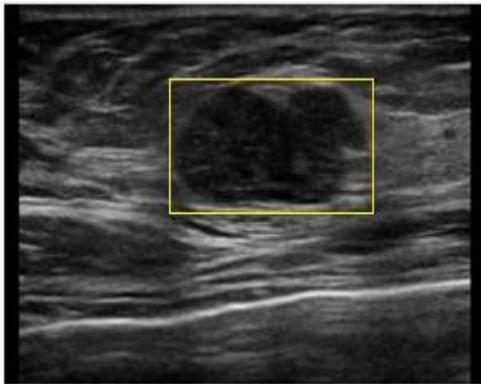
Figure 6.3 (a) : Outside Area of Interest (OAOI) of Benign case

Figure 6.3 (b): Inner Area of Interest (IAOI) of Benign case
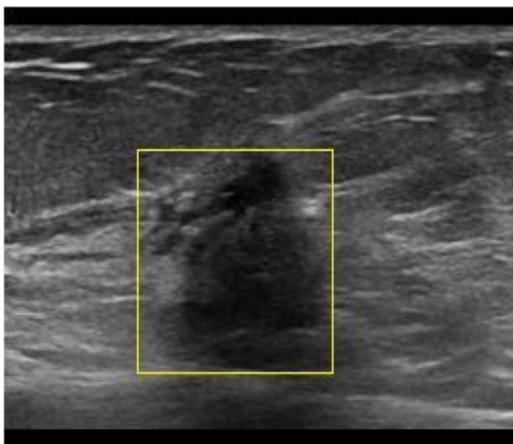
Figure 6. 4(a) : Outside Area of Interest (OAOI) of Malignant Lesion
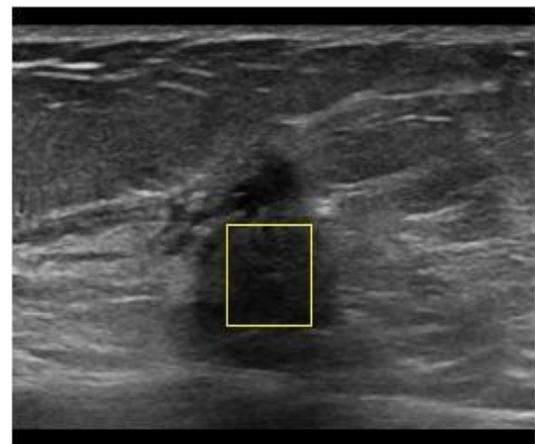
Figure 6. 4 (b) : Inner Area of Interest (IAOI) of   Malignant case

6.3   Experimental Workflow and Results

For evaluating the performance of the proposed CAD system design, rigorous experimentation has been carried out for the characterization of the breast lesions. A brief description of experiments is given in Table 6.1

Table 6.1: Description of experiments carried out for breast lesions classification

Experiment 1 : To obtain the classification performance of texture features obtained from IAOI

Experiment 2 : To obtain the classification performance of texture features obtained from OAOI

Experiment 3: To obtain the classification performance of morphological features and statistical features for two-class breast tissue density classification using SVM classifiers.

6.3.1. Experiment carried out for IOAI for two-class breast lesions classification

In this experiment, classification performance of Inner Area Of Interest (*IAOI*) containing different statistical features is evaluated for two-class breast lesions. The results of the experiment are shown in Table 6.2. It can be concluded from the table that for statistical features, the overall classification accuracy of 65% is achieved using the SVM classifier. It can also be concluded that the highest in individual class accuracy for Benign class is 54.1 % with GLDS feature and highest individual class accuracy for Malignant is 88.1 %, using GLDS feature model.

Table 6.2: Classification performance of texture features obtained from IOAI

Features	CM			OCA (%)	ICA_B(%)	ICA_M (%)
		B	M			
EDGE	B	9	15	45	37.5	50
	M	28	18			
SFM	B	11	13	65	45.8	77.7
	M	8	28			
NGTDM	B	8	16	65	33.3	86.1
	M	5	31			

Features	CM			OCA	ICAB	ICAM
FOS	B	7	17	60	29.1	80.5
	M	7	29			
GLCM	B	12	12	48.3	50	47.3
	M	19	17			
GLRLM	B	3	21	56.6	12.5	86.1
	M	5	31			
GLDS	B	13	11	55	54.1	88.1
	M	4	32			

Note: CM :Confusion matrix , OCA : Over all classification accuracy , B: benign class , M: Malignant Class , ICA_B: Individual class accuracy of Benign class , ICA_M : Individual class accuracy of Malignant class, FOS : First order statistics , GLCM : Gray length co-occurrence matrix , GLRLM : Gray level run length matrix , GLDS: Gray level difference statistics SFM : Statistical feature matrix , NGTDM : Neighborhood gray tone difference matrix.

6.3.2. Experiment carried out for OAOI for two class-class breast lesions classification

Table 6.3: Classification performance of texture features obtained from OAOI

Features	CM			OCA (%)	ICA_B(%)	ICA_M (%)
		B	M			
EDGE	B	12	12	66.6	50	77.7
	M	8	28			
SFM	B	19	5	75	79.1	72.2
	M	10	26			
NGTDM	B	3	21	58.3	12.5	88.8
	M	4	32			
FOS	B	10	14	46.6	58.3	50
	M	18	18			
GLCM	B	8	16	70	33.3	94.4
	M	2	34			
GLRLM	B	13	11	71.6	54.1	88.8
	M	4	32			
GLDS	B	10	14	51.6	46.6	58.3
	M	15	21			

Note: CM :Confusion matrix , OCA : Over all classification accuracy , B: benign class , M: Malignant Class , ICA_B: Individual class accuracy of Benign class , ICA_M : Individual class accuracy of Malignant

class, FOS : First order statistics , GLCM : Gray length co-occurrence matrix , GLRLM : Gray level run length matrix , GLDS: Gray level difference statistics SFM : Statistical feature matrix , NGTDM : Neighborhood gray tone difference matrix.

In this experiment, the classification performance of Outer Area Of Interest (OAOI) containing different statistical features is evaluated for two-class breast lesions classification using different classifiers. The results are shown in Table 6.3. It can be concluded from the table that the overall classification accuracy of 75 % is achieved using the SFM feature model. The highest individual class accuracy of 79.1 for primary benign class is achieved using SFM feature model and for primary malignant class the highest individual class accuracy achieved is 94.4 % using GLCM feature model.

6.3.3. Experiment carried out for two-class breast lesions classification using difference of statistical features of OAOI and IAOI.

Table 6.4: Classification performance of difference statistical features of OAOI and IAOI

Features	CM			OCA (%)	ICA$_B$(%)	ICA$_M$ (%)
EDGE		B	M	60	79.1	47.2
	B	19	5			
	M	19	17			
SFM	B	13	17	45	54.1	38.8
	M	22	14			
NGTDM	B	4	20	61.6	16.6	91.6
	M	3	33			
FOS	B	7	11	53.3	29.1	69.4
	M	11	25			
GLCM	B	9	15	66.6	37.5	86.1
	M	5	31			
GLRLM	B	17	7	85	70.8	94.4
	M	2	34			
GLDS	B	11	13	59.7	45.8	80.5
	M	7	29			

Note: CM :Confusion matrix , OCA : Over all classification accuracy , B: benign class , M: Malignant Class , ICA$_B$: Individual class accuracy of Benign class , ICA$_M$: Individual class accuracy of Malignant class, FOS : First order statistics , GLCM : Gray length co-occurrence matrix , GLRLM : Gray level run length matrix , GLDS: Gray level difference statistics SFM : Statistical feature matrix , NGTDM : Neighborhood gray tone difference matrix.

In this experiment, the classification performance of difference of statistical features of OAOI and IAOI is using different classifiers. The results are shown in Table 6.4. It can be concluded from the table that the overall classification accuracy of 85 % is achieved using difference of statistical features of OAOI and IAOI. The highest individual class accuracy for benign class is 79.1 % using EDGE feature model and for the malignant class 94.4 % is achieved using the GLRLM feature model

6.4   Concluding Remarks

From the results obtained in this chapter, it can be concluded that for two-class breast lesions, the difference vector of GLRLM feature model of OAOI and IAOI achieves highest classification accuracy of 85%

CONCLUSION AND FUTURE SCOPE

The present work has been carried out for two-class breast lesions classification of ultrasound cases. Accordingly, different CAD system designs have been proposed in the present work for two-class breast lesion classification to provide radiologists with a second opinion tool.

7.1 Conclusion- Design of an Efficient CAD System for Two-Class Breast Lesions Classification

For designing an efficient CAD system for two-class breast tissue density classification, various CAD system designs based on Morphological features , Statistical features, Laws' TEMs and Multi resolution texture features have been proposed in the present work. The performance of these CAD system designs has been compared in Table 7.1

Table 7.1: Performance comparison of CAD system designs for two-class breast lesions

TFV	Classifier	CAD Design	OCA(%)
Statistical Features	K-NN	CAD system design based on statistical features	83.1%
Transform Domain Features	PNN	CAD system design based on transform domain features	88.3%
Laws' Mask Feature of length 5	SVM	CAD system design based on Laws' Mask features	97.1%
Shape Based Features	SVM	CAD system design based on shape features	81.8%
Shape + FOS	SVM	CAD system design based on combination of Shape and First order statistics features	88.3%
Difference vector of GLRLM	SVM	CAD system design based on Difference vector of GLRLM of OAOI and IAOI	85%

Note: TFV : Texture feature Vector   OCA : Overall Classification Accuracy   FOS: First order Statistics   OAOI : Outer of interest IAOI : Inner Area of Interest

From the above table it can be concluded that CAD system design based on Laws' Mask texture features of dimension 5 computed achieves maximum classification of 97.1 % out of all the proposed CAD system designs for two-class breast lesions. It can that be concluded that the Laws' Mask Features are thus most efficient texture features to account for the textural changes exhibited by benign and malignant lesion when fed to SVM classifier for predicting the labels unknown testing instances of ultrasound images.

7.2 Limitations and Future Scope

The limitation of the present work is that it has been carried out on the database that consists of digitized ultrasound images and not real data. Following are the recommendations for future work:

(*i*) The present work has been carried out on images developed using ultrasound as the imaging modality; however images acquired from MRI can also be used in the future to test the proposed algorithms.

(*ii*) In the present work, ROIs from the ultrasound images are extracted manually. An algorithm for automatic ROI extraction can be developed by employing various pattern recognition concepts to identify the center of the breast lesions and then extract an ROI of some specified size automatically.

(*iii*) The performance of the proposed algorithms remains to be tested on images having ratio of area present inside the lesion and area present outside the lesion.

PUBLICATIONS FROM THE PRESENT WORK

Referred Journal Publications

1) Sahil Bhusri , Shruti Jain and Jitendra Virmani , "Breast Lesion Classification using amalgamation of Morphological and Texture Features", International Journal of Pharma and Bio Sciences, 2016 April; 7(2): (B) 617 – 624

2) Sahil Bhusri , Shruti Jain and Jitendra Virmani , "Classification of breast lesions using the difference of statistical features" Research Journal of Pharmaceutical , Biological and Chemical Sciences (RJPBCS) , 2016 July; 7(4)

Referred Conference Publications

1) Sahil Bhusri ,Shruti Jain and Jitendra Virmani , "Classification of Breast Lesions using Laws' Mask Texture features", in Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference March 2016 : 2528-2532

REFERENCES

[1] *Cancer Research UK* [online] ,2016 [cited 2016 ,March]
   http://www.cancerresearchuk.org/about-cancer/what-is-cancer/how-cancer-starts

[2] What is cancer? *MNT Knowledge Center*, [online], http://www.medicalnewstoday.
   com/info/cancer-oncology/ (Accessed: 6 December 2014).

[3] Breast cancer awareness month in October, *World Health Organisation*, [online]
   2012, http://www.who.int/cancer/events/breast_cancer_month/en/ (Citied: 6 December
   2014).

[4] Cancer stats: key stats, *Cancer Research UK*, [online], http://www.cancerresearch
   uk.org/cancer-info/cancerstats/keyfacts/ (Accessed: 8 December 2014).

[5] Breast Cancer facts and figures 2015-16, *American Cancer Society,* [online] 2015,
   http://www.cancer.org/acs/groups/content/@research/documents/document/acspc-
   046381.pdf

[6] J.N. Wolfe, "Breast patterns as an index of risk for developing breast cancer",
   American Journal of Roentgenology, vol. 126, pp.1130-1137, 1976.

[7] N.F. Boyd, L.J. Martin, S. Chavez, A. Gunasekara, A. Salleh, O. Melnichouk,
   M.Yaffe, C. Friedenreich, S. Minkin and M. Bronskill, "Breast tissue composition
   and other risk factors for breast cancer in young women: a cross sectional study",
   Lancet Oncology, vol. 10, pp. 569-580, 2009.

[8] H.M. Zonderland ."The role of ultrasound in diagnosis of breast cancer," *Semin
   Ultrasound CT MR* . ,vol. 21 , pp. 317-324 , 2000

[9] Minavathi , Murali .S , M.S.Dinesh , " Classification of Mass in breast Ultrasound
   Images Using Image Processing Techniques " International Journal of Computer
   Applications , vol. 42, issue 10 ,pp. 29-36 , 2012

[10] *MO Obajimi, OO Akute, AO Afolabi, AA Adenipekun, AO Oluwasola, EEU
   Akang, RU Joel, ATS Adeniji-Sofoluwe, O Funmi, G Newstead, R Schmitt, C Sennet
   ,''BI-RADS Lexicon : An Urgent Call for the standardization of Breast Ultrasound in
   Nigeria'' Annals of Ibadan Postgraduate Medicine* Vol. 3 (1) : pp. 82-88 , 2005

[11] Breast ultrasound , [online]   *http://www.webmd.com/breast-cancer/guide/breast-*

ultrasound

[12] Understanding your mammogram report-BI-RADS categories, *American Cancer Society*, [online], http://www.cancer.org/healthy/findcancerearly/examandtest descriptions/mammogramsandotherbreastimagingprocedures/mammograms-and-other-breast-imaging-procedures-mammo-report (Accessed: 15 December 2014).

[13] Y. L. Huang, D. R. Chen, Y. K. Liu, Breast cancer diagnosis using image retrieval for different ultrasonic systems ,in: Proceedings of IEEE ICIP,2004,pp.2957– 2960

[14] P. Miller and A. Astley, "Classification of breast tissue by texture analysis", *Image and Vision Computing*, vol. 10, pp. 277-282, 1992.

[15] Ultrasound Cases [online] , 2015 http://www.ultrasoundcases.info/category.aspx?cat=67.

[16] *Image Processing and Analysis in JAVA ,*Image J 1.49 vesion1.6.024 *http://imagej.nih.gov/ij/download/win32/ij149-jre6-64.zip*

[17] A. V. Alvarenga, A. F. C. Infantosi, W. C. A. Pereira, and C. M. Azevedo, "Assessing the performance of morphological parameters in distinguishing breast tumors on ultrasound images," *Medical engineering & physics*, vol. 32, no. 1, pp. 49–56, 2010

[18] T. Wan, R. Liao, and Z. Qin, "A robust feature selection approach using low rank matrices for breast tumors in ultrasonic images," in *18th IEEE International Conference on Image Processing*, 2011.

[19] R. Liao, T. Wan, and Z. Qin, "Classification of Benign and Malignant Breast Tumors in Ultrasound Images Based on Multiple Sonographic and Textural Features," in *International Conference on Intelligent Human-Machine Systems and Cybernetics*, 2011, vol. 1, pp. 71–74.

[20] W. Gomez, W. C. de A. Pereira, A. F. C. Infantosi, and A. Diaz-Perez, "Computerized diagnosis of breast lesions on ultrasonography," in *XXII   Congresso Brasileiro de Engenharia Biomedica (CBEB)*, 2010.

[21] X. Shi, H. D. Cheng, L. Hu, W. Ju, and J. Tian, "Detection and classification of masses in breast ultrasound images," *Digital signal processing*, vol. 20, no. 3, pp. 824–836, 2010.

[22] Huang YL, Chen DR (2005) Support vector machines in sonography: application to decision making in thediagnosis of breast cancer. Clin Imaging 29(3):179–184

[23] Gómez W, Pereira WCA, Infantosi AFC (2012) Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound. IEEE Trans Med Imaging 31(10):1889–1899

[24] Radhakrishnan M, Kuttiannan T (2012) Comparative Analysis of Feature Extraction Methods for the Classification of Prostate Cancer from TRUS Medical Images. IJCSI Int J Comput Sci Issues 9(1):1694–0814

[25] Huang YL, Chen DR, Jiang YR, Kuo SJ, Wu HK, Moon WK (2008) Computer aided diagnosis using morphological features for classifying breast lesions on ultrasound. Ultrasound Obstet Gynecol 32(4):565–572

[26] Wu WJ, Moon WK (2008) Ultrasound breast tumor image computer-aided diagnosis with texture and morphological features. Acad Radiol 15(7):873–880

[27] Wu WJ , Lin SW , Moon WK (2012) Combining support vector machine with genetic algorithm to classify ultrasound breast tumor images. Computer Med Imaging Graph 36(8):627–633

[28] Alvarenga AV, Infantosi AFC, Pereira WC, Azevedo CM (2012) Assessing the combined performance of texture and morphological parameters in distinguishing breast tumors in ultrasound images. Med Phys 39(12):7350–7358

[29] J. Suckling, J. Parker, D.R. Dance, S. Astley, I. Hutt, C.R.M. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S.L. Kok, P. Taylor, D. Betal and J. Savage, "The mammographic image analysis society digital mammogram database", in *Digital Mammography*, E.G. Gale et al., Eds., Berlin, Heidelberg: Springer, 1994, vol.1069, pp. 375-378.

[30] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, "A rapid approach for prediction of liver cirrhosis based on first order statistics", in *Proceedings of the IEEE International Conference on Multimedia, Signal Processing and Communication Technologies, IMPACT-2011*, Aligarh, India, 2011, pp. 212-215.

[31] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, "Prediction of cirrhosis based on singular value decomposition of gray level co-occurrence matrix and a neural network classifier", in *Proceedings of Development in E-systems Engineering, DeSE*, Dubai, 2011, pp. 146-151.

[32] M. Vasantha, V. Subbiah Bharathi and R. Dhamodharan, "Medical image feature extraction, selection and classification", *International Journal of Engineering*

Science and Technology, vol. 2, pp. 2071-2076, 2010.

[33] P. Mohanaiah, P. Sathyanarayanam and L. Gurukumar, "Image texture feature extraction using GLCM approach", *International Journal of Scientific and Research Publications*, vol. 3, pp. 1-5, 2013.

[34] D.H. Xu, A.S. Kurani, J.D. Furst and D.S. Raicu, "Run-length encoding for volumetric texture", *Heart*, vol. 27, pp. 25-30, 2004.

[35] F. Albregtsen, "Statistical texture measures computed from gray level run length matrices", *Image*, vol.1, pp. 3-8, 1995.

[36] G. Castellano, L.Bonilha, L.M. Li and F. Cendes, "Texture analysis of medical images", *Clinical Radiology*, vol.59, pp. 1061-1069, 2004.

[37] M. Amadasun and R. King, "Textural features corresponding to textural properties", *IEEE Transactions on Systems, Man and Cybernetics*, vol. 19, pp. 1264-1274, 1989.

[38] J.S. Weszka, C.R. Dyer and A. Rosenfeld, "A comparative study of texture measures for terrain classification", *IEEE Transactions on Systems, Man and Cybernetics*, vol. 6, pp. 269-285, 1976.

[39] J.K. Kim and H.W. Park, "Statistical textural features for detection of microcalcifications in digitized mammograms", *IEEE Transactions on Medical Imaging*, vol. 18, 1999.

[40] K.I. Laws, "Rapid texture identification", in *Proceedings of SPIE Image Processing for Missile Guidance*, 1980, pp. 376-380.

[41] M. Rachidi, A. Marchadier, C. Gadois, E. Lespessailles, C. Chappard and C. L. Benhamou, "Laws' masks descriptors applied to bone texture analysis: an innovative and discriminant tool in osteoporosis", *Skeletal Radiology*, vol. 37, pp. 541-548, 2008.

[42] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, "Characterization of primary and secondary malignant liver lesions from B-mode ultrasound", *Journal of Digital Imaging*, vol. 26, pp. 1058-1070, 2008.

[43] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, "Prediction of cirrhosis from liver ultrasound B-mode images based on Laws' mask analysis", in *Proceedings of IEEE International Conference on Image Information Processing*, *ICIIP-2011*. Himachal Pradesh, India, 2011, pp. 1-5.

[44] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, "Neural network ensemble based CAD system for focal liver lesions from B-mode ultrasound", *Journal of Digital Imaging*, vol. 27, pp. 520-537, 2014.

[45] G.H. Seng, H.Y. Chai and T.T. Swee, "Research on Laws' mask texture analysis system reliability", *Reasearch Journal of Applied Sciences, Engineering and Technology*, vol. 7, pp. 4002-4007, 2014.

[46] J.G. Daugman, "An information-theoretic view of analog representations in the striate cortex", in *Computational* Neuroscience, E.L. Schwartz, Ed. Cambridge: MIT Press, 1990, pp.403-424.

[47] H. Yoshida, D.D. Casalino, B. Keserci, A. Coskun, O. Ozturk and A. Savranlar, "Wavelet packet based texture analysis for differentiating between benign and malignant liver tumors in ultrasound images", *Physics in Medicine and Biology*, vol. 48, pp. 3735-3753, 2003.

[48] X. Li, Z. Tian, "Wavelet energy signature: comparison and analysis", in *Neural Information Processing*, I. King, J. Wang, L.W. Chan and D.L. Wang, Eds., Heidelberg, Berlin: Springer, 2006, vol. 4233, pp. 474-480.

[49] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal N, "SVM-based characterization of liver ultrasound images using wavelet packet texture descriptors", *Journal of Digital Imaging*, vol. 26, pp. 530-543, 2012.

[50] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, "Prediction of liver cirrhosis based on multiresolution texture descriptors from B–mode ultrasound", *International Journal of Convergence Computing*, vol. 1, pp. 19-37, 2013.

[51] C.C. Lee, and S.H. Chen, "Gabor wavelets and SVM classifier for liver diseases classification from CT images", in *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, Taipei, Taiwan, 2006, pp. 548-552.

[52] S.R. Amendolia, G. Cossu, M.L. Ganadu, B. Galois, G.L. Masala and G.M. Mura, "A comparative study of k-nearest neighbor, support vector machine and multi-layer perceptron for thalassemia screening", *Chemometrics and Intelligent Laboratory Systems*, vol. 69, pp. 13-20, 2003.

[53] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, "PCA-SVM based CAD system for focal liver lesion using B-mode ultrasound images", *Defence Science Journal*, vol. 63, pp. 478-486, 2013.

[54] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, "A comparative study of computer-aided classification systems for focal hepatic lesions from B-mode ultrasound, *Journal of Medical Engineering and Technology*, vol. 37, pp. 292-306,

2013.

[55] A. Yazdani, T. Ebrahimi and U. Hoffmann, "Classification of EEG signals using dempster shafer theory and a *k*-nearest neighbor classifier", in *Proceedings of 4th International IEEE EMBS Conference on Neural Engineering*, Antalya, Turkey, 2009, pp. 327-330.

[56] Y. Wu, K. Ianakiev and V. Govindaraju, "Improved kNN classification", *Pattern Recognition*, vol. 35, pp. 2311-2318, 2002.

[57] M.L. Zhang and Z.H. Zhou, "A kNN based algorithm for multilabel classification", in *Proceedings of IEEE International Conference on Granular Computing*, Beijing, China, 2005, vol. 2, pp. 718-721.

[58] D.F. Specht, "Probabilistic neural networks", *Neural Networks*, vol.1, pp. 109-118, 1990.

[59] Shruti Jain, Pradeep K. Naik, Sunil V. Bhooshan, "Non linear Modeling of cell survival/ death using Aritificial neural network",pp 565-568, Oct 07-09, 2011, International Conference on Computational Intelligence and Communication Networks (CICN2011),Gwalior, India

[60] D.F. Specht and H. Romsdahl, "Experience with adaptive probabilistic neural network and adaptive general regression neural network", in *Proceedings of the IEEE International Conference on Neural* Networks, Orlando, Florida, 1994, vol. 2, pp. 1203-1208.

[61] V.L. Georgiou, N.G. Pavlidis, K.E. Parsopoulos and M.N. Vrahatis, "Optimizing the performance of probabilistic neural networks in a bioinformatics task", in *Proceedings of the EUNITE 2004 Conference*, 2004, pp. 34-40.

[62] C.C. Chang and C. J. Lin, "LIBSVM, A library of support vector machines".

[63] J. Virmani, V. Kumar, N. Kalra and N. Khandelwal, "SVM based characterization of liver cirrhosis by singular value decomposition of GLCM matrix", *International Journal of Artificial Intelligence and Soft Computing*, vol. 3, pp. 276-296, 2013.

[64] A.E. Hassanien, N.E. Bendary, M. Kudelka and V. Snasel, "Breast cancer detection and classification using support vector machines and pulse coupled neural network", in *Proceedings of 3rd International Conference on Intelligent Human Computer Interaction IHCI 2011,* Prague, Czech Republic, 2011, pp. 269- 279.

[65] A.T.Azar and S.A. El-Said, "Performance analysis of support vector machine classifiers in breast cancer mammography recognition", *Neural Computing and*

Applications, vol. 24, pp. 1163-1177, 2014.

[66] S.W. Purnami, A. Embong, J.M. Zain and S.P. Rahayu, "A new smooth support vector machine and its applications in diabetes disease diagnosis", *Journal of Computer Science*, vol. 5, pp. 1003-1008, 2009.

[67] Y.J. Lee and O.L. Mangasarian, "SSVM: a smooth support vector machine for classification", *Computational Optimization and Applications*, vol. 20, pp. 5-22, 2001

TEXTURE FEATURES USED IN THE PRESENT WORK

A.1. Statistical features

A.1.1 First Order Statistics

For the individual pixel values x_i, the computed features are given as·

$$\text{Mean} = \frac{1}{N}\sum_i i\, x_{i,} \tag{A.1}$$

$$\text{Standard Devaition} = \frac{(\sum_i (x_i - \bar{x})^2}{\sqrt{N-1}} \tag{A.2}$$

$$\text{Third Moment} = \frac{\sum_i (x_i - \bar{x})^3}{N\sigma^3} \tag{A.3}$$

$$\text{Uniformity} = \sum_i p(i)^2 \tag{A.4}$$

$$\text{Entropy} = -\sum_i p(x_i) \log p(x_i) \tag{A.5}$$

$$\text{Smoothness} = 1 - \frac{1}{1+\sigma^2} \tag{A.6}$$

A.1.2 GLCM Features

$$\text{Angular Moment} = \sum_{i,j} P_{i,j}{}^2 \tag{A.7}$$

$$\text{Contrast} = \sum_{i,j} P_{i,j}\ (i-j)^2 \tag{A.8}$$

$$\text{Correlation} = \sum_{i,j} P_{i,j} \left[\frac{(i-\mu_i)(j-\mu_j)}{\sigma_i \sigma_j} \right] \tag{A.9}$$

$$\text{Variance} = \sum_{i,j} P_{i,j}(i - \mu_i)^2 \qquad (A.10)$$

$$\text{Inverse Difference Moment} = \sum_{i,j} \frac{P_{i,j}}{1 + (i - j)^2} \qquad (A.11)$$

$$\text{Sum Entropy} = -\sum_{i=2}^{2Ng} p_{x+y}(i)\, log(p_{x+y}(i)) \qquad (A.12)$$

$$\text{Entropy} = -\sum_{i,j} p_{i,j}\, log(p_{i,j}) \qquad (A.13)$$

$$\text{Difference Entropy} = -\sum_{i=2}^{2Ng} p_{x-y}(i)\, log(p_{x-y}(i)) \qquad (A.14)$$

$$\text{Difference Variance} = -\sum_{i=0}^{Ng-1} (i - f_6)^2 p_{x-y}(i) \qquad (A.15)$$

Where $f_6 = \sum_{i,j} |i - j| p_{i,j}$

A.1.3 .GLRLM Features

$$\text{Short Run Emphasis} = \sum_{i=1}^{G} \sum_{j=1}^{R} \frac{p(i,j|\theta)}{j^2} \bigg/ \sum_{i=1}^{G} \sum_{j=1}^{R} p(i,j|\theta) \qquad (A.16)$$

$$\text{Long Run Emphasis} = \sum_{i=1}^{G} \sum_{j=1}^{R} p(i,j|\theta) * j^2 \bigg/ \sum_{i=1}^{G} \sum_{j=1}^{R} p(i,j|\theta) \qquad (A.17)$$

$$\text{Low Gray Level Run Emp} = \sum_{i=1}^{G} \sum_{j=1}^{R} \frac{p(i,j|\theta)}{i^2} \bigg/ \sum_{i=1}^{G} \sum_{j=1}^{R} p(i,j|\theta) \qquad (A.18)$$

$$\text{High Gray level Run Emp.} = \sum_{i=1}^{G} \sum_{j=1}^{R} p(i,j|\theta) * i^2 \bigg/ \sum_{i=1}^{G} \sum_{j=1}^{R} p(i,j|\theta) \qquad (A.19)$$

$$\text{Short Run High Gray Emp} = \sum_{i=1}^{G} \sum_{j=1}^{R} \frac{p(i,j|\theta) \times i^2}{j^2} \bigg/ \sum_{i=1}^{G} \sum_{j=1}^{R} p(i,j|) \qquad (A.20)$$

$$\text{Short Run Low Gray Emp} = \sum_{i=1}^{G}\sum_{j=1}^{R}\frac{p(i,j|\theta)}{i^2 \times j^2} \Big/ \sum_{i=1}^{G}\sum_{j=1}^{R} p(i,j|\theta) \qquad \text{(A.21)}$$

$$\text{Long Run High Gray Emp} = \sum_{i=1}^{G}\sum_{j=1}^{R}\frac{p(i,j|\theta) \times j^2}{i^2} \Big/ \sum_{i=1}^{G}\sum_{j=1}^{R} p(i,j|\theta) \qquad \text{(A.22)}$$

$$\text{Gray level non Uniformity} = \sum_{i=1}^{G}\left(\sum_{j=1}^{R} p(i,j|\theta)\right)^2 \Big/ \sum_{i=1}^{G}\sum_{j=1}^{R} p(i,j|\theta) \qquad \text{(A.23)}$$

$$\textit{Run Leng non Uniformity} = \sum_{j=1}^{R}\left(\sum_{i=1}^{G} p(i,j|\theta)\right)^2 \Big/ \sum_{i=1}^{G}\sum_{j=1}^{R} p(i,j|\theta) \qquad \text{(A.24)}$$

$$\textit{Run Percentage} = \sum_{i=1}^{G}\sum_{j=1}^{R} p(i,j|\theta)/n \qquad \text{(A.25)}$$

A.1.4 GLDS Features

$$Homogenity = \sum_{i,j}\frac{P_{i,j}}{1+(i-j)^2} \qquad \text{(A.26)}$$

$$Contrast = \sum_{i,j} P_{i,j}\ (i-j)^2 \qquad \text{(A.27)}$$

$$Energy = \sqrt{P_{i,j}^2} \qquad \text{(A.28)}$$

$$Entropy = -\sum_{i,j} P_{i,j}\ log\ (P_{i,j}) \qquad \text{(A.29)}$$

$$Mean = \frac{1}{m}\sum_{i,j} i\ P_{i,j} \qquad \text{(A.30)}$$

A.2. Signal Processing Methods based Features

A.2.1. Laws' Texture Features

Laws' masks of lengths 3, 5, 7 and 9 are used to compute different features. A description of these masks is given below.

Table 1: Description of Laws' masks of different lengths

Length of 1-D filter	1-D filter coefficients	No. of 2D Laws' masks	No. of TR images
3	L3=[1, 2, 1] E3=[-1, 0, 1] S3=[-1, 2, -1]	9	6
5	L5= [1, 4, 6, 4, 1] E5= [-1, -2, 0, 2, 1] S5= [-1, 0, 2, 0, -1] W5= [-1, 2, 0, -2 1] R5= [1, -4, 6, -4, 1]	25	15
7	L7= [1, 6, 15, 20, 15, 6, 1] E7= [-1 -4, -5, 0, 5, 4, 1] S7= [-1, -2, 1, 4, 1, -2, -1]	9	6
9	L9= [1, 8, 28, 56, 70, 56, 28, 8, 1] E9= [1, 4, 4, -4, -10, -4, 4, 4, 1] S9= [1, 0, -4, 0, 6, 0, -4, 0, 1] W9= [1, -4, 4, -4, -10, 4, 4, -4, 1] R9= [1, -8, 28, -56, 70, -56, 28, -8, 1]	25	15

Note: TR: rotation invariant texture images.

As an example Laws' mask of length 5 is used for explanation purposes. The ROIs are convolved with each of the above twenty five 2D Laws' masks.

a) The Texture Image (*TI*) is obtained by convolving the input image I (i, *j*) with the 2-D mask

$$TI_{E5E5} = I_{i,j} \otimes E5E5 \qquad \text{(A.2.1)}$$

b) The contrast of the texture image obtained from above equation   is normalized

$$Normalize\ (TI_{mask}) = \frac{TI_{mask}}{TI_{L5L5}} \qquad \text{(A.2.2)}$$

c) The Texture Energy Measurement (TEM) filters are used to pass the Texture image

$$TEM_{i,j} = \sum_{u=-5}^{5}\sum_{v=-5}^{5} Normalize(TI_{i+u,j+v}) \qquad \text{(A.2.3)}$$

d) To obtain 15 rotationally invariant TEM's that are denoted as *TR* are obtained by collaborating the 25 TEM descriptors

$$TR_{E5L5} = \frac{TEM_{E5L5}+TEM_{L5E5}}{2} \qquad \text{(A.2.4)}$$

e) Five statistical parameters are determined, that are the Mean, Standard Deviation, Skewness, Kurtosis, Entropy. Here M×N is the dimension of the image

1) *Mean (m)*: It describes the mean intensity value with in texture image.

$$Mean = \frac{\sum_{i=0}^{M} \sum_{j=0}^{N} (TR_{i,j})}{M \times N}$$

(A.2.5)

2) Standard Deviation (SD): It is used to measure the variability.

$$SD = \sqrt{\frac{\sum_{i=0}^{M} \sum_{j=0}^{N} (TR_{i,j} - Mean)^2}{M \times N}}$$

(A.2.6)

3) *Skewness:* It measures of the asymmetry of the probability distribution of a random variable that is real valued.

$$Skewness = \frac{\sum_{i=0}^{M} \sum_{j=0}^{N} (TR_{i,j} - Mean)^3}{M \times N \times SD}$$

(A.2.7)

4) *Kurtosis:* It measures of the probability distribution shape of a random variable that is real valued.

$$Kurtosis = \frac{\sum_{i=0}^{M} \sum_{j=0}^{N} (TR_{i,j} - Mean)^4}{M \times N \times SD^4} - 3$$

(A.2.8)

5) *Entropy:* It measures the randomness of the elements of the image.

$$Entropy = \frac{\sum_{i=0}^{M} \sum_{j=0}^{N} (TR_{i,j})^2}{M \times N}$$

(A.2.9)

A.3. Transform Domain based Features

A.3.1. Wavelet based Texture Features

Normalized energy is calculated for each sub image.

For approximate sub image at i^{th} level of decomposition

$$Normalized\ Energy = \frac{||A_i||_F^2}{area\ (A_i)}$$

(A.3.1)

For detailed sub image in k^{th} direction and at i^{th} level of decomposition

$$Normalized\ Energy = \frac{||D_i^k||_F^2}{area\ (D_i^k)}$$

(A.3.2)

A.3.2 FPS Feature

$$Radial\ Sum = \sum_{r1^2 < u^2 + v^2 < r2^2} |F(u,v)|^2 \qquad\text{(A.3.3)}$$

$$Angular\ Sum = \sum_{\theta_1 \leq \tan^{-1}\frac{v}{u} \leq \theta_2} |F(u,v)|^2 \qquad\text{(A.3.4)}$$

A.3.3 Gabor Filter Based Methods

$$Mean = \frac{1}{m} \sum_{i,j} i\ P_{i,j} \qquad\text{(A.3.5)}$$

$$Variance = \sum_{i,j} P_{i,j} (i - \mu_i)^2 \qquad\text{(A.3.6)}$$