SUPPORT VECTOR MACHINE BASED TEXTURE FEATURE EXTRACTION TECHNIQUE FOR CLASSIFICATION OF BREAST CANCER FROM ULTRASOUND IMAGES

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

MASTERS OF TECHNOLOGY

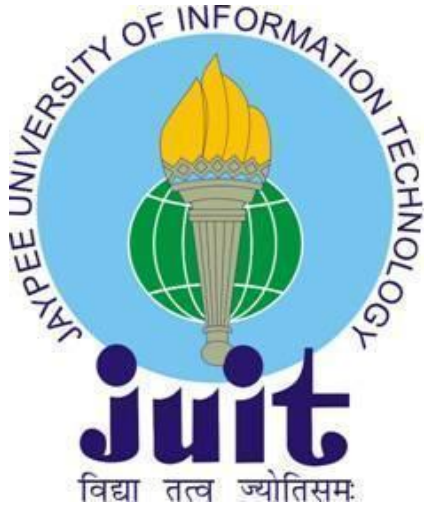
By

Shreya Sharma

Enrollment Number: 152013

Under the Supervision of

Dr. Shruti Jain



DEPARTMENT OF ELECTRONICS AND COMMUNICATIOENGINEERING
JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY
WAKNAGHAT, SOLAN - 173234, INDIA
MAY-2017

TABLE OF CONTENTS

DECLARATION BY THE SCHOLAR

I hereby declare that the work reported in the M-Tech thesis entitled **"Support Vector Machine Based Texture Feature Extraction Technique for Classification of Breast Cancer from Ultrasound Images "** submitted at **Jaypee University of Information Technology, Waknaghat India,** is an authentic record of my work carried out under the supervision of **Dr. Shruti Jain.** I have not submitted this work elsewhere for any other degree or diploma.

Shreya Sharma

Shreya Sharma.

Department of ECE

Jaypee University of Information Technology, Waknaghat , India

Date: 01/05/2017.

iv

JAYPEE UNIVERSITY OF INFORMATIO TECHNOLOGY
(Established by H.P. State Legislative vide Act No. 14 of 2002)
P.O. Waknaghat, Teh. Kandaghat, Distt. Solan - 173234 (H.P.) INDIA
Website: www.juit.ac.in
Phone No. (91) 01792-257999
Fax: +91-01792-245362

CERTIFICATE

This is to certify that the work reported in the M.Tech project report entitled **"Support Vector Machine Based Texture Feature Extraction Technique for Classification of Breast Cancer from Ultrasound Images"** which is being submitted by **Shreya Sharma** in fulfillment for the award of Masters of Technology in Electronics and Communication Engineering by the Jaypee University of Information Technology, is the record of candidate's own work carried out by her under my supervision. This work is original and has not been submitted partially or fully anywhere else for any other degree or diploma.

01/05/17.

Dr. Shruti Jain

Associate Professor
Department of Electronics & Communication Engineering
Jaypee University of Information Technology, Waknaghat

ACKNOWLEDGEMENT

I am highly indebted to Dr. Shruti Jain for her guidance and constant supervision as well as for providing necessary information regarding the project & also for her constant support as we ascend toward the completion of this project .

I would like to express my gratitude towards Mr. Sahil Bhusri for his kind co-operation in spite of being occupied with his work, he managed to guide me and help me with the problems which I faced during the completion of this project.

I would like to express my special gratitude to my family and my friend Aman Deep for their immense support, help and motivation throughout the work. My thanks and appreciations also go to my colleague in developing the project and people who have willingly helped me out with their abilities.

LIST OF ACRONYMS & ABBREVIATIONS

AJCC: American Joint Committee on Cancer

ANN: Artificial Neural Network

BMP: BitMap

BC: Breast Cancer

CAD: Computer Aided Design

CADe: Computer-Aided Detection System

CADx: Computer-Aided Diagnosis System

CM: Confusion Matrix

CR: Carcinoma

CT: Computed Tomography

DNA: Deoxyribonucleic acid

DCIS: Ductal Carcinoma In Situ

DICOM: Digital Imaging and Communications in Medicine

EBCTCG: Early Breast Cancer Trialists' Collaborative Group .

FA: Fibroadinoma

FITS: Flexible Image Transport System

FN : False Negative

FP: False Positives

FPS: Fourier Power Spectrum.

GIF: Graphic Interchange Format

GWT: GaborWavelet Transform

HRT: Hormone Replacement Therapy

HAR:	Hole Area Ratio.
ICA:	Individual Classification Accuracy
IDC:	Invasive Ductal Carninoma.
ILC:	Invasive Lobular Carcinoma.
IGMC:	Indra Gandhi Medical College
JPEG:	Joint Photographic Experts Group
KNN:	K-nearest_neighbors
LCIS:	Lobular Carcinoma In Situ
MATLAB:	Matrix Laboratory
MMTV:	Mouse Mammary Tumor Virus
MRI:	Magnetic Resonance Imaging
MS:	Metastasis
OCA:	Overall Classification Accuracy
PCNN:	Pulse-Coupled Neural Network
PET:	Positron Emission Tomography Scan.
PNG:	Portable Network Graphics
PNN:	Probabilistic Neural Network
ROI:	Region Of Interest
SF:	Statistical Feature.
SNM-RBF:	Support Vector MACHINE Radial Basis Function
SSVM:	Smooth Suport Vector Mechine
SVM:	Support Vector Machines
TD:	Transform Domain Methods.
TFV:	Texture Feature Vector
TIFF:	Tag Image File Format

TNM: Tumor Node Metastases

TP: True Positive.

WPT: Wavelet Packet Transform

LIST OF FIGURES

LIST OF TABELS

ABSTRACT

Breast Cancers is such a ailment that has got excellent attention within many years. In breast cancer the breast lesion are differentiated into two instructions i.e. Benign and Malignant. Computer-Aided Detection (CAD) system is designed to resource radiologists in detecting lesions which could imply the presence of breast cancers. The ROI is extracted from the ultrasonic photos, the usage of *imageJ* software after which the unique photograph processing techniques are carried out i.e. preprocessing, feature extraction and feature classification using MATLAB. SVM classifier is significantly used for category.

Classification of breast ultrasound images using Statistical and Transform domain feature extraction techniques were data is partitioned by hold-out method and classified using Support Vector Machine (SVM) classifier. SVM trains a model that assigns unseen new objects into a specific category. The best obtained result out of all the features used is calculated using Fourier Power Spectrum (FPS) feature.

CHAPTER 1

INTRODUCTION

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. Breast cancer (BC), after the lung cancer is the common cause for deaths in the world. Breast cancer has been divided into two different classes Benign and Malignant. The cells in the breast starts to grow unmanageably, this is how (BC) begins. In many cases tumor are formed because of these cells visualized on a X-Ray or can be sensed as a lump. If the cells grow into the outer portion of tissues or open out to unsuitable portions of the body it is considered malignant. BC may originate from any part of the breast. Maximum cases of BC originates in the ducts carrying milk to the nipple, whereas some starts in the glands where the breast milk is formed .Various other type of BC do exist which are less common. The lumps are not most cancers, so considered as benign. Benign breast tumors are aberrant growths which do not enlarge beyond the breast and are not existence threatening. Some of the benign lumps would possibly boom the danger of BC in ladies. Normal breast cells turn out to be most cancers due to versions in DNA. Some DNA variations are hereditary in the circle of family as well.

Rather than being inherited, transformation takes place in cells of the breast during a woman's life. As we have seen that mostly the lumps are benign i.e. non cancerous. This can be removed surgically and does not infect the surrounding tissues of the body. The malignant types of tumor, if not detected and treated earlier will grow and destroy the healthy nearby cells of the surrounding parts. In this case the cancer breaks out from the tumor and spreads to the blood vessels. In this paper, work is based on two class classification of breast ultrasound images using different Statistical features and Transform Domain Methods. Each set has been normalized by using min-max approach. The normalized values are then partitioned in training and testing datasets by hold-out approach. The classifier used is SVM for the designing CAD system.

There are some symptoms that indicate breast cancer are:

- A lump in a breast
- Pain in the armpits or breast
- A swelling (lump) in one of the armpits
- The size or the shape of the breast changes
- Redness of the skin of the breast
- Menstrual cycle
- Redness, scaliness, or thickening of the nipple or breast skin
- Nipple discharge (other than breast milk)
- Family History and Genetic Susceptibility.

The formation of breast cancer has been shown in the diagram below :

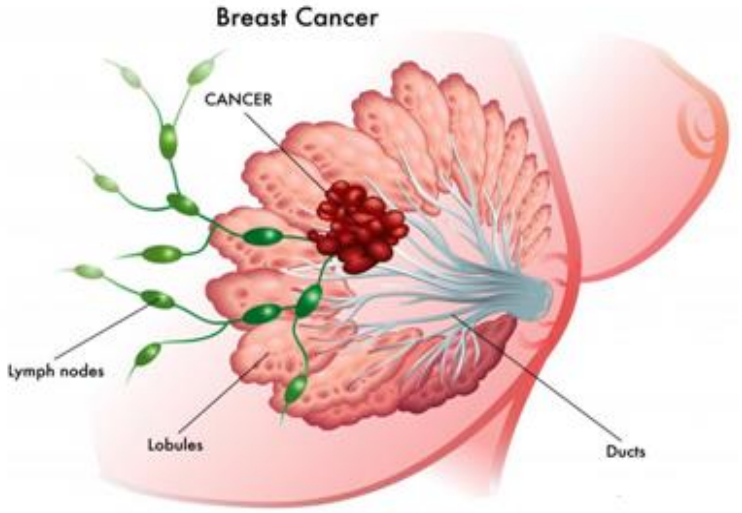


Fig:1.1:FormationOf Breast Cancer[1]

Benign type of breast cancer is the type of cancer in which the cancer is not broken out and is still inside the infected area.

Whereas in Malignant type of cancer the cancer cells break out from the lobule or ducts and infect the nearby healthy cells of the body, making their way to the other organs like neck, bones, liver etc.

Primary occurring benign cases is also called Fibroadenomas. Non-cancer tumor which are made up of glandular breast tissue and connective tissues are called Fibroadenomas. Some tumors(fibroadenomas) are very small and cannot be felt can be seen only if breast

tissues are removed and examined under a microscope. It can usually move under the skin and not tender.

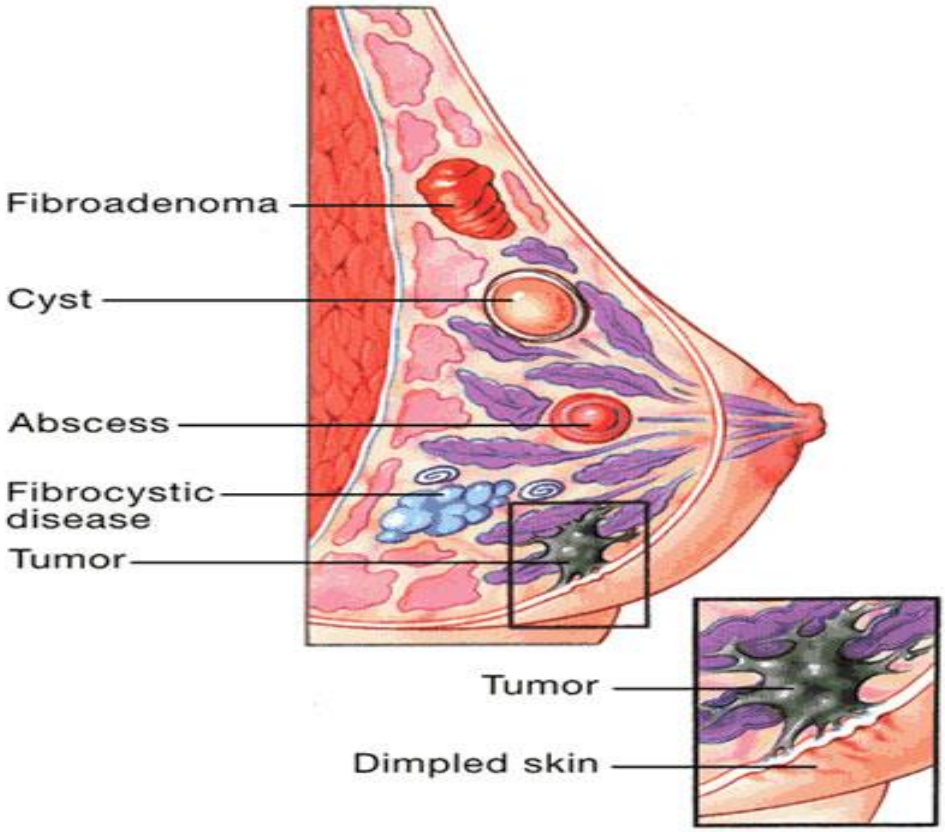


Fig1.2: Fibroadenoma in breast[1]

Malignant type of breast cancer is further divided into two different types: Carcinomas and Sarcomas.

Carcinoma is type of cancer that starts in cells that lines organs and tissues. Whereas Sarcomas are cancers that arises from the connective tissues components.

The primary occurring Malignant tumor is called Carcinoma. Carcinoma occurs most commonly and is divided into three types:

- Ductal Carcinoma in situ (DCIS)
- Invasive Ductal Carcinoma(IDC)
- Invasive Lobular Carcinoma(ILC)

DCIS: The infected cells which are found in the lining of the non invasive cancer of the breast milk ducts.

IDC: Is also called as infiltrating Ductal Carcinoma, This form of most cancers begins to develop in the duct and invade the fatty tissue of the breast outside the duct.

ILC: It is from time to time known as infiltrating lobular carcinoma. Cancer that has damaged through the wall of the lobule and has started invading the tissues of the breast is known as ILC.[2]

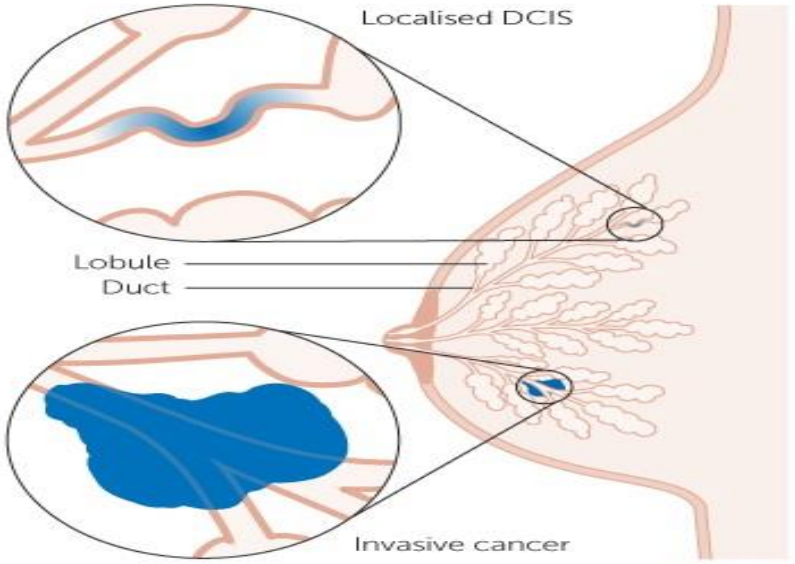


Fig1.3: General view of carcinoma

The secondary occurring Malignant tumor is called Metastatic tumor. As explained above Cancer that has spread to the other parts of the body is called Malignant tumor this includes lungs, liver, bones etc. The cancer cells invade the nearby healthy cells and penetrate into the circulatory or lymph system thus forming new small tumors.

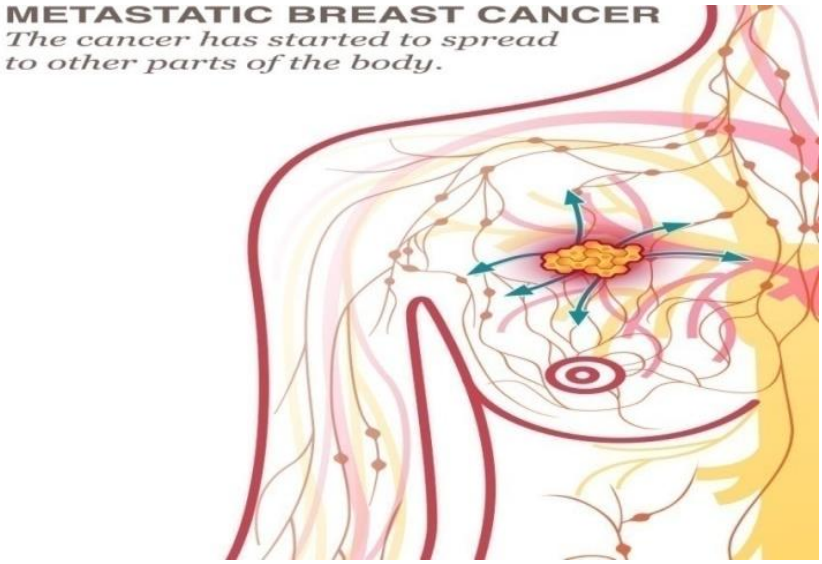


Fig1.4: Metastatic Brest Cancer[2]

Causes of breast cancer may include:

- Genetic Problems
- Getting Older
- Lumps in Breast
- Dense Breast Tissue
- Obesity
- Alcohol Consumption
- Radiation Exposure

Classification of stages of Breast Cancer:

Scale of 0 to IV are used to express the Stages of breast cancer . In order to find out in many cancers mostly those which are likely to spread out in the entire body it is very important to know the stage of the cancer. So for better and accurate treatment it is very important to find out the stages of the cancer, it starts with stage 0,Where stage 0 is non invasive cancer that remains in its original location and stage IV showing invasive cancer, that has spread out to the various parts of breast.

Stage 0: DCIS is used to characterise stage 0 type of breast cancer. Cancer cells are not present in this stage or non-cancerous cells they breaking out of the part of the breast where they started, or invading neighboring normal tissue.

Stage I: The infected cells they invade the nearby healthy cells in surrounding breast tissues in this stage. Stage I is divided into two subparts, stage IA and IB.

Stage IA describes invasive breast cancer in which:

- the tumors measurement is up to two centimeters in length.
- the cancer is still inside the breast.

Stage IB describes invasive breast cancer in which :

- Small group of cancer cells which are larger than 0.2 millimeter but not more than 2 centimeter are found in the lymph nodes.

Microscopic invasion is possible in stage 1 type of breast cancer.

Stage II: stage II is subcategories into two types of stages, stage IIA and IIB.

- **Stage IIA** Describes invasive breast cancers wherein the lumps are two cm or smaller and unfolded to the axillary lymph nodes or while the tumor is more than two cm but not greater than five cm and has no longer unfold to the axillary lymph nodes.

- **Stage IIB** Explains the invasive BC in which tumor is greater than two cm but not greater than five cm. If this happens then the cancer has spread to one to three axillary lymph nodes or the tumor is greater than five centimeters but still is in the axillary lymph nodes.

Stage III : is subcategories into IIIA,IIIB,IIIC

- **Stage IIIA** Explainsthe invasive BC in which the tumor is larger than five cm, small group of cancer cells is in a range of 0.2 millimeter to two millimeters. If the tumor is broadend than five cm, most cancers has spread to one to three axillary lymph nodes or to the lymph nodes close to the breastbone.
- **Stage IIIB** Explains the invasive breast cancer in which the tumor may be any size and has spread to chest walls causing swelling or ulcer. It may spread to nine axillary lymph nodes.
  Inflammatory breast cancer is seen in stage IIIB.

- **Stage IIIC** Explains the invasive breast cancer wherein the cancer has unfold to more than ten axillary lymph nodes or most of the cancer has unfold to lymph nodes above or beneath the collarbone.

Stage IV: Explains invasive breast cancer which has unfold across the breast and close by lymph nodes to different organs of the body. The word “metastatic” describes stage IV BC. Cancer may be of degree IV at first analysis or it can be a recurrence of a older breast cancers that has unfold to other parts of the body.[3]

Various Imaging tests used to evaluate breast disease:

Mammograms

The X-ray of the breast is referred to as the mammogram. To observe the breast modifications in girls who have no symptoms or signs of a breast problem, Screening mammograms are used. Diagnostic mammograms are used to get a closer appearance of a change seen on a screening mammogram. And with the help of diagnostic mammogram greater photos may be taken of the location that can be most cancers.

Breast ultrasound

Sonography additionally known as Ultrasound, which makes use of sound waves to define a part of the body. It can be used for detecting some breast modifications, inclusive of people in whom it can be felt but can not be visible on a mammogram. It also facilitates in telling the distinction among fluid-stuffed cysts and solid masses.

Magnetic resonance imaging (MRI)

Radio waves and robust magnets are used for MRIs rather than x-rays. The absorbed radio wave power is launched in a body by using the sort of frame tissue of positive diseases. A picture with maximum detail is acquired. Gadolinium, a assessment liquid is injected to a vein before the scan to expose higher details in breast MRI for detection of most cancers.

Biopsy

A biopsy is conducted when mammograms and other imaging tests shows change in the breast which may be cancer. To conform the presence of cancer biopsy is done. To conduct a biopsy, samples of the doubtful area are taken and tested. This sample is called biopsy specimen.

Ultrasound imaging, less expensive and is portable, as well is an alternative to mammography, where as the X-ray Mammography is non-radioactive, non-invasive, real time display having low cost and better penetrating capacity.

Need for using CAD

CAD (Computer Aided Design) technology, by increasing sensitivity to rates comparable to those obtained by double reading, in a cost-effective manner it gives a second opinion to the radiologists.

As CAD system is used to improve quality of images such that the radiologists can easily diagnosed the ultrasonic images of carcinoma resulting in better results. CAD systems fulfills the following requirements: it improve the performance of radiologists providing high sensitivity in the diagnosis, a low number of false positives (FP), have high processing speed, present high level of automation, low cost, the ability to detect different shapes of nodules, and software security assurance.

The computational systems developed for assistance of radiologists are: CADe (computer-aided detection system) and CADx (computer-aided diagnosis system). CADe systems is the system in which lesion detection is done through medical images while CADx system is used for the measurement of lesion characterization, for example, determining the malignancy and staging of the cancer . CADx systems are not in scope of this type of work so discarded. CADe systems have the following goals:

1. Improves the accuracy in diagnosis;
2. Helps in early detection of cancer;
3. Decreases the time of the radiologist to conform the results.

CADe systems is an important tools for medical radiology as most of the systems yet do not have all the necessary requirements which can be considered useful by most radiologists.

It Improve the radiologists performance by providing high sensitivity in diagnosis. Sensitivity of such systems are given by the formula:

$$\text{Sensitivity} = \frac{\mathbf{TP(true\ positive)}}{\mathbf{(TP+FN)(true\ positive+false\ negative)}}$$

Where: TP, the system shows positive results to a sample that actually had the disease, and FN,the system shows negative results when the sample had the disease.[3]

As we proceed further we will see in the upcoming chapters about the work that has been done. In chapter 2 literature review has been done of various authors using different methods for detection of Breast Cancer. In chapter 3 all the methods, classifiers are discussed regarding the work done for this project. Where as in chapter 4, Gabor Wavelet for three class classification has been discussed along with its work done in detail showing the results. For chapter 5, Two class classification of Breast Lesion using Statistical and Transform Domain Features has been discussed broadly showing all the literal results followed by conclusion and refrences.

CHAPTER 2

LITERATURE REVIEW

NAME OF PAPER	YEAR	METHOD	CLASSIFICATION	RESULT	NAME OF AUTHOR
Research journal of pharmaceutical,sidogical & chomical Services: Classification of breast lesions using the difference of statistical features.	2016 [5]	Ultrasound image database: 130cases having 54 cases of benign & 76 cases of malignant class.	SVM	85% for differential diagnosis between benign & malignant breast lesion using the ultrasound images with individual class accuracy values of 70.8% & 90.4% for benign & malignant lesion respectively.	sahil bhusri, shruti jain et-al.
Rare breast cancer, clinical presentation, diagnosis & treatment.	2016 [9]	Clinical Data: inclusion criteria were rare patient>18 years with localised breast cancer, locally advanced or metastatic.	KNN	The main symptom was a solid sub areolar mass in 36 cases(76%). Tumor was associated with gynecomastia in two cases(4%)	
Adjuvant bisphosphonate treatment in early breast cancer:Meta-analysis of individual patient data from randomised trials. (EBCTCG)	2015 [40]		KNN	97% of all 19291women in 32 completed trails that recorded recurrance data.	
Presence of insuline like growth factor binding protine correlates with tumor promoting effects of matrix metalloproteinasen in breast cancer.	2015 [8]	MMTV-neu, C3(1)-Tag, Mmp9 All three mice stairs were used on the FVB/n background	SVM		jal-hyunpark et-al.

NAME OF PAPER	YEAR	METHOD	CLASSIFICATION	RESULT	NAME OF AUTHOR
Multidemensional tumor detection in automated whole breast ultrasound using topographic watershed	2014 [32]	1- Data Acquistion:- The ABUS scanner equipped a 1425BV linear array transducer with variable frequency.	SVM	The proposed CAD system was evaluted with an ABUS database including 104 abnormal passes & 34 normal passes.	chung-minglo,senior member IEEE.
Improved mass detection in 3D automated breast ultrasound using region based feature & multi-view info.	2014 [31]	1. Voxel-Based candidate Generation 2. Candidate Segmentation 3. Feature Extraction	KNN	Proposed approach reached senstivity of 95%, 90% &70% with avg. 4.3, 3.8 & 1.6 false +ve per volume.	chuyang ye et-al.
Feature selection for breast cancer detection from ultrasound images	2014 [11]	1. Problem statement 2. Feature Extraction 3. Feature Selection	SVM	The RF data & biopsy- confirmed diagnosis result of a total of 504 breast tumor & 50 are malignant	mohd. Ashiqueridwan nayeem et-al
lesion detection in breatst ultrasound images using tissue trasnition analysis.	2014 [13]	1.candidate genration 2. descriptor computation for candidate pixels	PNN	clinical dataset of 135 images show that the proposed approach can achieve heigh sensitivity 95% with modest false positive per image.	somabiwas et-al.
improvimg classification performance of breast lesions on ultrasonography.	2014 **[33]**	1.BUS dataset 2.BUS segmentaion 3.morphological and texture features. 4.feature selection and classification performance.	SVM	a total of 26 morphological and 1465 texture features were computed from 641 BUS images (413 benign and 228 malignant lesions)	wilfridogomez flores et-al

NAME OF PAPER	YEAR	METHOD	CLASSIFICATION	RESULT	NAME OF AUTHOR
Identification of malignant breast tumors based on acoustic Attenuation mapping of conventional ultra sound images.	2013 [6]	1. Acoustic Attenuationmap estimation 2. Attenuation Analysis	KNN	233 images of 80 different lesions gave the result.	
Evaluation of breast tissue characterisation by ultrasound computer tomography using a 2D/3D image registration with mannugrams.	2013 [42]		SSVM	80% accurate result.	torsten hopp et-al.
A new approach to ultrasonic detection of malignant breast tumors	2013 [43]	1.Data Collection 2.Feature Extraction	KNN	Search was echaustive & was performed with the goal of maximum area under ROC curve.	nishant uniyal et-al.
An automated breast ultrasound system for elastrography.	2012 [44]	Data Acquisition: the imaging done of the ABUS is equipped with a high resolution encoder (8000 steps per 360 degree.	SVM		reza zahiri azar et-al
Computer Aided Lesion Diagnosis in automated 3-D breast ultrasound using coronal spiculation.	2012 [45]	1 Dataset 2 Intensity Normalisation 3 lesion segmentation 4 Feature Computation	SVM	System differentiates benign from malignant breast tumors well.	bram platel et-al.
Multiple domain knowledge based MRF model for tumor segmentation in breast ultrasound images.	2012 [46]	1.The MAP MRF segmentation framework 2.define the posterior energy function 3.construct the freq constraints	SVM	Proposed method is applied to a breat ultrasound database with 131 cases & boundary error metrics.	chung-ming lo,senior member IEEE.

NAME OF PAPER	YEAR	METHOD	CLASSIFICATION	RESULT	NAME OF AUTHOR
Formal Design methods for reliable computer aided diagnosis: A review	2012 [47]	System Engineering	KNN	Modeling is not new for CAD but modeling for system is less explored. So selected system designs are discussed.	oliver faust et-al.
Ultrasonic multifeature analysis procedure for computed aided diagnosis of solid breast lesions	2011 [49]	Data Acquisition Acoustic Feature Morphometric Features	PNN	Sensitivity of Mammograph is significantly less (60%) in women young than 50 year old compared to women older than 50 (86%).	s.kaisar alam et-al.
Breast tumor identification in ultrasound images using the normalised cuts with partial grouping constraints.	2008 [41]	1-Normalised cuts 2-Constrained Normalised cuts 3-Procedures & Implementation	SVM	We applied our method to segments over 500 cropped ultrasonic images & compared the results with manual delieation result obtained by physician	shao-yu chen et-al.
Computer aided diagnosis using morphological features for classifying breast lesions on ultra sound.	2008 [48]	ultrasound images was captured at the largest diameters of each tumor > 1cm. patient age range from 22 to 67.	SVM	CAD system, using all morphological features & low demensions principles vector were 0.90 & 0.91	y.l. huang et-al.
male breast cancer, clinical presentation,diagnosis ant treatment	2016 **[10]**	1.clinical data	SVM	clinical features .	A.sangui netti et-al.
Multifeature gradient vector flow snakes for adaptive segmentation of the ultrasound images of breast cancer.	2013 [38]		PNN	98% accurate result.	annupan rodtook et-al.

CHAPTER 3

METHODOLOGY

Data Collection

In the process of data collection, the data (images) has been collected from various sources. Different types of images can be used in CAD system and these are Mammogram, Breast ultrasound, Magnetic resonance imaging (MRI), computer tomography (CT or CAT) scan, positron emission tomography (PET) scan, Bone scan etc.

Various tests are conducted for various stage. For stage 0, stage I and stage II where the cancer has not yet spread to the other parts of the body CT scan, PET scan and bone scan can be done and rest imaging tests can be done for the other two stages i.e. stage III and stage IV.

Data Pre-processing:

The collected data (images) is been checked at this stage and with the help of the radiologist, cancerous part in an image is conformed. After the conformation of the cancerous part in an image, ROI selection is done in two different methods for benign and malignant types of tumors separately.

The image is taken and the cancerous part is marked with the software imageJ. This software helps to mark and segment the abnormal area. MATLAB: It is a multi-paradigm numerical computing environment. It is also fourth generation language for programming. MALAB works on matrix manipulations, data and function plotting, algorithm implementation, user interface creation and interfacing with other language programs (C, C++, C#, Java, FORTRAN and Python). The MATLAB scripting language is the base of MATLAB application. Common window as an interactive mathematical shell or executing text files containing MATLAB code are common usage of the MATLAB application.

Feature Extraction:

A process which is used for detecting and representing certain feature of interest within an image for further processing is called feature extraction. The data is converted to a

reduced set of feature representations when the input data is very large for processing and imagined to be unnecessary. Features often contain information relative to color, shape, texture or context. Many techniques have been used for features extraction from images. Feature extraction module is divided into two types: Morphological and Texture. The Texture method is divided into three different methods: Signal Processing, Transform Domain and Statistical.

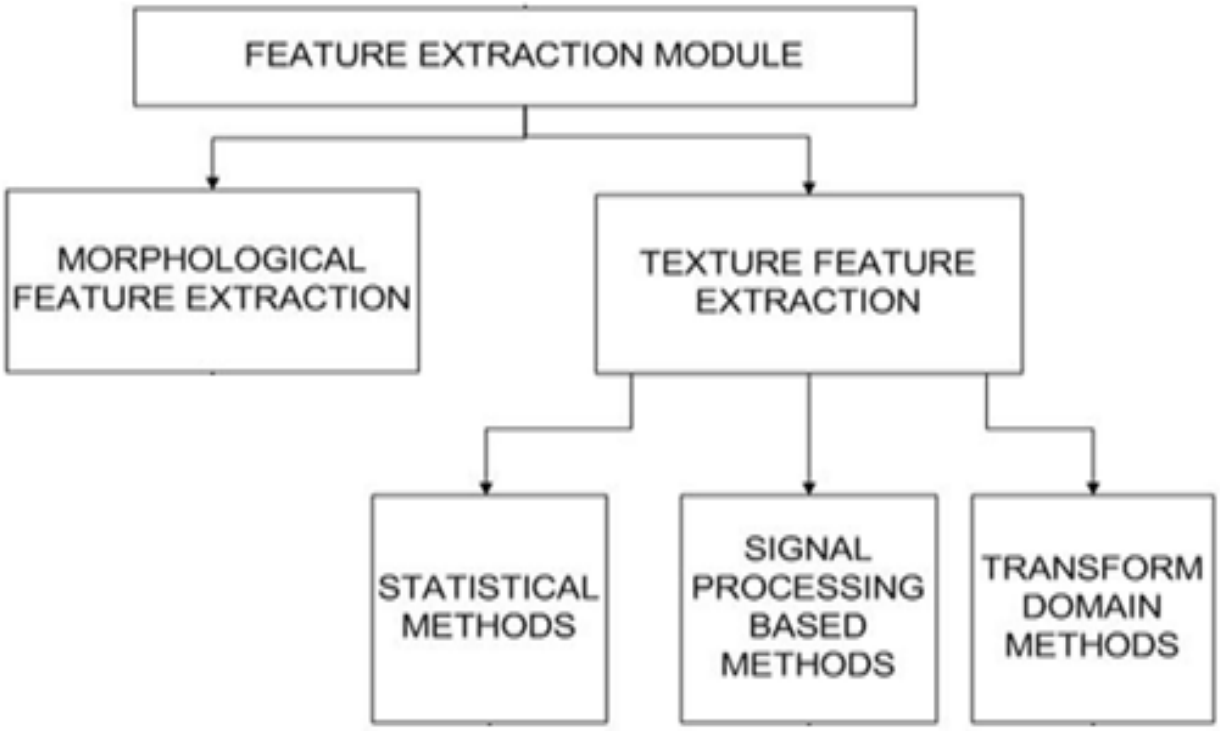


Fig 3.1: Feature extraction module flow chart.

Morphological Feature Extraction: Morphological methods consists of the shape based properties which includes Area, Perimeter, Convexity, Eccentricity, Extent, Hole Area Ratio (HAR) and Solidity are calculated for benign and malignant tumors

- Area: Area of the lesion is calculated.
- Perimeter: Perimeter of the lesion is calculated.
- Diameter: it is the Diameter of the circle having an equivalent area as the region.
- Major Axis and Minor Axis: The diameters of the ellipse, having major axis as the longest diameter where as minor axis are the smallest diameter.

- Convexity: this is a ratio of the overall contour to the perimeter of the convex envelop.
- Eccentricity: Ratio of minor axis to the major axis with value lying between the 0 and 1.
- Extent: Ratio of pixels in the bounding box that are also present in the region.

  Extent = Area / Bounding Area
- Euler No: Relationship between the number of contiguous part to the number of holes in the shape.

Texture Feature Extraction: It is divided into three sub methods namely: statistical method, signal processing based method, transform domain method.

Statistical Method is subdivided into four features:

- First order statistical: The FOS features are extracted from the gray level intensity histograms of an image. We have five different FOS features: average gray level, standard deviation, kurtosis, smoothness and entropy.
- Second Order Statistical-GLCM Feature: Generally how many times combinations of different grey level pixels occur in an image used for second order texture calculations is defined by GLCM. Relation between groups of 2 pixels is considered. These pixels are also known as reference and neighbor pixel. The neighbor pixel is the one right to the reference pixel. Inside a window, every pixel is referred as the reference pixel which is starting at the top left corner and ending towards the lower right hand corner. The GLCM includes features of calculating contrast, entropy, sum entropy, difference entropy, correlation, inverse difference moment, sum average, variance, sum variance, difference variance angular moment etc
- Higher Order Statistical-GLRLM Feature: Way's in which we extract the higher order statistic texture feature is called the GLRL method. Gray level run consists of successive sets of pixels having the same gray level collinear in a accustomed direction. Length run is the number of pixels in run and the count of such a run occurring in an image is called the length run value.

GLRL consists of each element is represented by a two-dimension matrix $p(i, j \mid \theta)$, where j : length run, i : gray level, and θ : direction.

- Other Statistical Features: Other Statistical Features are further subdivided into four different features:

I. Edge Feature (Absolute Gradient): Information present in the edges is higher as compare to the different parts of the image. Spatial variation in an image is calculated by the gradient value. If there is an instant change in pixels then gradient will be high else it will be low. It calculates two features: absolute gradient and absolute gradient variance.

II. NGTDM: In NGTDM, the calculation of busyness, coarseness, complexity, contrast, strength takes place. This method considers a difference between the gray levels between pixels.

III. SFM: It calculates contrast, coarseness, periodicity and roughness of pixels at different distance within an image.

IV. GLDS: The contrast, energy, entropy, homogeneity and mean are calculated by GLDS on the basis of the co-occurrence of the pixel pair that has variation in gray levels divided by particular distance.

Signal Processing Based Method: Five statistical parameters i.e. mean, standard deviation, skewness, kurtosis and entropy using Laws' masks of lengths 3, 5, 7 and 9 are computed. In this method image can be filtered with some specific masses to accesss texture properties, these masks are created by combination of different one dimension kernel vectors. There are five types of masks can be used : EDGE, Level< wave, Ripple, Spot.

Transform domain method: Transform domain method is sub divided into three different categories:

- Gabor Wavelet Transform (GWT): It is a 2D wavelet serving as a basis of

Fourier transform which can be used to represent data where uncertainty represented is minimized. It is best used for feature extraction as it reduces the 2-D enunciation in space and frequency.

- Wavelet Packet Transform (WPT): This allows achieving time frequency localization and multi scale resolution by focusing in the neighborhood. It represents both transient and stationary behavior of the signal using very few transform coefficients. WPT is best used for denoising, compression and classification.
- Fourier Power Spectrum (FPS): For FPS, two spectral features namely radial sum and angular sum were computed.

Data Partitioning: The collected data (images) is divided according to their shapes and sizes as Benign or Malignant, further testing and training is done on each separately.

Classification Module:

Classification module is a data extraction technique which is used to predict the future data set for data instances. It differentiates the testing samples into various classes. It can be categories into two different classes such as supervised and unsupervised classification. In case of supervised classification the classes has been defined already, and not defined for unsupervised classification. Different classifiers are :

- k-NN
- PNN
- ANN
- SVM
- SSVM

k-NN (k- Nearest Neighbor) classifier: it is used in statistical estimation and pattern recognition. It is considered as top 10 data extraction methods. K-NN classifies data sets based on similarities with the neighbor thus also considered as lazy learner algorithm. Its processing differs with respect to k values. Analysis of Stored data helps in generating the results. Intermediate values are neglected in k-NN. It is applicable to the data from any

distribution. It is a very good classifier if the data is too large. Disadvantages of k-NN are that there is no training stage all the work is done during the testing stage and it requires large number of samples as well.

PNN (Probabilistic Neural Network) classifier: It's far predominantly a classifier. It is supervised Bayesian based feed-ahead neural network used for estimating the magnificence of unknown instances. The PNN classifier contains of four layers: input layer, pattern layer, summation layer and output layer. The values from the checking out dataset are exceeded to the 'n' neurons inside the input unit. The values from the input layer are similarly forwarded to the sample units within the sample layer in which responses for each unit are calculated on basis of possibility density function.
The pattern unit values are forwarded to the layer of summation where the responses are summed and get the response of each category. To get the classes of the unknown instances the maximum response from all categories is taken to the decision layer. For PNN, the choice is crucial to choose spread parameter or the kernel with parameter. By performing repeated experiment for values of SP, the optimal values are used for SP to design a PNN classifier.

ANN (Artificial Neural Network) classifier: there are two components of ANN: nodes and weights. The techniques which are artificially intelligent such as Neural network, genetic algorithm and fuzzy logic are very powerful equipments. The tool is used to detection and analyzing the relationship for massive unknown data. Neural networks are programmed rather than taught to perform.
SVM (support vector mechanism) classifier : it is a method for classification and regression, it even leads to high performance in practical applications. The SVM is one of the most accurate and efficient group of algorithms in machine language.SVM used a hypothesis space of function (linear) in higher dimensional feature space which is known as kernel function. It is well suited for high dimension data, it is used for linear and non linear applications as well.

SSVM(Smooth Support Vector Machine) classifier: SVM has been solved traditionally in dual space. SSVM solved the optimization problem in dual space. SSVM was introduced in order to adopt SVM.

CHAPTER 4

CLASSIFICATION OF BREAST LESIONS USING GABOR WAVELET FOR THREE CLASS

A CAD system is designed for detection of lesion which may indicate the presence of breast cancer. The extraction of the ROI in this paper was done using the software imageJ with different image processing techniques such as preprocessing, feature extraction, feature classification using MATLAB. GaborWavelet gives higher accuracy and result as compared to the previous researches done. The methods followed were:

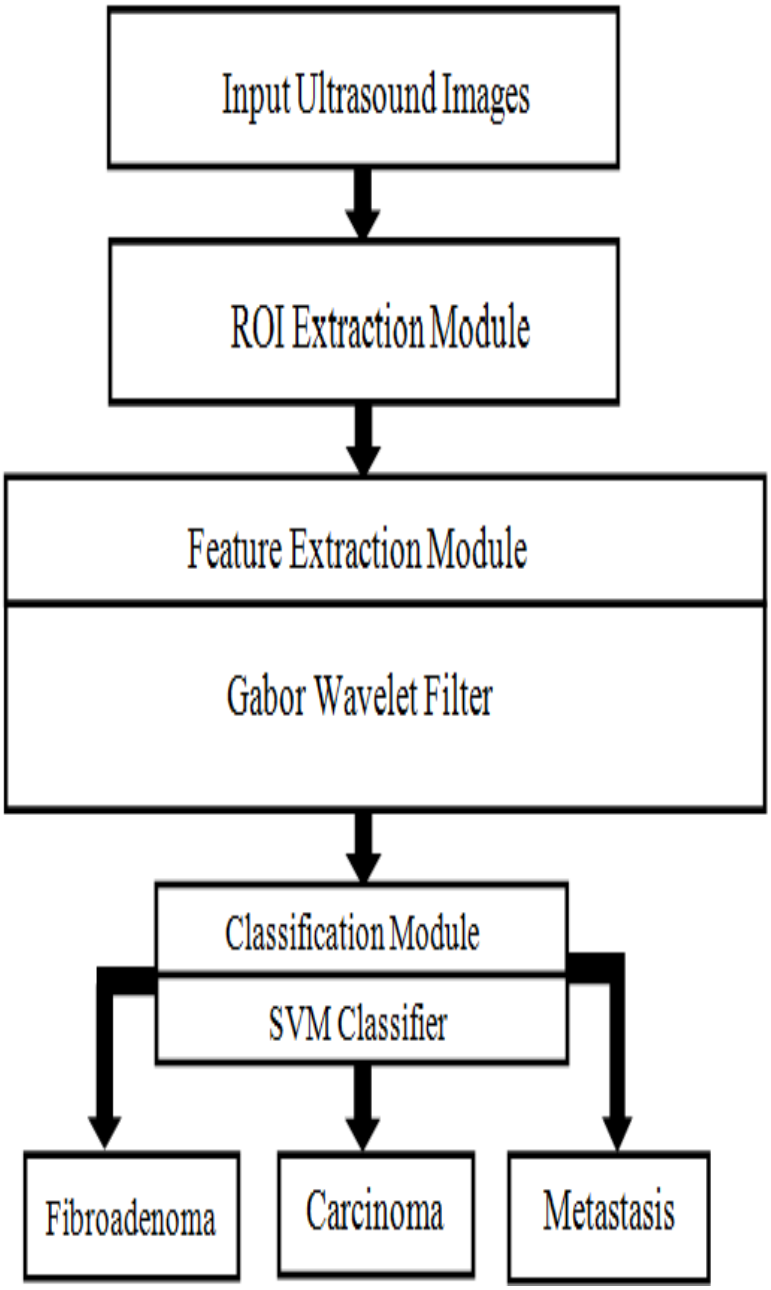


Fig 4.1: Overview of the system

ROI Extraction:

Data collection of the ultrasound images are taken for analysis of Fibroadenoma, Carcinoma, Metastasis. Except for the cases of biopsies and cases having blood vessels, total data was taken for 100 cases. 25 cases of Fibroadenoma, 27 of Carcinoma and 49 cases of Metastasis consisted in the data base.

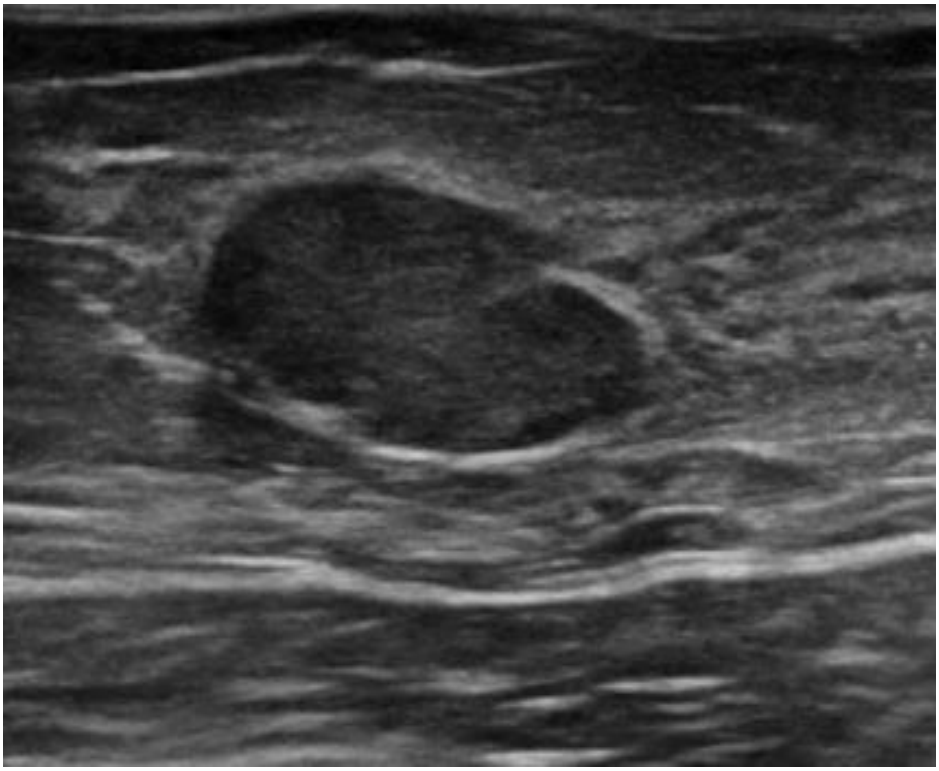


Fig 4.2: Image of Fibroadenoma[34]

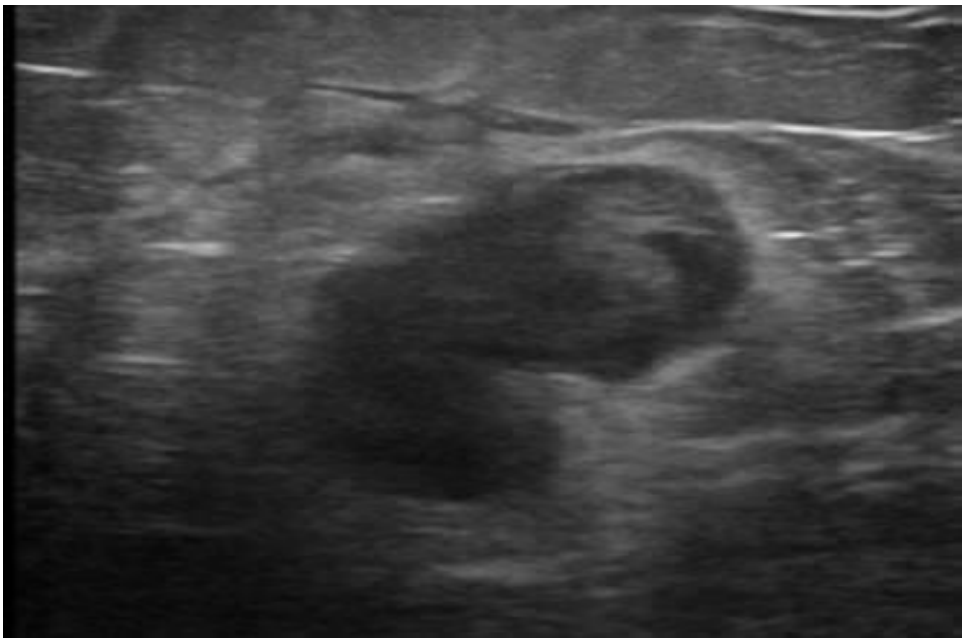


Fig 4.3: Image of Carcinoma[34]

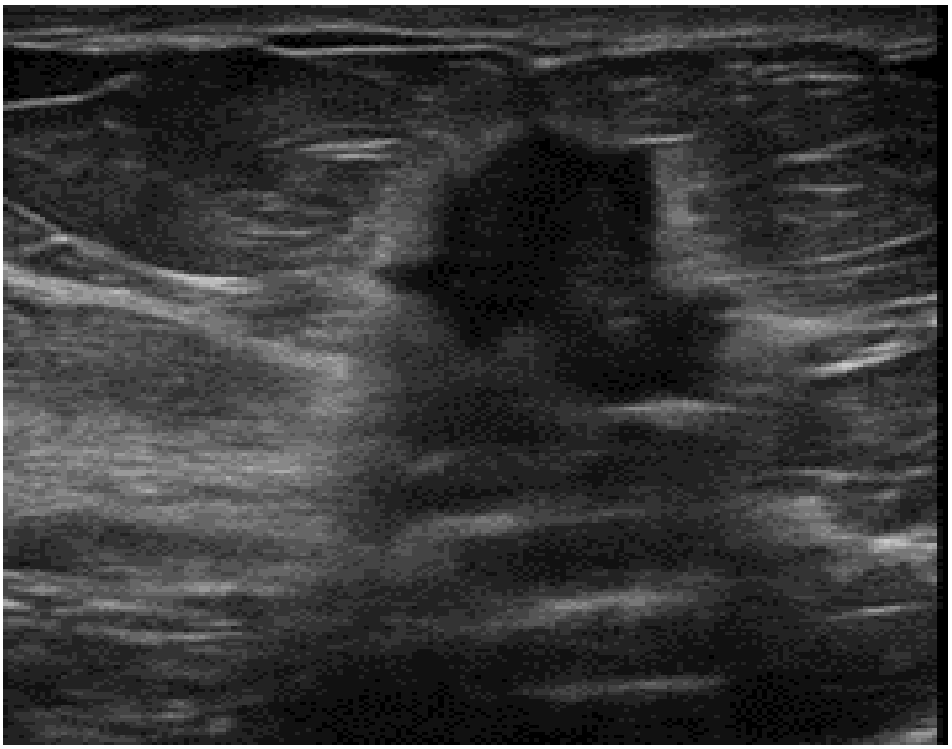
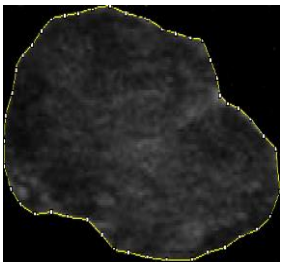
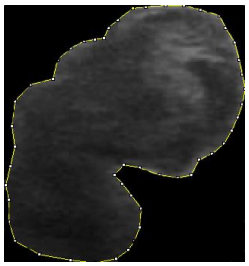


Fig 4.4: Image of Metastasis[34]

The ROI of the deformities and irregularities of the images are segmented using imageJ software. ImageJ: it is java based image processing public domain which is based on image processing. This software presentations, edit, examine, system, shop in addition to print eight-bit color and grayscales, 16 bit integer and 32 bit floating point photos. It may additionally examine numerous record formats (TIFF,PNG,GIF, JPEG, BMP, DICOM and FITS) in conjunction with the raw data. It also supports photograph stacks, a series of images that proportion a single window and it's far multithreaded,so the multi CPU performed parallel operation which is time consuming. This software program can also calculate vicinity and pixel value information of user defined choice and depth threshold items. ImageJ can gives measurements of distance and angles. By using this, the density histograms and line profile plots. The logical and arithmetical operation between photos, contrast manipulation, convolution, fourier analysis, polishing, smoothing, side detection and median filtering is also supported by ImageJ software program. It can support any number of images simultaneously. This software helps marking and segmenting the infected area. The segmented region is confined into a rectangular box adjoining the boundaries of abnormality as shown in Figures given below:





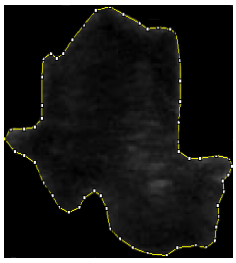


Fig.4.2.1:ROI of Fibroadenoma **Fig.4.3.1**:ROI of Carcinoma **Fig.4.4.1**: ROI of Metastasis [50]

Feature extraction:

For image processing and pattern recognition, feature extraction plays an important step that is special form of dimensionality reduction. As discussed earlier in this chapter, Gabor Wavelet is used as Feature Extraction module under Transform Domain Feature for better results. Gabor Wavelet provides optimal resolution in time as well as frequency domain. For Gabor Wavelet we have 3×7 dimension, where Θ is 7, and 3 are the scales (0,1,2) which is equal to 21 mask. Two Dimensional Gabor Wavelet Transform:

The application of 2D-GWT results in a set of frequency and orientation selective filters which capture energy at specific frequency and orientation. The 2-D-GTW, considering three scale(0,1 and 2) and seven angels(22.5^0, 45^0, 67.5^0, 90^0, 112.5^0,135^0,157.5^0) resulting in a group of 21 wavelets(7×3).Mean and standerd deviation are computed as features forming a texture feature vector (TFV) of length 42.

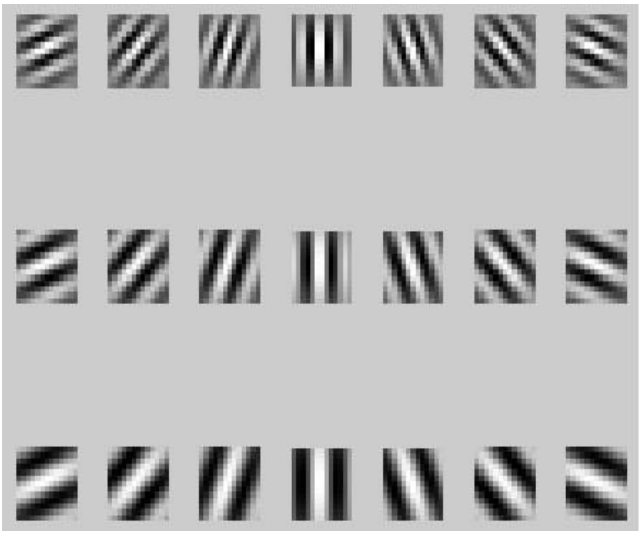


Fig 4.5: Real part of Gabor filter family of 21 wavelets

The algorithm for Gabor Wavelet feature is :

Step1: Collection of data

Step2: Data preprocessing was done by selection of ROI's.

Step3: Import ROI images.

Step4: Conversion of ROI images to Gray level images.

Step5: Apply Gabor Wavelets.

Step6: Convolve Gabor Wavelets with gray level images.

Step7: Calculate the features.

Formula used for calculating features:

$$Mean = \frac{1}{m}\sum_{i,j} i\, P_{i,j} \quad (1)$$

$$Variance = \sum_{i,j} P_{i,j}(i - \mu_i)^2 \quad (2)$$

Classifier

SVM is used as a classifier for the extraction of the result in this CAD system. SVM creates a feature space, which is a finite dimensional vector space, whose each dimension represents feature of a certain object. Objective of SVM is the training of a model, assigning new unseen objects into a specified category. This is done by creating a linear partition of feature space into two categories. On the basis of feature in the new unseen objects, it places an object above or below the separation plane, which gives a categorization.

Results

98% result is computed in this paper using Gabor Wavelet as Feature Extraction Module and SVM as a Classifier.

The computed result is shown in table number 4.1:

Result of SVM classifier:

TFV(*l*)	**CM**				**Sensitivity**	**OCA**
		FA	**CR**	**MS**		
TFV(50)	**FA**	17	0	0	100%%	98%
	CR	0	16	1	94.1%	
	MS	0	0	16	100%	

Note:TFV: Texture Feature Vector , CM: Confusion Matrix OCA: Overall classification Accuracy , sensitivity, FA: Fibroadinoma, CR: Carcinoma, MS: Metastasis, *l*:length of TFV

Table 4.2: Comparison of results with existing results

AUTHOR	**FEATURE**	**CLASSIFIER**	**ACCURACY**
Bhusri S *et al*	Texture feature	SVM	85%
Huang YL *et al*	Morphological features	SVM	90%
Proposed work	Texture feature FPS	SVM	98%

Conclusion

Various CAD system designs have proved to be useful for the radiologists as it provides a second opinion tools for breast lesions classification of ultrasounds where a clear discrimination cannot be made easily. Different CAD system designs employing the morphological features.

CHAPTER 5

TWO CLASS CLASSIFICATION OF BREAST LESIONS USING STATISTICAL AND TRANSFORM DOMAIN FEATURE

Computer- Aided Detection (CAD) system is designed for detecting lumps which may indicate presence of breast cancer. This paper presents the classification of breast ultrasound images using Statistical and Transform domain feature extraction techniques were data is partitioned by hold-out method and classified using Support Vector Machine (SVM) classifier. SVM trains a model that assigns unseen new objects into a specific category. The best obtained result out of all the features used is calculated using Fourier Power Spectrum (FPS).

Methods used: The experimental flow of the system follows a sequence as shown in Figure 1. Initially data collection is done from database. The ROI's are segmented using MATLAB 14.7 software (as explained in the 3rd chapter). The ROI's are further used for the extraction of features. Different features are calculated by different feature extraction techniques. Partitioning of data is done by hold-out technique. Lastly classifiers are used on the test data so as to get accuracy and sensitivity.

Steps followed in this paper:

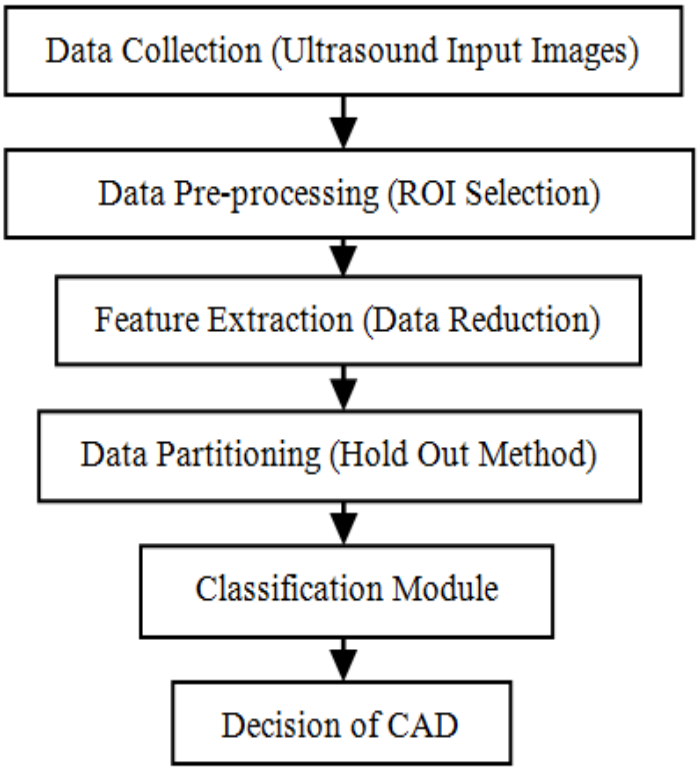


Figure 5.1: Flowchart of the proposed work

Database Ultrasound Images

The ultrasound images are taken from []for analysis of benign and malignant cases. The data is taken for 100 cases discarding the cases of biopsy and cases having blood vessels. The data base contains 25 cases of benign and 75 cases of malignant.

Extraction of ROI:

The unhealthy or abnormal area in the ultrasonic images are marked with the help of experienced radiologist and segmented with the help of MATLAB 14.7 software .The infected area is segmented out in form of small squares of 32 × 32 pixels. The ultrasound cases and segmented squares are displayed in the Fig 5.1 and Fig 5.2.

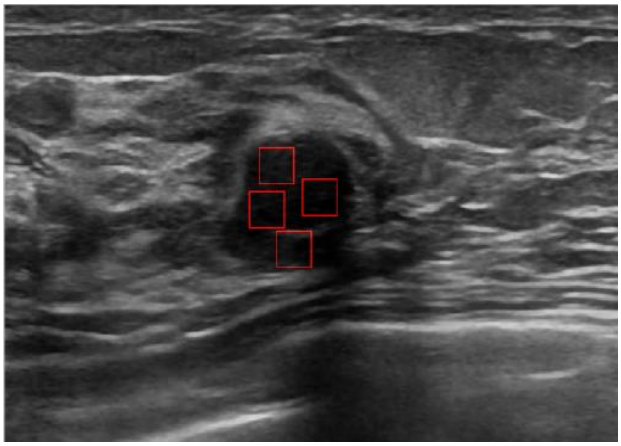


Fig 5.2 : Ultrasound of Fibroadenoma(benign)[34]

Fig 5.2.1: Cropped ROI of Fibroadenom

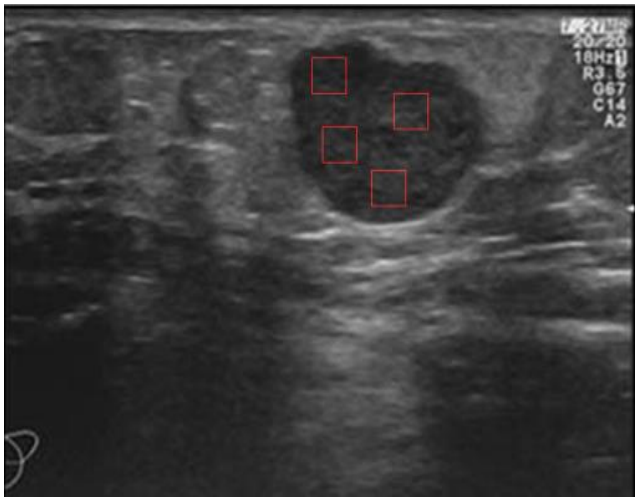


Fig 5.3: Ultrasound of Carcinoma (Malignant)[34]

Fig 5.3.1 Cropped ROI of Carcinoma

Feature extraction Module: A process which is used for detecting and representing certain feature of interest within an image for further processing is called feature extraction. The data is converted to a reduced set of feature representations when the input data is very large for processing and imagined to be unnecessary. Features often contain information relative to color, shape, texture or context. The Texture method is divided into three different methods: Signal Processing, Transform Domain and Statistical.

Statistical features (SF) were divided into four main categories : First order Statistics (FOS), Second order Statistics which calculates Gray Level Co-occurrence Matrix (GLCM), Higher order Statistics which calculates The Gray Level Run Length (GLRL) and Other Statistics which calculates Edge Features, Neighborhood Gray Tone Difference Matrix (NGTDM), Statistical Feature Matrix (SFM), and Gray Level Difference Statistics (GLDS).

i. **FOS :** The FOS features are extracted from the gray level intensity histograms of an image. We have five different FOS features : average gray level, standard deviation, kurtosis, smoothness and entropy.

ii. **GLCM Features:** Generally how many times combinations of different grey level pixels occur in an image used for second order texture calculations is defined by GLCM. Relation between groups of 2 pixels is considered. These pixels are also known as reference and neighbor pixel. The neighbor pixel is the one right to the reference pixel. Inside a window, every pixel is referred as the reference pixel which is starting at the top left corner and ending towards the lower right hand corner. The GLCM includes features of calculating contrast, entropy, sum entropy, difference entropy, correlation, inverse difference moment, sum average, variance, sum variance, difference variance angular moment etc

GLCM runs on the bases of following algorithm :

Step 1: Different extracted ROI images (database) were imported.

Step 2: Conversion of database images to gray level images.

Step 3: Assume initial value of co-occurrence matrix as zero.

Step 4: Assign index to co-occurrence matrix.

Step 5: Calculation of co-occurrence matrix.

Step 6: Calculation of pixels (reference pixel and neighbor pixel) of the row matrix.

Step 7: Normalization (Z Normalization) of data. Normalization in GLCM is to transform the matrix into a close approximation of probability table. The normalization is only an approximation because the gray levels are integer value so they are discrete in nature but probability work on the continuous values.

$$\text{Z normalization} = \frac{\text{basic value} - \text{mean}}{\text{standard deviation}} \qquad (1)$$

Step 8: Calculation of GLCM features. There are different features of GLCM: entropy, sum entropy, correlation, inverse difference moment, variance, contrast etc. Formulae used for calculation of features are:

1. Entropy is calculated by:

$$\text{Entropy} = -\sum_{i,j} P_{i,j} \log(P_{i,j}) \qquad (2)$$

2. For calculation of Sum Entropy:

$$\text{Sum Entropy} = -\sum_{i=2}^{2Ng} P_{i,j}(i) \log(P_{x+y}(i)) \quad (3)$$

3. For calculating correlation:

$$\text{Correlation} = \sum_{i,j} P_{i,j} \left[\frac{(i-\mu_i)(j-\mu_j)}{\sigma_i \sigma_j}\right] \quad (4)$$

4. For calculating Inverse Difference Moment:

$$\text{Inverse Difference Moment} = \Sigma_{i,j} \frac{P_{i,j}}{1+(i-j)^2} \quad (5)$$

5. For calculation of variance:

$$\text{Variance} = \Sigma_{i,j} P_{i,j}(i-\mu_i)^2 \quad (6)$$

6. For calculation of contrast:

$$\text{Contrast} = \sum_{i,j} P_{i,j}(i-j)^2 \quad (7)$$

where *i* and *j* are horizontal and vertical coordinates respectively. μ is expected value and *P* is Probability Mass Function (the function that gives the probability that a discrete random variable is exactly equal to some value).

iii. **GLRL Features** : Way's in which we extract the higher order statistic texture feature is called the GLRL method. Gray level run consists of successive sets of pixels having the same gray level collinear in a accustomed direction. Length run is the number of pixels in run and the count of such a run occurring in an image is called the length run value. GLRL consists of each element is represented by a two-dimension matrix $p(i, j \mid \theta)$, where j : length run, i : gray level, and θ : direction.

iv. **Edge Features**: Information present in the edges is higher as compare to the different parts of the image. Spatial variation in an image is calculated by the gradient value. If there is an instant change in pixels then gradient will be high else it will be low. It calculates two features: absolute gradient and absolute gradient variance. The algorithm followed for Edge feature is shown below:

Step 1: Collection of dataset.

Step 2: Data preprocessing was done by selection of ROI's

Step 3: Import ROI images.

Step 4: Conversion of ROI to gray level images.

Step 5: Calculation of horizontal and vertical edges using sobel filter. (Sobel Filter is used to calculate the approximation of the gradient of the image intensity function. It creates an image emphasizing edges in image processing and computer vision).

Step 6: Compute probability distribution of gray level images.

Step 7: Compute mask to selected ROI's.

Step 8: Discard elements which counts pixels outside ROI.

Step 9: Normalization of matrix.

Step 10: Calculation of the features (absolute gradient). For calculation of absolute gradient formula is given in eq 8.

$$\nabla_j A^k \equiv \frac{\partial A^k}{\partial x^j} + A^i \Gamma_{ij}^k \qquad (8)$$

where Γ_{ij}^k is Christoffel symbol, *A* is basis vector and *k* is free index.

v. **NGTDM**: In NGTDM, the calculation of busyness, coarseness, complexity, contrast, strength takes place. This method considers a difference between the gray levels between pixels. The steps of algorithm are :

Step 1: Collection of dataset.

Step 2: Data preprocessing was done by selection of ROI's

Step 3: Import ROI images.

Step 4: Conversion of ROI to gray level images.

Step 5: Calculation of NGTDM vector by comparing pixels to its neighborhood gray tone values of pixels.

Step 6: Calculation of normalization coefficient.

Step 7: Define neighborhood kernels (Kernel is the small matrix which is useful for blurring, sharpening, edge detecting etc).

Step 8: Convolve the kernels of ones to select pixels for which the kernel lies entirely within ROI.

Step 9: Compute NGTDM matrix by convolving these kernels with gray level images.

Step 8: Calculation of the features (busyness, coarseness, complexity, contrast (same as in eq.7)).

Formula for calculating image complexity is:

$$\text{Complexity} = \frac{\text{RMSE}}{\text{CR}} \qquad (9)$$

where RMSE = root mean square error between lossy compressed image and original image.

CR = Compression ratio.

vi. SFM: It calculates contrast, coarseness, periodicity and roughness of pixels at different distance within an image. Algorithm for SFM:

Step 1: Collection of dataset.

Step 2: Data preprocessing was done by selection of ROI's

Step 3: Import ROI images.

Step 4: Conversion of ROI to gray level images.

Step 5: Define coordinates of statistical matrix.

Step 4: Shift images by deleting rows and columns.

Step 6: Define ROI in original and shifted images.

Step 7: Common pixels in both areas are considered.

Step 8: Compute statistical feature matrix.

Step 9: Calculation of the features (contrast (same as in eq.7), periodicity and absolute roughness).

Formulae for feature calculation:

$$\text{Roughness} = \frac{1}{\text{n}}\sum_{\text{i}=1}^{\text{n}}|\text{y}_\text{i}| \qquad (10)$$

where y_i = the vertically distance from i^{th} data point to the mean line.

n = the order of equally spaced point in the pixels.

vii. **GLDS:** The contrast, energy, entropy, homogeneity and mean are calculated by GLDS on the basis of the co-occurrence of the pixel pair that has variation in gray levels divided by particular distance.

Step 1: Collection of dataset.

Step 2: Data preprocessing was done by selection of ROI's

Step 3: Import ROI images.

Step 4: Conversion of ROI to gray level images.

Step 5: Calculate the probability distribution of gray level differences for different dimensions and angles.

a. Compute mask to select ROI.
b. Setup gray level difference vector.
c. Initialize gray level difference count vector.
d. Calculate gray level difference.

Step 6: Calculate gray level difference at 0, 45, 90 and 135 degree.

Step 7: Normalize gray level difference matrix for all angles.

Step 8: Calculation of the features (contrast (same as in eq.7), homogeneity, energy, entropy (same as in eq. 2), and mean).

1) The homogeneity is calculated by:

$$\text{Homogenity} = \sum_{i,j} \frac{P_{i,j}}{1+(i-j)^2} \tag{11}$$

2) Energy is computed by following formula :

$$\text{Energy} = \sqrt{P_{i,j}^2} \tag{12}$$

3) Mean is calculated using:

$$\text{Mean} = \frac{1}{m} \sum_{i,j} iP_{i,j} \tag{13}$$

where *i, j* and *P* are same as mentioned for eq. 2 to eq.7.

Transform domain methods (TD) were divided into three categories : Gabor Wavelet transform (GWT), wavelet packet transform (WPT) , and FPS.

i. **GWT**: It is a 2D wavelet serving as a basis of Fourier transform which can be used to represent data where uncertainty represented is minimized.

ii. **WPT**: This allows achieving time frequency localization and multi scale resolution by focusing in the neighborhood.

iii. **FPS** : For FPS, two spectral features namely radial sum and angular sum were computed. Formulas of radial sum and angular sum are given in Eq 14 and Eq 15 respectively.

FPS feature algorithm has been discussed below:

Step 1: Collection of dataset.

Step 2: Data preprocessing was done by selection of ROI's.

Step 3: Import ROI images.

Step 4: Conversion of ROI to gray level images.

Step 5: Apply mask to gray level images.

Step 6: Compute FFT of rows and column.

Step 7: Calculation of FPS features.

$$Radial\ Sum = \sum_{r_1^2 < u^2 + v^2 < r_2^2} |F(u,v)|^2 \quad (14)$$

$$Angular\ Sum = \sum_{\theta_1 \leq \tan^{-1}\frac{v}{u} \leq \theta_2} |F(u,v)|^2 \quad (15)$$

where *F* is the function; *u* and *v* are the directions of coordinates; *r* are the radius.

Classification Modules

There are two type of characterization of classification; Supervised Classification and Unsupervised Classification. When the classes are defined for the training sets in classification, then it is supervised classification and opposite is for unsupervised classification. The classifier used in our work is SVM classifier. The SVM is supervised type of classifier. For implementation of SVM classifier a library is used in MATLAB, called LibSVM .SVM classifier proceeds on the principle of decision planes, where the boundaries of decision are defined. It works on both linear and non-linear classification

with higher speed and accuracy. Kernel Based Classifier; the training data of non-linear aligning to higher dimensional feature space from input space has been done using kernel functions. There are different types of kernels. For classification task in this paper we have used the *Gaussian radial basis function*. The appraisal of kernel parameter γ and parameter of regularization C is always a diagnostic step for having desired abstract performance. By doing the expanded search that is carried out in the parameter space for the values of C ϵ {2-4, 2-3… 215}, γ ϵ {2-12, 2-11… 24}, the admirable values of γ and C are obtained

RESULT AND DISCUSSIONS

The results for two class classification of breast lesions were computed by SF and TD. The results of classification using TD method using FPS feature is shown in table 1. The sensitivity calculated for malignant is 92.8% where as for benign is 81.2%. The overall classification accuracy calculated is 91.5% using SVM classifier.

Table 5.1: Sensitivity and Accuracy Calculation for TD

Feature	CM			Sensitivity	OCA
		M	B		
FPS	M	117	9	92.8%	91.5 %
	B	3	13	81.2%	

***Note*:** CM: Confusion Matrix OCA: Overall classification Accuracy,
M: Malignant, B: Benign, FPS: Fourier Power Spectrum.

Result obtained for SF using SVM classifier for two class classification is presented in Table 5. 2

Table 5.2: Sensitivity and Accuracy Calculation using SF						
Feature	**CM**			**SM**	**SB**	**OCA**
		M	**B**			
EDGE	**M**	121	5	96.03%	0%	85.2%
	B	16	0			
GLDS	**M**	121	5	96.03%	43.2%	90.4 %
	B	9	7			
GLCM	**M**	0	126	0%	100%	11.2%
	B	0	16			

Note: CM: Confusion Matrix OCA: Overall classification Accuracy, M: Malignant, B: Benign, SM: Sensitivity for Malignant, SB: Sensitivity for Benign

The formulas for parameters used in calculation of sensitivity and overall classification accuracy are:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{16}$$

Sensitivity is also known as individual classification accuracy or true positive rate.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{Total Number of Cases}} \tag{17}$$

accuracy is known as Overall Classification Accuracy (OCA).

In table 5.2, the results are obtained with EDGE feature for SM is 96.03%, SB is 0% and OCA calculated is 85.2%. For GLDS, SM calculated is 96.03%, SB is 43.2% and OCA calculated is 90.4%. For GLCM feature the SM is 0%, SB is 100%, OCA is 11.2%, for NGDTM, SFM and GLRL results were misclassified so not considered. For betterment of our work, the comparison between obtained results with former results is presented in Table 5.3.

Table 5.3: Comparison of results with existing results			
AUTHOR	**FEATURE**	**CLASSIFIER**	**ACCURACY**
Bhusri S *et al*	Texture feature	SVM	85%
Huang YL *et al*	Morphological features	SVM	90%
Proposed work	Texture feature FPS	SVM	91.5%

On bases of the literature survey that was done, a comparison table is made for different authors using different features and classifiers which lead to appreciable results. S. Bhusri et al uses texture features which includes SFM, NGTDM, FOS, GLCM, GLRLM and GLDS. They have also used SVM as a classifier. 90% accuracy was obtained by Huang YL using SVM classifier. In this paper PCA with Morphological features were used where 19 different feature were extracted from the contour. In our present paper ROI has been extracted from the collected data set of the images, texture features were extracted using statistical methods and transform domain methods. The results calculated for FPS feature gives the highest accuracy of 91.5% using SVM classifier

CONCLUSIONS AND FUTURE RESEARCH

The present work has been worked out for two-class breast lesions classification of ultrasound cases for enhancement of the images and for higher accuracy. This paper focuses on achieving higher accuracy as compared to the results of different authors. Breast cancer has been classified into two categories and FPS has attained the best results. Benign for Edge and Malignant for GLCM are misclassified so the author can focus on enhancing the results for Edge and GLCM features. Different Texture Feature Extraction methods can be used to attain better results by using these features as well.

CONCLUSION AND FUTURE WORK

For designing an efficient CAD system for two-class breast tissue density classification, various CAD system designs based on Morphological features and texture based features have been proposed in the present work. For Gabor Wavelet for three class classification, 98% result is computed using Gabor Wavelet as Feature Extraction Module and SVM as a Classifier.

Two class classification of breast lesion using statistical and transform domain features the results were obtained by using EDGE feature for SM computing a result of 96.03%, SB computes a result of 0% and OCA calculated is 85.2%. For GLDS, SM calculated is 96.03%, SB is 43.2% and OCA calculated is 90.4%. For GLCM feature the SM is 0%, SB is 100%, OCA is 11.2%, for NGDTM, SFM and GLRL results were misclassified so not considered.

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:

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.

As the work is done on Gabor Wavelet features for three class classification, work could be done using various other features as well and by using various other crop techniques, calculations can be done for three class classification also.

REFRENCES

[1] http://www.nationalbreastcancer.org/causes-of-breast-cancer
[2] http://www.cancer.org/acs/groups/cid/documents/webcontent/003180-pdf.pdf
[3] Understanding your mammogram report-BI-RADS categories, American Cancer Society, [online], http://www.cancer.org/healthy/findcancerearly/examandtestdescriptions/mammogramsandotherbreastimagingprocedures/mammograms-and-other-breast-imaging-procedures-mammo-report
[4] S. Bhusri, S. Jain, and J. Virmani, classification of breast lesion using the difference of statistical features, July – August 2016 RJPBCS 7(4) Page No. 1365
[5] S.Y. Chen1, H.-H. Chang1, S.-H. Hung3,W. C. Chu1,Breast Tumor Identification in Ultrasound Images Using the Normalized Cuts with Partial Grouping Constraints, International Conference on BioMedical Engineering and Informatics, 2008, pp. 28-32.
[6] X. Liu, Z. Huo, J.Zhang, Automated segmentation of breast lesions in ultrasound images, EMBS 2005, 2006, pp.7433–7435.
[7] S. Bhusri, S.Jain and J.Virmani, Breast Lesions Classification Using The Amalgagation Of Morphological and Texture Features. International J of Pharma and Bio Sciences(IJIPBS). Apr-jun 2016.pp. 617 – 624.
[8] S. Bhusri, S.Jain , and J Virmani, Classification of Breast Lesions Using the Difference of Statistical Features, Research Journal of Pharmaceutical, Biological and Chemical Sciences(RJPBCS), july-aug 2016,pp. 1366.
[9] S.K. Alam, E.J. Feleppa, M.Rondeau, A. Kalisz, and B.S. Garra,Ultrasonic multi-feature analysis procedure for computer-aided diagnosis of solid breast lesions,2011. vol. 33, no. 1, pp. 17–38.
[10] Y.L. Huang, D.R. Chen, Y.K. Liu, Breast cancer diagnosis using image retrieval for different ultrasonic systems, IEEE,2004,pp.2957– 2960.
[11] K. Drukker, M.L. Giger, and E.B. Mendelson,Computerized analysis of shadowing on breast ultrasound for improved lesion detection, 2003, pp. 1833–1842.
[12] Y.L. HUANG, D.R. CHEN, Y.R. JIANG, S.J. KUO, H.K. WU and W. K. MOON, Computer-aided diagnosis using morphological features for classifying breast lesions on ultrasound, Wiley InterScience, 2008; .DOI:10.1002/uog.5205; Ultrasound Obstet Gynecov; 32: 565–572
[13] T. Tan, B. Platel, H. Huisman, C.I. Sánchez, R. Mus, and N. Karssemeijer,Computer-Aided Lesion Diagnosis in Automated 3-D Breast Ultrasound Using Coronal Spiculation, IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 31, NO. 5, MAY 2012, pp. 1034 – 1042.
[14] S. Sharma, S.Jain, S. Bhusri, classification of breast lesions using Gabor WAVELET Filter for Three classes, Internaltion Confrence on computer for sustainable global development, BVICAM, 2017. Pp.628-6284.
[15] Amandeep, S.Jain, S.Bhusri, CAD system for non small cell lung carcinoma ysing laws' mask analysis, internation conference on computing for sustainable global development, BVICAM,2017,pp. 6285-6288.
[16] Wu WJ, M. WK, Ultrasound breast tumor image computer-aided diagnosis with texture and morphological features,(2008) ,15(7):873–880.
[17] MAR. Nayeem, MA.M.Joadder, S.A. Shetu, Feature Selection for Breast Cancer Detection from Ultrasound Images,IEEE,2014. 978-1-4799-5180-2/14

[18] 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.
[19] A.V.Alvarenga, AFC Infantosi, WC Pereira, Assessing the combined performance of texture and morphological parameters in distinguishing breast tumors in ultrasound images, Azevedo CM (2012) . Med Phys 39(12):7350–7358.
[20] S. Bhusri, S.Jain and J.virmani, Classification of breast lesions based on laws' feature extraction techniques, 16-18 March 2016, Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference, BVICAM,New Delhi.
[21] S.Rana, S.Jain, and J.Virmani, SVM-Based characterization of focal kidney lesions from B-Mode ultrasound images, research J of pharmaceutical, biological and chemical sciences (RJPBCS). July-Aug 2016,vol. 7(4), pp.837.
[22] S.Rana, S.jain, and J.virmani, classification of focal kidney lesion using wavelet-based texture description, internation J of pharma and bio sciences, July-Sep 2016, pp.646-652.
[23] S.Jain, Design and Simulation of Fuzzy system using two stage CMOS Operational amplifier,J of Active and passive Electronic Devices. 2014, pp.329-338.
[24] A.Dhiman, A. Singh, S.Dubey and S. Jain, Design of lead II ECG Waveform and Classification performance for Morphological features using Differenct Classifiers on lead II, research J of pharmaceutical, biological and chemical sciences, July-aug 2016, pp. 1226-1231.
[25] S. Sharma, S.Jain, S.Bhusri, classification of beast lesion using gabor wavelet filter for three classes, 11th INDIACom-2017,march 1st-3rd, 2017, pp 6282-6284, BVICAM, New Delhi.
[26] S.Jain,DS.Chauhan, implementation of fuzzy system using different voltages of OTA for JNK pathway leading to cell survival/death,Network biology, 2015,pp62-70.
[27] S.Jain and PK Naik, System Modeling of cell survival and cell death: A deterministic model using Fuzzy System, International J of Pharma and BioSciences. 2012. pp. 358-373.
[28] S.Jain and DS. Chauhan, Linear and Non Linear Modeling of Protein Kinase B/ AkT, International Conference on Information and Communication Technology for Sustainable Development (ICT4SD - 2015), Ahmedabad, India, pp 81-88.
[29] S.Jain, PK Naik and SV. Bhooshan, Nonlinear Modeling of cell survival/ death using artificial neural network, International Conference on Computational Intelligence and Communication Networks (CICN2011), Gwalior, India, Oct 2011, pp 565-568.
[30] S.Rana, S.Jain and J.Virmani, Classification of Kidney Lesions using Gabor Wavelet Texture Features, Computing for Sustainable Global Development (INDIACom), 2016, pp. 2528-2532.
[31] O. Faust, U. R. Acharya, and T. Tamura, Formal Design Methods for Reliable Computer-Aided Diagnosis: A Review, IEEE 2012, pp. 15-28.
[32] C.M. Lo, R.T. Chen, Y.C. Chang, Y.W. Yang, M.J. Hung, C.S. Huang, and R.F. Chang, Multi-dimensional Tumor Detection in Automated Whole Breast Ultrasound using Topographic Watershed, IEEE, 2014,pp. 1-9.
[33] C.Ye, V. Vaidya, and F. Zhao,Improved Mass Detection in 3D Automated Breast Ultrasound Using Region Based Features and Multi-view Information, IEEE, 2014, pp. 2865-2868.
[34] http://www.ultrasoundcases.info

[35] Minavathi, S.Murali, M.S.Dinesh, Classification of Mass in breast Ultrasound Images Using Image Processing Techniques, International Journal of Computer Applications , vol. 42, 2012 ,pp. 29-36.
[36] S.Jain and DS Chauhan, Mathematical Analysis of Receptors For Survival Proteins, International J of Pharma and Bio Sciences (IJPBS), 2015, Vol-6. pp. 164-176.
[37] S.K. Alam, E.J. Feleppa, M. Rondeau, A. Kalisz and B.S. Garra ,Ultrasonic Multi-Feature Analysis Procedure for Computer-Aided Diagnosis of Solid Breast Lesions, ULTRASONIC IMAGING, 2011, vol. 33, pp. 17-38.
[38] Domingues, A.V. Alvarenga, J.S. Cardoso and W.C.A. Pereira ,SVM Classification of Breast Tumors on Ultrasound Images Using Morphological Features, XXIII CBEB 2012, pp. 1-5.
[39] 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.
[40] R.Z. Azar, C.Leung, T.K. Chen, K. Dickie,J. Dixon, K. Chan, L. Pelissier, An Automated Breast Ultrasound System for Elastography,IEEE,2012.
[41] S.Chen, H.Chang, S.Hung, W.C. Chu, Breast Tumor Identification in Ultrasound Images Using the Normalized Cuts with Partial Grouping Constraints, International Conference on BioMedical Engineering and Informatics,pp.28-32,2008.
[42] T.Hopp, A.Stromboni, N.Duric, N.V.Ruiter, Evaluation of breast tissue characterization by ultrasound computer tomography using a 2D/3D image registration with mammograms, IEEE, pp.647-650,2013.
[43] N.Uniyal, H.Eskandari, P.Abolmaesumi, S.Sojoudi, P.Gordon, L.Warren, R.N. Rohling, S. E. Salcudean, and M.Moradi, A new approach to ultrasonic detection of malignant breast tumors, IEEE, pp.96-99, 2013.
[44] R.Z.Azar, C.Leung, T.K.Chen, K.Dickie, J.Dixon, K. Chan, L. Pelissier, An Automated Breast Ultrasound System for Elastography, IEEE, 2012.
[45] T.Tan, B.Platel, H. Huisman, C.I. Sánchez, R.Mus, and N.Karssemeijer, Computer-Aided Lesion Diagnosis in Automated 3-D Breast Ultrasound Using Coronal Spiculation, IEEE, pp.1034-1042, 2012.
[46] M.Xian, J.Huang, Y.Zhang, X.Tang, MULTIPLE-DOMAIN KNOWLEDGE BASED MRF MODEL FOR TUMOR SEGMENTATION IN BREAST ULTRASOUND IMAGES, IEEE, pp.2021-2024, 2012.
[47] O.Faust, U. R.Acharya, and T.Tamura, FormalDesignMethodsforReliableComputer-Aided Diagnosis: A Review, IEEE, reviews in biomedical engineering, vol.5, 2012.
[48] Y.HUANG, D.CHEN, Y.JIANG, S.KUO, H.WU and W. MOON, Computer-aided diagnosis using morphological features for classifying breast lesions on ultrasound, Ultrasound Obstet Gynecol, vol.32, pp.565–572, 2008.
[49] S.K.Alam, E.J.Feleppa, M.Rondeau, A. Kalisz and B.S.Garra, Ultrasonic Multi-Feature Analysis Procedure for Computer-Aided Diagnosis of Solid breast lesion, ULTRASONIC IMAGING, pp.17-38, 2011.
[50] Image processing and analysis in JAVA,imagej 1.49 version 1.6.024 http://image j.nih.gov/ji/download/win32/*il49-jre6-64.zip*

LIST OF PUBLICATIONS

1. Shreya Sharma, Shruti Jain, Sahil Bhusri , " Classification of Breast Lesions using Gabor Wavelet Filter for Three Classes" March 1st - 3rd , 2017, pp 6282-6284, Proceedings of the 11th INDIACom- 4th 2017International Conference on "Computing for Sustainable Global Development", BVICAM, New Delhi.

2. Shreya Sharma, Shruti Jain, Sahil Bhusri, "Two Class Classification of Breast Lesions using Statistical and Transform Domain features", Journal of Global Pharma Technology (JGPT), 9(6), 2017.

Submission Info

SUBMISSION ID	806842770
SUBMISSION DATE	29-Apr-2017 11:32
SUBMISSION COUNT	1
FILE NAME	thesis_Shreya_152013.d...
FILE SIZE	1.94M
CHARACTER COUNT	58422
WORD COUNT	10393
PAGE COUNT	65
ORIGINALITY	
OVERALL	19%
INTERNET	12%
PUBLICATIONS	13%
STUDENT PAPERS	10%
GRADEMARK	
LAST GRADED	N/A
COMMENTS	0
QUICKMARKS	