

Automated BIRAD Scoring of Breast Cancer Mammograms

Maryam Nasser

Degree of Doctor of Philosophy

Computer science

Kingston University London

2019

Professor Jamshid Dehmeshki

Abstract:

A computer aided diagnosis system (CAD) is developed to fully characterize and classify mass to benign and malignancy and to predict BIRAD (Breast Imaging Reporting and Data system) scores using mammographic image data. The CAD includes a preprocessing step to de-noise mammograms. This is followed by an active contour segmentation to deform an initial curve, annotated by a radiologist, to separate and define the boundary of a mass from background. A feature extraction scheme was then used to fully characterize a mass by extraction of the most relevant features that have a large impact on the outcome of a patient biopsy. For this thirty-five medical and mathematical features based on intensity, shape and texture associated to the mass were extracted. Several feature selection schemes were then applied to select the most dominant features for use in next step, classification. Finally, a hierarchical classification schemes were applied on those subset of features to firstly classify mass to benign (mass with BIRAD score 2) and malignant mass (mass with BIRAD score over 4), and secondly to sub classify mass with BIRAD score over 4 to three classes (BIRAD with score 4a,4b,4c).

Accuracy of segmentation performance were evaluated by calculating the degree of overlapping between the active contour segmentation and the manual segmentation, and the result was 98.5%. Also reproducibility of active contour

using different manual initialization of algorithm by three radiologists were assessed and result was 99.5%.

Classification performance was evaluated using one hundred sixty masses (80 masses with BRAD score 2 and 80 mass with BIRAD score over4). The best result for classification of data to benign and malignance was found using a combination of sequential forward floating feature (SFFS) selection and a boosted tree hybrid classifier with Ada boost ensemble method, decision tree learner type and 100 learners' regression tree classifier, achieving 100% sensitivity and specificity in hold out method, 99.4% in cross validation method and 98.62 % average accuracy in cross validation method.

For further sub classification of eighty malignance data with BIRAD score of over 4 (30 mass with BIRAD score 4a,30 masses with BIRAD score 4b and 20 masses with BIRAD score 4c), the best result achieved using the boosted tree with ensemble method bag, decision tree learner type with 200 learners Classification, achieving 100% sensitivity and specificity in hold out method, 98.8% accuracy and 98.41% average accuracy for ten times run in cross validation method.

Beside those 160 masses (BIRAD score 2 and over 4) 13 masses with BIRAD score 3 were gathered. Which means patient is recommended to be tested in another medical imaging technique and also is recommended to do follow-up in six months. The CAD system was trained with mass with BIRAD score 2 and over 4 also

it was further tested using 13 masses with a BIRAD score of 3 and the CAD results are shown to agree with the radiologist's classification after confirming in six months follow up.

The present results demonstrate high sensitivity and specificity of the proposed CAD system compared to prior research. The present research is therefore intended to make contributions to the field by proposing a novel CAD system, consists of series of well-selected image processing algorithms, to firstly classify mass to benign or malignancy, secondly sub classify BIRAD 4 to three groups and finally to interpret BIRAD 3 to BIRAD 2 without a need of follow up study.

Keywords: Breast Imaging Reporting and Data System (BIRADS) score, feature extraction, feature selection, Principal component analysis, Classification.

Table of Content:

Abstract.....	2
Table of content.....	5
List of tables.....	9
List of figures.....	10
List of diagrams.....	14
Acknowledgment.....	15
Chapter 1	
1-Introduction	18
1.1-Breast medical imaging technology.....	20
1.1.1-Breast MRI.....	21
1.1.2-Breast Mammography.....	28
1.1.3-Breasts Ultrasound and ultra sound Doppler.....	34
1.2-Breast cancerous abnormalities.....	42
1.2.1-Micro calcification.....	42
1.2.2-Adenopathy.....	43
1.2.3-Asymetry in tissue.....	45
1.2.4-Asymetry in density.....	46
1.2.5- Architectural distortion.....	47
1.2.6-Mass.....	48
1.3-Abnormality quantification using BIRAD score.....	51
1.4-CAD.....	56
1.5-Summary.....	58

1.6-Current CAD systems.....	59
1.7-Aim and objectives.....	79

Chapter 2

2. Image enhancement.....	82
2.1-Methodology	82
2.2-Image enhancement.....	91
2.2.1-Noise removal.....	92
2.3-Segmentation of ROI.....	102
2.3.1-Thresholding.....	102
2.3.2-Edge base segmentation.....	104
2.3.3-Region based segmentation.....	105
2.3.4-Active contour segmentation.....	107
2.3.4.1-Active contour without edge(Chan-Vese).....	109
2.4-Feature extraction.....	112
2.4.1-Intensity based features	113
2.4.2-Shape based features.....	115
2.5-Discussion and conclusion.....	122

Chapter 3

3-Pattern recognition.....	125
3.1-Feature selection and dimension reduction.....	125
3.1.1-Sequential Floating Forward Feature Selection.....	128
3.1.2-Kruskal Wallis feature selection.....	129
3.1.3-MRMR feature selection.....	131

3.1.4-PCA.....	133
3.2-Classification.....	135
3.2.1-Classification tree and regression tree.....	135
3.2.2- Classification KNN.....	138
3.2.3-Classification SVM.....	142
3.2.4-Classification ensemble	147
3.3-Meta Data.....	151
3.4-discussion for pattern recognition.....	154

Chapter 4

4-Experimental results.....	157
4.1-Introduction.....	157
4.2- Methodology.....	158
4.2.1-Image enhancement.....	158
4.2.2-Segmentation results.....	164
4.2.2.1- Segmentation result and its reproducibility.....	166
4.2.3-Feature extraction.....	168
4.2.4-Feature selection.....	173
4.2.5-Classification.....	176
4.2.6-CAD results on BIRAD score of 3.....	190
4.2.7-Meta data for sub classification of BIRAD score 4.....	191
4.2.7.1-Micro classification.....	192
4.2.7.2-Specularity index.....	193
4.2.8-Classification mass with BIRAD score over 4.....	195

4.3-Discussion conclusion.....204

Chapter 5

5-General discussion and conclusion.....208

References.....216

List of abbreviations.....229

List of tables:

Table 1: BIRAD scoring system.....	51
Table 2: Presents summary of all reviewed CAD systems.....	75
Table 3: Medical features.....	113
Table 4: Features and their formula.....	118
Table 5: Ensemble methods specifications.....	150
Table 6: MSE and PSNR value for three types of noises.....	161
Table 7: Presents effect of noise on classifier accuracy in both level of classification.....	161
Table 8: Min, max and average value for 35 extracted features.....	170
Table 9: SFFS and KruskalWallis number of weighted features and their impact on value.....	176
Table 10: Best results for accuracy of different classifiers in cross validation and hold out method (with different feature selection algorithms)	178
Table 11: Results for classifiers with higher accuracy (run with different number of learners)..	180
Table 12: Min, max and average accuracy results for classifier with higher accuracy (when run 10 times)	181
Table 13: Classifications accuracy results for different classifiers when all extracted features are fed in to classifier.....	183
Table 14: Classifiers average accuracy (run 10 times) when all features are fed to classifier.....	189
Table 15: Accuracy results for classifiers (extracted features from mass with BIRAD score over4 are fed in to different classifiers)	199
Table 16: Average accuracy results (10 times run) for classifier with higher accuracy (classification for mass with BIRAD score over 4)	202

List of figures:

Figure 1: MRI device.....	21
Figure2: MRI background enhancement.....	24
Figure 3: T1W and T2W of a mass (MRI)	27
Figure 4: MRI curves.....	28
Figure 5: Mammography unit.....	29
Figure 6: Different grades of breast density.....	31
Figure 7: Four grades of breast density.....	31
Figure 8: Breast zones in mammogram.....	33
Figure 9: Some abnormalities in breast mammogram	34
Figure 10: Ultrasound machine.....	33
Figure11: Different quadrant of breast (in ultrasound)	40
Figure 12: Different abnormalities in ultrasounds	41
Figure13: Different micro calcifications.....	43

Figure 14: Different distribution of micro calcification.....	43
Figure 15: Adenopathy.....	45
Figure16: Asymmetry in tissue.....	46
Figure 17: Architectural distortion.....	47
Figure18: Mass with different BIRAD scores.....	49
Figure 19: Mass with different BIRAD scores (over 4).....	54
Figure 20: Original mammogram and mammograms after employing noise on them.....	93
Figure 21: Mammograms with noise and mammograms after applying median filter.....	96
Figure 22: Mammograms with noise and mammograms after applying adaptive median filter	99
Figure23: Mammograms with noise and mammograms after applying Wiener filter	101
Figure 24: Threshold segmentation of ROI.....	103

Figure 25: First initialization of ROI and edge based segmentation105

Figure 26: First initialization of ROI and Cahn-Vese segmentation.....112

Figure 27: Mass with different shapes, margin's shape and densities.....117

Figure28: Mass with different BIRAD score(over4)153

Figure 29: Different initialization and segmentation of ROI165

Figure 30: Reproducibility.....168

Figure 31: Manual feature selection.....169

Figure 32: Confusion matrix and ROC curve for SFFS feature selection with method combined with boosted tree classifier and Ada boost ensemble method with 100 learners (classify mass with BIRAD score 2 and over4)182

Figure 33: Confusion matrix and ROC curve for boosted tree classifier and ada boost ensemble method and 200 learners (classify mass with BIRAD score 2 and

over4 when no feature selection method is used)	189
.....	189
Figure 34: Mass with micro calcification.....	193
Figure 35: Extract margin of ROI.....	194
Figure 36: Confusion matrix and ROC curve for boosted tree classifier and bag ensemble method and 200 learners (classify mass with BIRAD over 4)	
.....	202

List of diagrams:

Diagram 1: Steps for automatic diagnosis of detected mass in mammogram.....	83
Diagram 2: Training classifier based on mass with IRAD score 2 and over 4.....	161
Diagram 3: Training classifier based on mass with IRAD over 4(a, b and c)	161
Diagram 4: Evaluation flow of proposed CAD.....	162
Diagram 5: Evaluation of CAD by mass with noise.....	163

Acknowledgment:

I would like to thank my first supervisor, Prof. Dehmeshki, for all his helps. He devoted lots of his time for me during all these years. He taught me teaching methods as well as guides me morally and professionally to complete my study. Therefore, I would like to express the deepest appreciation to him. without his support and persistent help this dissertation would not have been possible.

I would like to thank Prof. Tim Ellis, my associate professor, who always advised me and guide me through my studies.

I would like to thanks my spouse, Morad, without his support and motivation I couldn't do this course.

My mother, always accompany me during my visits to London and took care of my son so I could come to university, without her it was impossible.

I would like to thank Dr. Neda Nasser, specialist in radiology and director of "Farokhi Yazdi medical imaging center". She helped me a lot in collecting data and she taught me all the clues for diagnosing and reporting a mass.

I appreciate Dr. Hasanizadeh, specialist in radiology in "Haghighat medical imaging center", she is one of the top 2 specialists in reading mammograms in Iran. She devoted lots of her times in 2 first years of my studies. So I could get

enough knowledge in all medical imaging techniques, breast abnormalities to get enough motivation and confidence to select such project.

Last but not the least, I appreciate the time that my dear examiners devote and studies my project.

I dedicate this thesis to my son, Sam, I wish he learns that beside all my difficulties during 6 years of part time PhD research, how I passed all the obstacles and overcome this project. Being a mother, spouse, student and teacher, I managed my time and finished the aim that once I started, whilst enjoying every step of it. No matter what happens, you should finish the job that you have aimed.

Chapter 1

1. Introduction

Breast cancer is one of the most common cancers in women. Fortunately, death caused by breast cancer decreased over the past 20 years [James,2008]. Experts believe use of medical screening techniques such as Mammography, ultrasound and MRI can assist specialists to diagnose any cancerous abnormality in early stages [James,2008,Boris,2002, Balleyguier 2007]. The most common abnormalities in breast which may lead to cancer are: micro calcifications, asymmetrical breast tissue, asymmetrical density, tumor (mass), architectural distortion and adenopathy [Boris 2002, Balleyguier 2007, American College of Radiology 1998, Liberman 2002]. Specialists by reading medical image, diagnose the abnormality and assign a BIRAD score (Breast Imaging Reporting and Data System) associated to the abnormality. According to the BIRAD, patient will continue the treatment procedure [Balleyguier 2007, American College of Radiology 1998, Liberman 2002, Leconte 2003]. For diagnosing any possible abnormality, specialist mostly starts the process by ultrasound or Mammography of patient and according to the found abnormality they may or may not use other medical imaging techniques [Balleyguier 2007, American College of Radiology 1998, Liberman 2002, Leconte 2003]. There are some difficulties in diagnosing procedure that may lead to false diagnose or inconsistency due to complexity of abnormality.

Besides, the process of reading images are time consuming as the breast tissue in each patient is different, even for the same patient in different periods and times, breast texture may differ which leads to high probability of miss diagnosis [American College of Radiology 1998, Liberman 2002, Leconte 2003, Berg 2004]. All these mentioned facts make the diagnosing procedure difficult and thus requires specialist's full concentration.

The accuracy of this procedure highly depends on specialist's adequate time and concertation. Specialists read many medical images in their working time with high concentration. Lots of thing may happen that can distract specialist attention or they may have a bad day or mood. All these points could affect their diagnosing; hence their diagnosing accuracy may not be consistent [Karssemeijer 2004, Freer 2001, Samulski 2010].

To overcome such problem, specialist can use Computer Aided Diagnosis(CAD); which is explained more in section 1.4; as a second opinion to assist them to make their final decision. Considering all these points the aim of this project was to design a CAD tools that automatically analyze a detected abnormality. Prior to introducing the proposed CAD system. In next following sections, different medical imaging techniques, their specifications and diagnosing procedures, different type of abnormalities and BIRAD scoring system are explained. Finally, this chapter was completed by introducing and comparing current existing CAD

system in order to understand the resolved problems and those that are not fully solved.

1.1 Breast medical imaging techniques

As mentioned in introduction for early and accurate diagnosing any breast abnormality, specialist may use one or combination of medical imaging techniques. In order to fully comprehend specification of each medical imaging device, imaging trend, reporting and detecting an abnormality, beside studding text books, several meeting has been made with specialist in “Haghighat medical imaging center”¹ (Dr. Hasanizadeh ²) and “Farokhi medical imaging center”³ (Dr. Neda Nasser⁴). All figures presented in this thesis are the data that I have collected from these medical imaging centers, except the ones that is mentioned under the images. I have the ethical approve for using these data from both medical imaging centers. As a result of such investigation, Below, imaging techniques (MRI, Mammography and ultrasound) and their efficiency and deficiencies are described.

¹ Haghighat medical imaging center, Iran, Tehran, #706, east Janbazan street, Nabovat Square. Tel: +98-2-77957519

² Parto Hassanizadeh, specialist in Radiology, Tehran University, Iran, Tehran, 16azar street, Enghelab street, Enghelab square.

³ Farokhi Yazdi medical clinic, Iran, Tehran, Farokhi Yazdi square, Shahis Keshvari street, Pasdaran. Tel: +98-21-22842000

⁴ Neda Nasser, specialist in radiology, Shahid Beheshti university, Iran, Tehran, Daneshjoo Boulevard, Velenjak. Tel: +98-912- 100-2064

1.1.1 Breast Magnetic Resonance Imaging (MRI)

One of the most accurate and popular medical imaging technique is MRI which uses a combination of large magnet, radiofrequencies and a computer to produce images of an organ structure [Doyle 2005].

MRI machine is a tube shaped machine that creates a magnetic field around the patient. The magnetic field pulse with radiofrequency, alter the hydrogen atoms natural alignment in the body, the radio waves knock the nuclei of the atom in the body, out of their normal position and they send out radio signals and these signals are received by the computer witch compute them and convert them into 2d images [James 2008]. Image of an MRI devise is presented in (Fig.1).



Figure 1: MRI device. Image from reference 2, Andersson,1997

MRI has its own ability and deficiencies such as:

Some of most important MRI key features could be mentioned as follow

[Lincott Williams & Wilkins Publishers 2010, Aghaei 2016]:

- Detection of a probable abnormality found in the patient's Mammography.
- Finding probable abnormality in the woman with dense breast tissue, whom their mammography could not be reliable
- Evaluate the exact size and location of breast cancers.
- Double check existence of asymmetry of tissue and density found in the Mammography and etc.
- End of the breast could be completely checked in MRI, which in other techniques it may not

It also has its own deficiencies such as:

- It could not always distinguish the difference between cancerous abnormalities and non-cancerous abnormality, referred to "false positive" test results.
- MRI is unable to identify micro calcifications that can indicate breast cancer [James 2008].

Reading breast MRI

In Haghghat medical imaging center Siemens MRI is used (Avanto 1.5 Tesla) For breast MRI, patient usually lies face down. Her breast positioned through opening space in the table. Technologist watches the procedure through the

window, monitoring breast position and any potential patient's movements [Båth 2007].

A breast MRI usually requires a contrast Dye that is injected before or during the procedure. The dye helps to create clear image and outline abnormalities more easily. Tissue that absorbs color will be appeared. For example, mass absorbs color but fibro adenoma does not. There exist different brands for the injection but in Haghghat medical imaging center they use Dotarem. Patients less than 80Kg will take just one dose but patients over 80 Kg will be injected two dosages before the screening and then 6 Spairs are taken, each one approximately takes 1 – 1.4 minute. It is important that the patient doesn't have any sudden movement or deep inhale or exhale during this procedure. At the end Spairs are subtracted in order to find any changes. Specialist starts with sub 3 but if they find a sudden change in sub 3 they will also check sub 1 and sub 2. They also use Tirms sequences to check the abnormality status in it [Doyle 2005]. Beside these points specialists consider several other things such as: background enhancement, morphology, dynamic curve, T2w, T1w and location, which they will be discussed below.

Background enhancement

As mentioned above specialist need to consider background enhancement as a part of their diagnosing procedure. Based on amount of absorbed injected dye,

specialists consider 4 background enhancements: minimal, mild, moderate and severe enhancement background [James 2008]. MRI with severe enhancement background are not very reliable. Patients with dense mammogram, which is not reliable for Mammography, are recommended to have MRI and patients with severe MRI also are recommended to have Mammography or ultrasound [Karssemeijer 2004, Freer 2001, Samulski 2010, Doyle 2005].

Figure 2 (a) shows MRI of minimal background and (b) shows MRI of severe background enhancement.

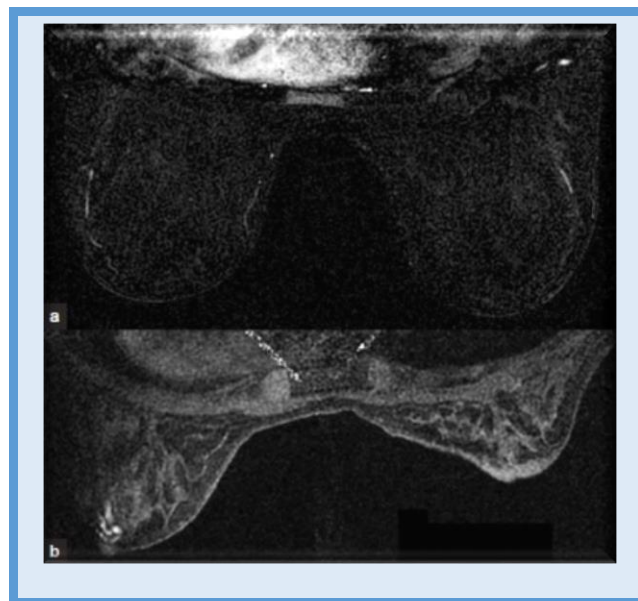


Figure2 :(a) Minimal background, (b)Sever background in MRI. Image from reference 2,

Andersson,1997

Morphology

Beside background enhancement morphology also is consider in diagnosing procedure. Morphology refers to shape and margin and internal enhancement

of the mass. Mass should be characterized according to shape of mass and its margin. A mass could be: round, oval, lobulated, or irregular. Margins could be: smooth, irregular, or speculated.

Internal enhancement: homogenous, heterogeneous, rim, dark internal septation, enhancing internal septation, or central. distribution of a non-mass enhancement could be: focal, linear, ductal, segmental or diffuse [James 2008].

More irregular and speculated masses have a higher likelihood of malignancy. Specific internal enhancement patterns are often associated with certain entities: rim-enhancement is seen with high-grade invasive ductal carcinoma, cysts with inflammation, and fat necrosis; dark internal septation may be seen with fibro adenomas; enhancing internal septation are often seen with malignancy; central enhancement is seen with high-grade ductal carcinoma and vascular tumors [Doyle 2005].

T1-weighted (T1W) T2-weighted (T2W) MRI

Beside background enhancement and morphology, T1-weighted (T1W) T2-weighted (T2W) MRI are also considered in diagnosing procedure [Aghaei 2016] [Doyle 2005].

Image acquisition is performed in an axial plane with 2 mm sections. Sagittal and coronal reconstructions are made from this dataset. Sagittal image acquisition

is usually preferred for biopsy procedures. The primary pulse sequences are fat-suppressed axial T1-weighted (T1W) without and with contrast (T1 shows the anatomy of the breast: fibroglandular tissue which is shown white and fat which is shown black and fat-suppressed axial T2-weighted (T2W) or short TI inversion recovery (STIR). A minimum of two post-contrast T1-weighted series should be obtained, with initial post-contrast images within 4 min and delayed post-contrast images within 8 min after contrast administration. Before the injection (in T1W) they look for any abnormal texture, distortion or mass, hence they will specify the size, location, texture, shape and margin of it. Specialist will consider T1 fat-suppressed images, maximum intensity images (MIP) or T1 subtracted images (T1 sub) to detect any abnormality if they could not verify the detected abnormality as benign they will also consider T2 images without fat suppression to search for fatty content of the lymph node or the fluid composition of the cyst. If they find an abnormality in T1 they will check it in 3-dimensional view (L, A, AND F). Also they will check the abnormality in the same position in T2. TIRM sequences are also considered in which edema and lesions are white and fats are black and fibroglandular tissue are white. Masses are seen black in T1 but they look white in T2. Below (Fig. 3-a) T1w and (Fig. 3-b) T2w of a well-defined mass is presented.

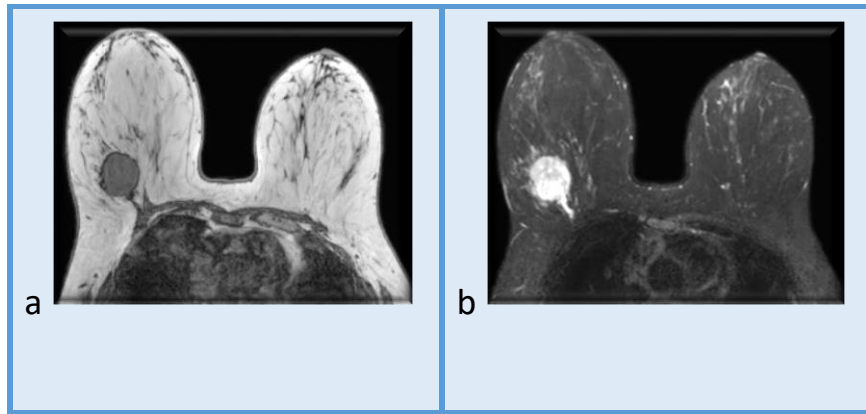


Figure 3: a-T1W of a mass with BIRAD score 2, b- T2W of a mass with BIRAD score 2

Dynamic curve

If they find any mass they will annotate region of interest (ROI) with mouse in the middle of the mass in the computer, where it is most enhanced, and for its six stages of injection they provide a dynamic curve for it, showing speed of absorbing and losing the dye, which helps the specialist in diagnosing the mass. It has 3 types of curves. Curve type 1, progressive absorbing, is normal curves which show a benign mass and curve type 3, washout, shows malignant mass [Liincott Williams & Wilkins Publishers 2007,Aghaei 2016].

Type-I curves are slowly enhancing, in which gradual steady enhancement occurs. Malignancy is seen in approximately 6% of lesions with a Type-I curve.

Type-II curves show early strong enhancement (increase over a 1–2 min period) with a subsequent plateau phase. Malignancy is seen in approximately 6–29% of lesions with a Type-II curve.

Type III or “washout” curves show early strong enhancement (over 1–2 min), with subsequent decline in enhancement. This produces a characteristic peak dubbed the “the cancer corner,” and is strongly associated with malignancy. Malignancy is seen in approximately 29-77% of lesions with a Type-III curve.

Both Type-II and Type-III curves should be considered suggestive of malignancy.

Figure 4 shows three types of curves.

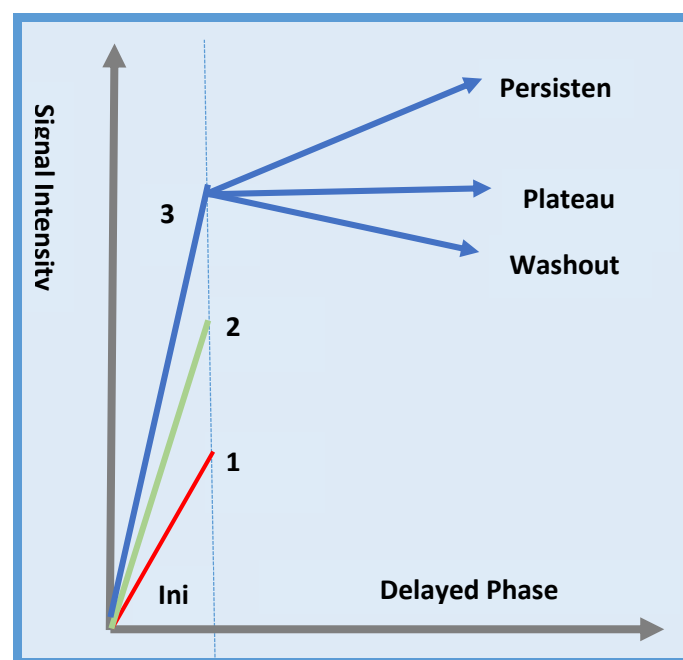


Figure 4: Three types of curves are presented, type 1 curve in 94% of cases are benign mass, type 2 curve in 69-94% cases are benign mass and type 3 curve are mostly suspicious of malignancy.

1.1.2 Breast Mammography

Mammography uses low dose X-ray system to examine breast. Imaging with x-ray involves exposing a part of body to a small dose of radiation to produce pictures of the inside of the body. In Digital Mammography x-ray is replaced by solid state detectors that covert x-ray into electrical signals. The electrical signals

are used to produce images of the breast on the computer or printed on the special fields [Lehman 2002, Mendez 2004, Baker 1996]. Mammogram is used to detect any early symptoms of cancers in early stages specially mass and micro calcifications, cyst, and fibro adenoma. Women with age of 40 and over are recommended to have Mammography every year and they are considered high risk patients. A Mammography unit is rectangular box that houses the tube in which x-ray are produced, with special accessories that allow only the breast to be exposed to the x-ray. Attached to the unit is a device that holds and compress the breast and position it in a way that images can be obtained in different positions. Figure 5 shows a Mammography unit [Arevalo 2016].



Figure 5: Mammography unit. Image from reference 2, Andersson,1997

Ability and deficiency of mammogram

- Asymmetry in tissue and density may not be completely reliable in Mammography and they need to be checked in other type of medical imaging techniques.

- If the technician does not position the patient exactly or does not pull the breast completely, end of the breast could not be seen completely in the mammogram.
- Mammogram is the best way to detect any existence of micro calcifications [Lehman 2002, Mendez 2004, Baker 1996].

How specialists read a mammogram

In Haghghat imaging center Hologic Selenia machine is used. Breast density is a key point in Mammography. Breast density is divided into two parts: fat tissue and fibro glandular tissue. More fat tissue, makes Mammography more reliable. Based on density of the breast they may or may not rely on Mammography [Kerlikowske 1996, Ikeda 2007]. They considered 4 grades for the density of breast (grades 1-4).

Grade 1: Fatty breast, refers to breast with more than 75% fat tissue.

Grade 2: Scattered fibro glandular breast, is a breast with 50% to 75% fat tissue.

Grade 3: Heterogeneously dense, is a breast with 25% to 50% fat tissue.

Grade 4: Extremely dense, is a breast with less than 25% fat tissue.

If size of fibro glandular tissue is more than 75% (stage 4) they may not rely on Mammography and uses another imaging technique [Boris 2002, Balleyguier 2007]. In figure 6-a breasts with grade 1 and figure 6-b breast tissue with grade 4 is presented. Figure 7 all 4 grades of breast density are presented.

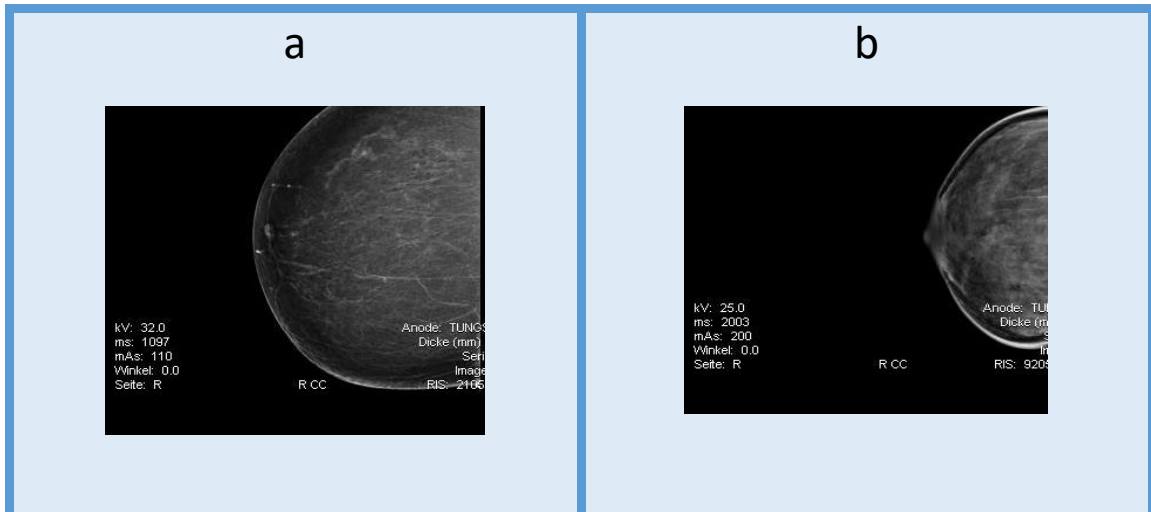


Figure 6: a) Breast with density grade 1 with 75% fat tissue which is reliable for mammography, b) Breast with density grade 4 with less than 25% fat tissue which is not reliable for mammography

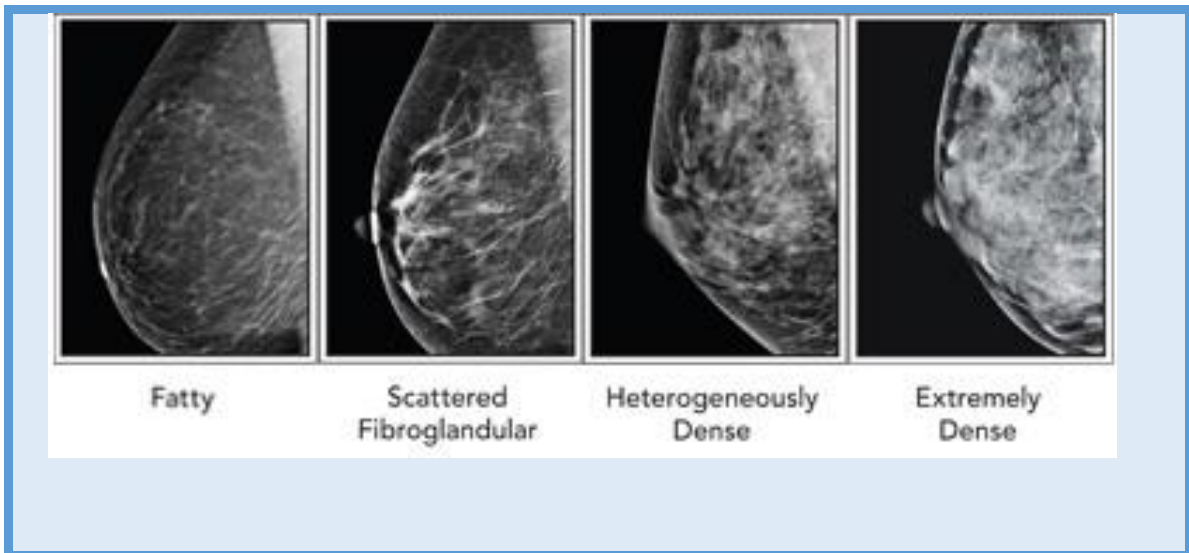


Figure 7: Four grades of breast density. Image from reference 2, Andersson,1997

In Mammography the patient will be asked to stand in front of the machine. In Mammography there are pictures of both breasts (right and left) in two directions: Cranial-Caudial (CC) which is view from above (Fig.9-b) and Medio Lateral-Oblique(MLO) (Fig.9-a), which is angled view (45 degree) Of the breast. So there are four images: RCC, LCC, RMLO, LMLO. Specialists compare right and left breasts in both directions. Comparing vertical images of right and left breast,

they can find any probable abnormality (Adenopathy, micro calcification or tumor) hence they will check the same part of breast in other direction to double check the existence of the abnormality and also to find any other probable abnormality [Lewin 2001].

Asymmetry in tissue or density may not be completely recognized in Mammography and they need to double check in other type of breast medical image [Rufus 1998].

Positioning an abnormality in breast Mammography

Positioning an abnormality in breast is of great importance for these reasons:

- It is used by Oncologist for further treatment.
- If they want to check the abnormality in other imaging techniques, they should know the exact position of the abnormality.
- Mostly there is not just one abnormality. In order to be able to differentiate them in more than one imaging technique exact positioning is really important.

For positioning an abnormality, specialists divide breast image into different zones which is presented in (Fig.8).

From up to down part of the breast: superior mammary zone, mid mammary zone and inferior mammary zone. From front to back of the breast: anterior,

medial and posterior. From left to right: medial and lateral zones [Metz 1998,A.C.S. 2010].

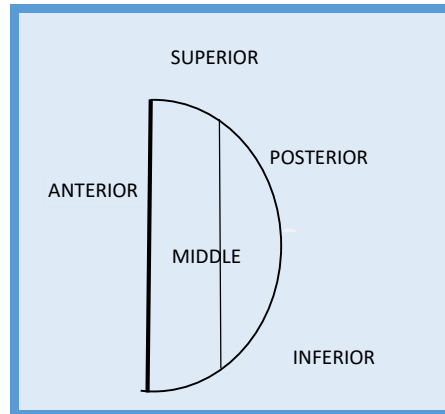


Figure 8: Breast zones in mammogram

Figure 9 shows mammogram with asymmetry in tissue in RMLO, asymmetry in density in LCC, mass in RCC and cluster micro calcifications in LCC.

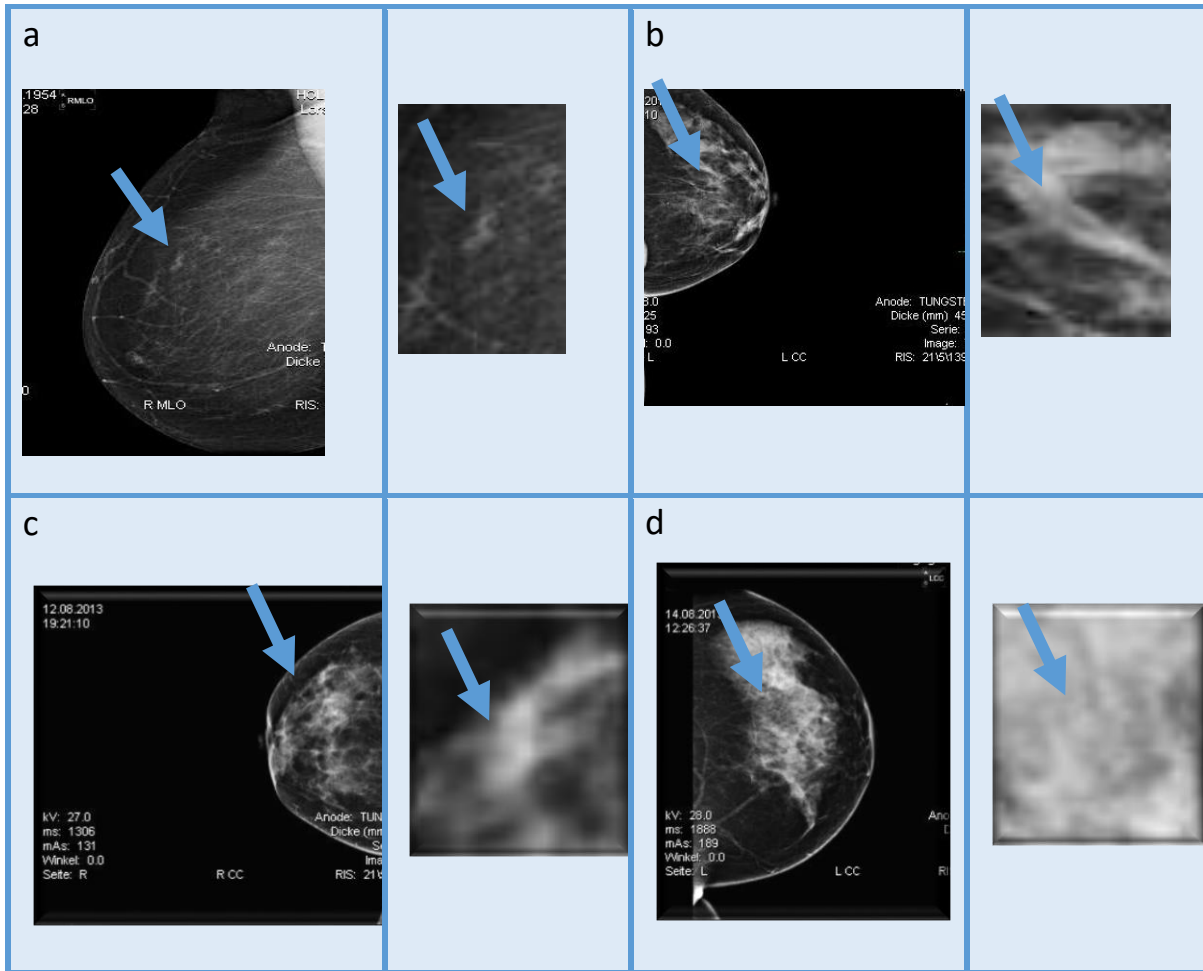


Figure 9: Some abnormality in breast mammogram a- Asymmetry in density, b- Asymmetry in density, c- Mass, d- Micro calcifications

1.1.3 Breast Ultrasound, ultrasound Doppler

Ultrasound is the most common medical imaging technique used for diagnosing any abnormality in breast. An ultrasound uses high-frequency sound waves that are transmitted through breast tissue from a hand-held unit called a transducer [Balleyguie2007]. These sound waves bounce off breast tissues. The "echoes" created are recorded by a computer that makes an image of the breast tissue and displays it on a monitor. No radiation is used, and very little pressure is

required. Ultrasounds produce sharp, high-contrast images. In dense breast tissue, the ultrasound can create an image that often allows a doctor to distinguish between a fluid-filled cyst and a solid mass. A breast ultrasound examination is not considered a screening test, but an investigative technology used for taking a closer look at areas of the breast that the doctor still has questions about after doing a mammogram and clinical breast exam. An ultrasound test may be useful if patient's mammogram shows an indistinct mass, or if a lump can be easily felt during a clinical breast exam or a probable asymmetry in tissue or density. In the past ultrasound was limited because the examiner could not cross correlate the mammographic finding with the sonographic information [Berg 2004] [American College of Radiology 1998]. To be able to cross correlate sonography with Mammography one should be familiar with normal mammographic and sonography of the breast and have optimal sonographic equipment [James 2008].

In past there were few breast sonographic applications, hence they require the least sophisticated equipment. If patient wishes to have optimal image of the mass and high rate of localizing solid mass, patient will need sophisticated machines with frequencies ≥ 10 MHz (frequency between 7 to 15 MHz) and linear transducer [Athira 2016].

As many breast structures are small, high frequency in sonography is important. Normal breast structure as duct and terminal duct lobular units are small [Athira 2016, Kannan 2016]. High spatial resolution allows specialist to recognize normal breast architecture and identify small malignant mass in ductal system and help specialist to better characterize the mass, see the margin, speculation and architectural distortion of the mass. Also in big breast for having exact image of back of breast lower frequency of the transducer can help the specialist. Beside high frequency for having image quality excellent contrast resolution is required. The reason that contrast resolution is important is that patient must be able to distinguish variety of masses from background parenchyma. When the breast is fatty focal masses such as cyst, lymph, fibro adenoma may be difficult to identify and when breast is dense, fat necrosis, surgical or radial scars may be hard to locate. High resolution can improve contrast resolution because the assignment of gray shade is more precise. Reducing the dynamic range may improve contrast resolution; this method exaggerates difference gray shades. Variety of programs improves contrast resolution by enhancement of specify gray shades on a point by point basis or region by region basis. Beside image contrast, specialist should optimize resolution. These methods include increasing line density of image, increasing the persistence and adjusting focal zones. The main disadvantage of this method is lower frame rate [Berg 2004].

Color or power Doppler ultrasound is a method for assessing vascularity. Breast vascularity is low. This means specialist should use lower frequency for power Doppler comparing to the frequency used for gray scale ultrasound and the filter and scale should be low. The Doppler gain is optimized by initially increasing the gain until the entire screen is filled with color then slowly reducing the gain until the color appears only in pulsating vascular structure. If no color is detected using these methods, then the sample size should be increased which will reduce color resolution. Color may be present outside of the vessel's walls. Doppler is useful to delineate vessel, it's also useful to verify if a hypo echoic or anechoic mass is cystic or solid [Doyle 2005].

Dynamic clips are used to demonstrate special relationship of multiple lesions. They are ideal to show color flow in pseudo aneurysms or intravenous contrast enhancement of solids masses. Before existence of 3d ultrasound dynamic clips were the best way to demonstrate relationship of multiple cyst to a solid mass or to shadow debris or calcifications moving in a complex cyst.

Figure 10 present an ultrasound machine.



Figure 10: Ultrasound machine. Image from reference 2, Andersson,1997

Ability and deficiency of ultrasound

Some significant deficiencies of ultrasound are as below: [Doyle 2005] [Berg,2004]

- Micro calcifications could not be seen in ultrasound vividly.
- Tissue distortion and asymmetrical density also could not be diagnosed in ultrasound.

And some of its abilities are as below:

- Doing an ultrasound is not as expensive as other type of medical imaging techniques.
- An ultrasound uses high-frequency sound waves so it is not harmful for the body.
- Cyst and solid masses can be differentiated easily in ultrasound while it could not be diagnosed easily in Mammography.

- All patients can have ultrasound test. Some patient may not have MRI, for example patients with Protez , sensitivity or pacemaker [Berg 2000].

Doing an ultrasound test

In Haghghat imaging center machines with frequencies 14 MHz is used. Ultrasound imaging requires a skilled operator who can examine suspect areas of the breast by positioning the transducer in several positions. The operator must decide when to reposition the transducer, or the patient, in order to get the best images. There are different models for positioning the patient and transducer's movement. For having a full view of all aspects of the breast they lay down the patients in two dimensions: face up and 45 degrees. At each position specialist move the transducer in each quadrant of breast in two dimensions: radial and anti-radial. Hence they can have a full view of the breast tissue and in case they find an abnormality they will try to have a complete view of it in different directions. Women with big breast will be asked to be sited so the specialist can see the superior section of breast more completely [Doyle 2005].

Positioning abnormality in ultrasound

For positioning an abnormality specialist use the "o'clock" method and quadrant method. The o'clock method views the breast as circular clock with the nipple in the center of the circle. 12 o'clock is directly above the nipple, 3 o'clock is left of nipple, 6 o'clock is below the nipple and 9 o'clock is right of nipple. The quadrant method divides the breast into 4 quadrants. These quadrants are defined by a

horizontal and vertical line through the nipple. They label 4 region of the breast: upper outer quadrant(UOQ), upper inner quadrant(UIQ), lower outer quadrant(LOQ)and lower inner quadrant (LIQ) [Berg 2004]. In Haghightat imaging center beside these two methods they also use Mammography breast zones (anterior, posterior and middle) for exact positioning of the abnormality. Below, in figure 11 different quadrant of the breast is shown.

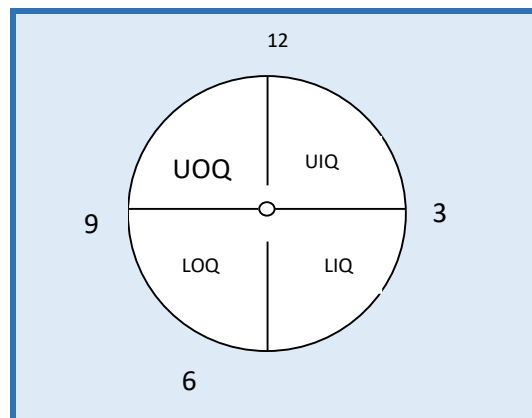


Figure 11: Different quadrant of breast, upper outer quadrant(UOQ), upper inner quadrant(UIQ), lower outer quadrant(LOQ)and lower inner quadrant (LIQ)

Figure 12 ultrasounds with different diseases are presented.

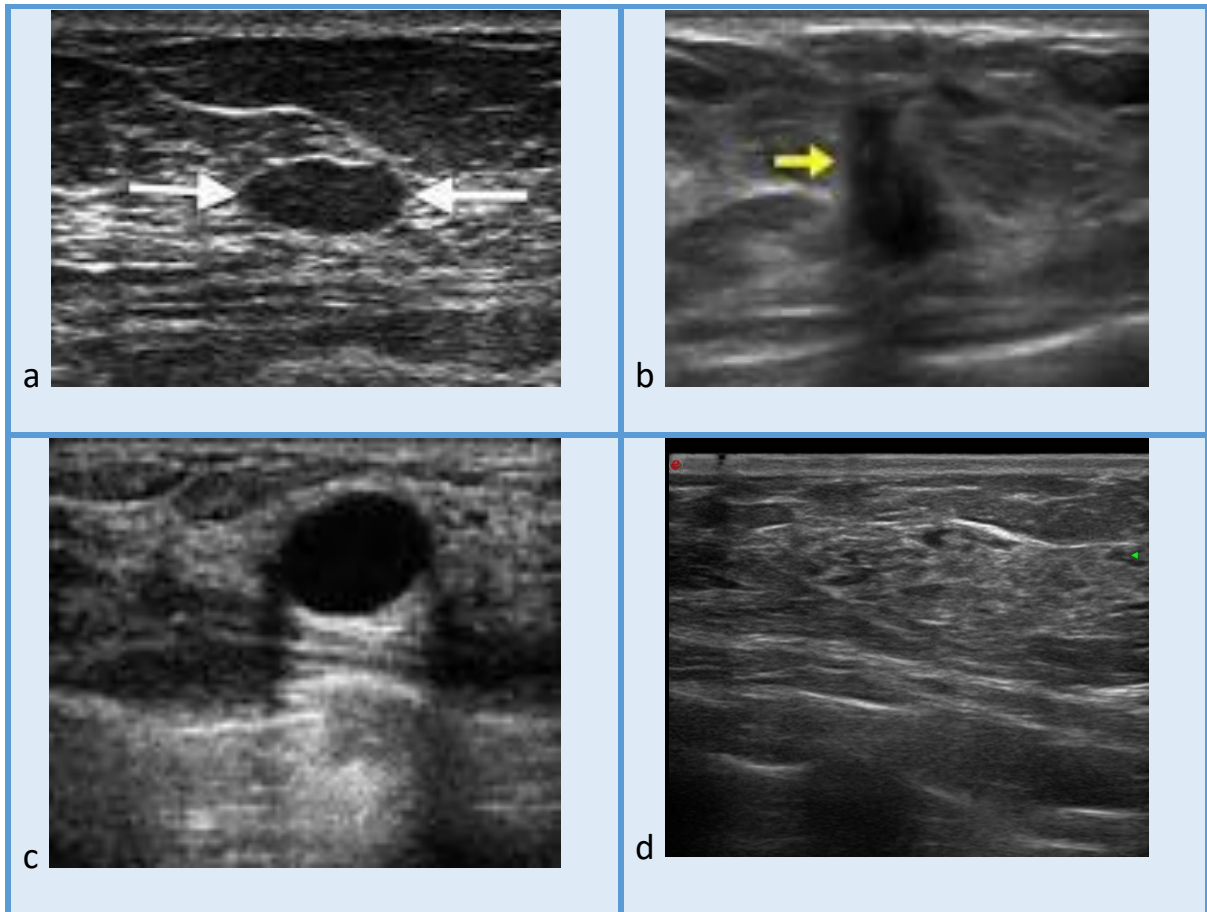


Figure 12: Ultrasounds with different diseases. a-Round oval benign mass, b-Taller than wide mass, c-Cyst, d-Adenopathy

Summery

In section 1.1 breast medical imaging techniques, their specification, breast imaging trend and diagnosing procedure are explained. It was mentioned in introduction that specialist starts the diagnosing procedure by ultrasound as it is harmless and not expensive. Women with the age of over 40 are considered high risk so its recommended that they do Mammography every year as their routine checkup. In next section different types of breast abnormalities will be explained in more details.

1.2 Breast's cancerous abnormalities

In previous section breast medical imaging techniques, imaging trend and diagnosing procedure are explained. In this sectional different breast cancerous abnormalities will be explained in order to find the most critical and common one among women.

1.2.1 Micro calcifications

As mentioned in introduction, one of breast cancerous abnormality is micro calcifications. One of the main ways to detect cancerous micro calcifications in early stage is Mammography [Pijnappel 2004,Gulsun 2003]. Micro calcifications are actually tiny specks of mineral deposits such as calcium and they can be distributed in various ways. Sometimes micro calcifications are found scattered throughout the breast tissue, and they often occur in clusters. Different types of distribution of micro calcification are presented in (Fig. 14). Most of the times, micro calcification's deposits are due to benign causes [Samulsk 2010,Cheng 1998]. However, certain type of presentation of micro calcifications are more likely to be associated with malignant breast cancer. "Powedish" micro calcifications with either a fine, indiscernible, or 'cotton ball' appearance, most frequently results in a 'low-grade' cancer (Fig. 13-a). When the micro calcifications have the 'Crushed Stone' characteristic, appearing either as coarse, granular, angular, broken-needle-tip, arrowhead, or a spearhead shape, the

breast cancer classified as low to intermediate-grade(Fig.13-b). But if the micro calcifications have a 'Casting' appearance, the breast cancer classified as high-grade. Casting micro calcifications typically appear in two variations (Fig.13-c) [Balleyguier 2007]. Different types of presentation of micro calcifications are presented in (Fig.13).

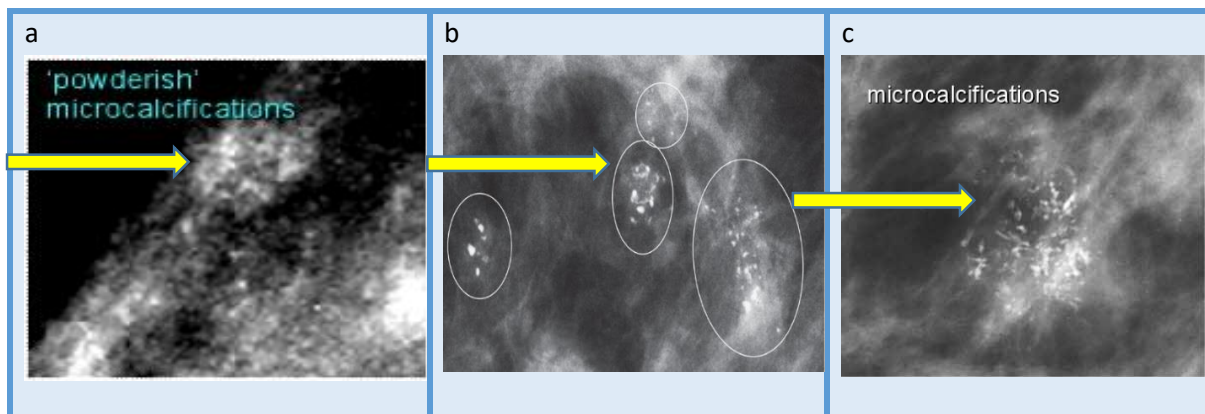


Figure13: Different micro calcifications: a) Powderish micro calcifications, b) Crushed stone micro calcification, c) Casting micro calcifications. Image from reference 3, A. C. S. (AMS), 2006

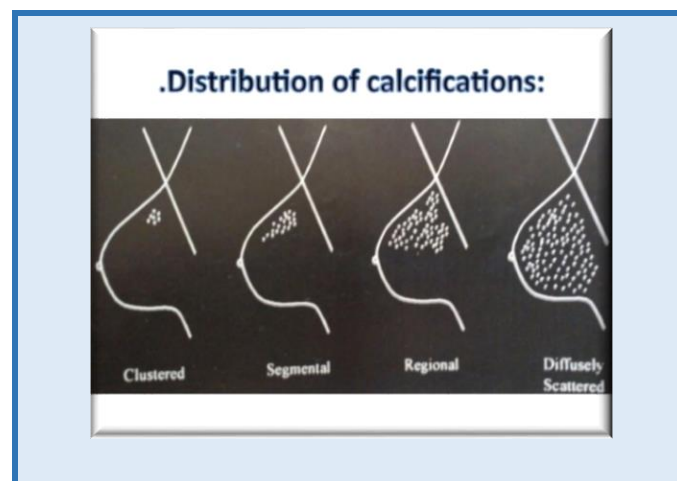


Figure 14: Different distribution of micro calcification. Image from reference 2, Andersson,1997

1.2.2 Adenopathy

Adenopathy refers to an enlarged lymph gland, and a very small percentage of women undergoing a breast cancer screening mammogram will have this

abnormality [Anderson 1997, A. C. S. 2006, Cheng 1998]. An axillary lymph node that seems enlarged on a mammogram could contain cancer, but some lymph nodes can normally be quite large and they are benign. However, mammographic features of benign and malignant lymph nodes are quite often indistinguishable. Sometimes the presence of intra nodal calcifications can be more suggestive of malignancy. An axillary lymph node is suspicious if its size is greater than 2 cm and there is no fatty helium. When a lymph node has fatty helium, the thickness of outer cortex shouldn't be more than 5 millimeters. When the cortex is 6mm or thicker, chances of cancer spread into the lymph node are significant. If there is no fatty helium in the lymph node, then its smallest short-axis width should be larger than 10 millimeters. Ultrasound is often used as a follow-up when enlarged lymph glands are detected. Some of these will be referred for either a fine needle or excision biopsy. Sometimes a suspected enlarged lymph node turns out to be either a lymphoma, fibro adenoma, or a hematoma. Common benign causes of benign lymph Adenopathy might also include reactive nodal hyperplasia, or collagen vascular disease. An acute bacterial infection or tuberculosis might also bring about the condition. If the lymph Adenopathy is actually caused by a malignant carcinoma, it is often associated with breast cancer development in the previously unaffected breast

[Lehman 2002, Mendez 2004, Baker 1996, Cheng 1998]. Figure 15 present a Mammography with adenopathy.

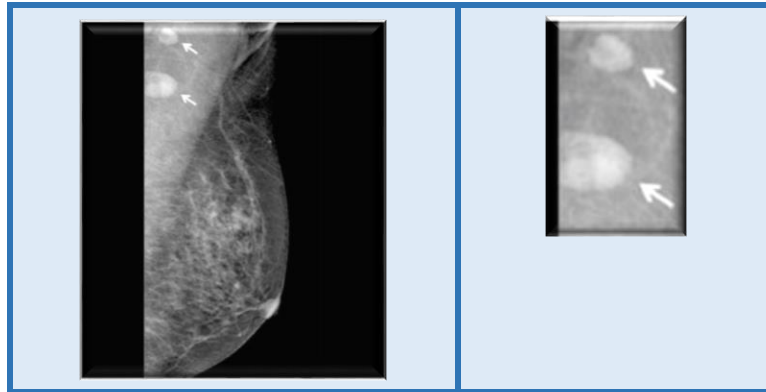


Figure 15: Present adenopathy in breast

1.2.3 Asymmetry in tissue

Another kind of abnormality is asymmetry in tissue and density. The term asymmetric breast tissue refers to a greater volume or density of breast tissue in one breast rather than in the corresponding area in the contra lateral breast. Asymmetry in tissue is a fairly vague finding in which there is no focal mass, no distorted architecture, no central density, and no associated breast calcifications [Mendez 2004]. Asymmetrical breast should only be concern, when it is also associated with a clinically palpable breast asymmetry. Although asymmetry is often a normal finding, additional evaluation may sometimes be required. Specialist will check the same part of breast in another direction and according to that they will ask for immediate follow up through ultrasound or MRI, etc. Because some times when they check the same part of breast in the ultrasound

they will find that the found asymmetry in Mammography was just overlapped tissue of breast. [Cheng 1998, Markey 2002, Mehul 2005]

1.2.4 Asymmetry in density

A density that is seen on only one standard mammographic view is referred to as Asymmetry in density. Although this finding may represent benign superimposed fibro glandular tissue, still additional imaging may reveal a true lesion. True lesions may appear on only one view because they are either obscured by overlapping dense parenchyma or located posterior and thus outside the field of view [Cheng 1998, Markey 2002, Mehul 2005, Lehman 2001]. Figure 16 present asymmetry in tissue. For detecting any possible asymmetry (density or tissue) specialist review all four mammogram images (RCC, LCC, RMLO and LMLO) from top right corner to the left bottom corner of the images. If they find an asymmetry in the breast, comparing two images (left and right), they check the same position of breast in another direction in order to proof the existence of asymmetry.



Figure 16: Asymmetry in tissue

1.2.5 Architectural distortion

In architectural distortion, a focal area of breast tissue appears distorted with no definable central mass. In architectural distortion there are speculations radiate from a common point, and there is an area of focal retraction and tethering of normal parenchyma. Architectural distortion may be associated with breast cancer because cancer infiltration can disrupt parenchymal architecture before there is evidence of a mass [Ikeda 2007].

Architectural distortions often are an accompanying feature of mass breast cancer as well. Almost 80% of mammographic ally detected breast masses which also have architectural distortions, turn out to be invasive breast cancers [Carney 2004]. Figure 17 show architectural distortion.

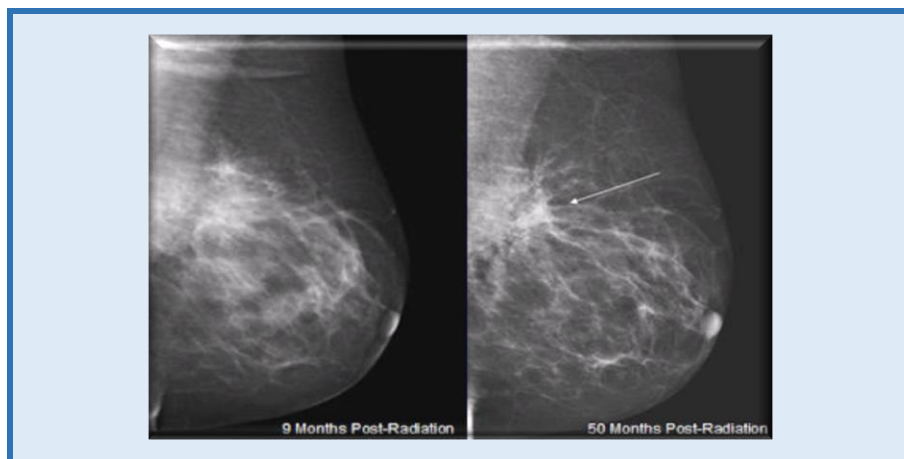


Figure 17: Architectural distortion

1.2.6 Mass

Among all cancerous abnormalities, mass is indicative of abnormality in breast.

A mass is a 'space-occupying' lesion seen on at least two projections and most mammographically detected breast masses tend to have curved, convex borders. The mass itself is typically described according to three features: shape, margin and density [Lehman 2002, Mendez 2004, Baker 1996].

If the shape of the mass is round or oval (Fig.18-e) the mass is probably benign but if the mass has an irregular shape, then it is suggestive of malignancy (Fig.18-a). If the margin's shape is smooth (Fig.18-d) it's probably benign. Likewise, there is moderate suspicion if the margin has many small lobes (Fig.18-c). Also if the margin is indistinct or speculated, then it's highly suspicious of malignancy [Ruschin 2007, Hilleren 1991, Lewin 2001] (Fig.18-a).

Considering the margin's density when the margin is circumscribed and well defined the mass is probably benign (Fig.18-d). If the margin is obscured more than 75% by adjacent tissue, it is moderately suspicious of malignancy (Fig.18-a) [Bhangale 2015, Alam 2014, Elter 2009]. Density is usually classified as either fatty, low, ISO-dense, or high density. The mass is probably benign for fatty and low densities (Fig.18-e), moderately suspicious of malignancy for an ISO-density (Fig.18-c), and highly suspicious of malignancy for high densities (Fig.18-b) [Ruschin 2007, Hilleren 1991, Lewin 2001, Kerlikowske 1996].

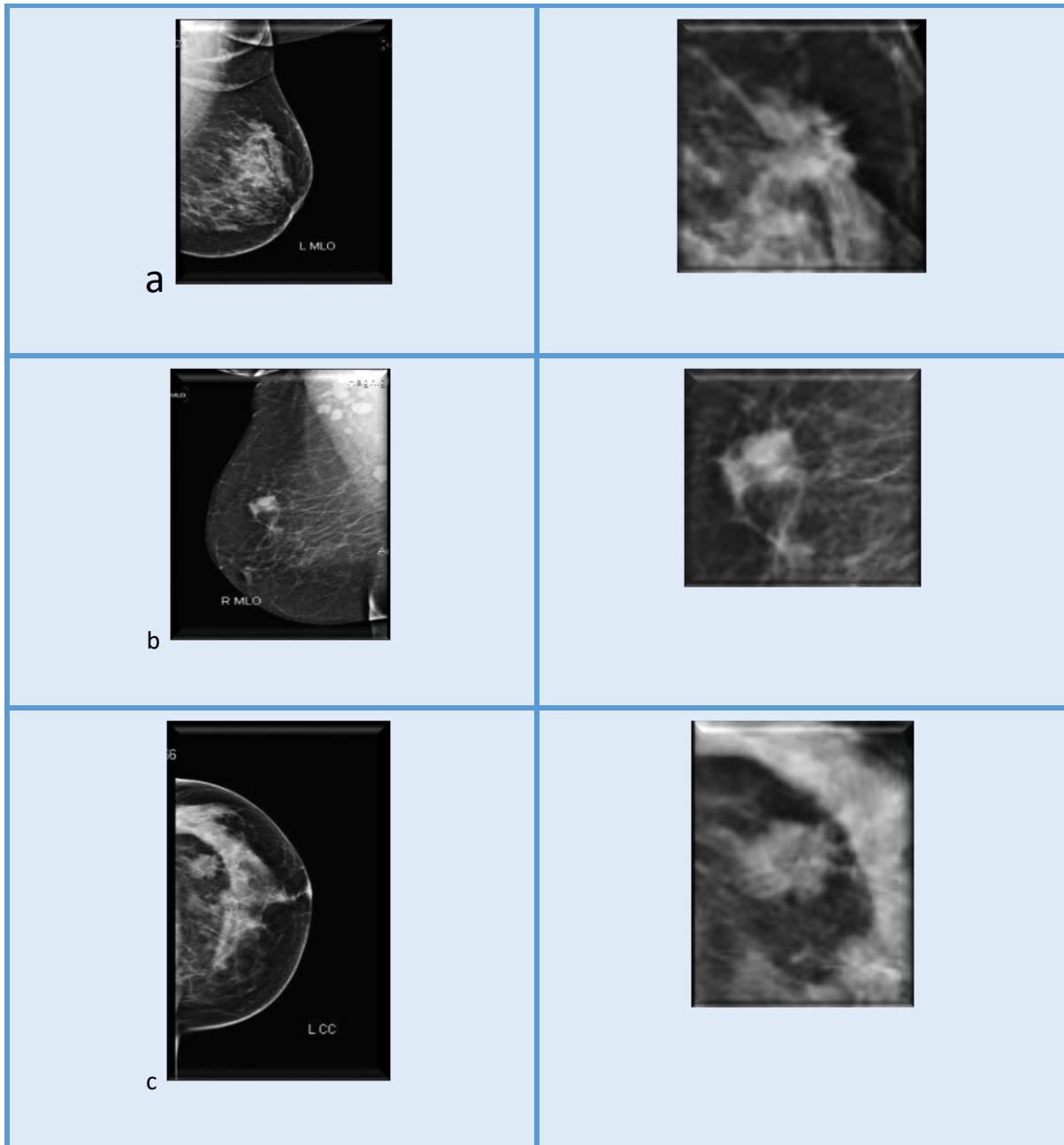


Figure18: Mass with different BIRAD scores: a) Irregular, spiky and ill define mass, b) High density and irregular, c) Lobulated and ISO dense mass, d) Well define, regular and smooth, e) Well define, regular, low density and smooth

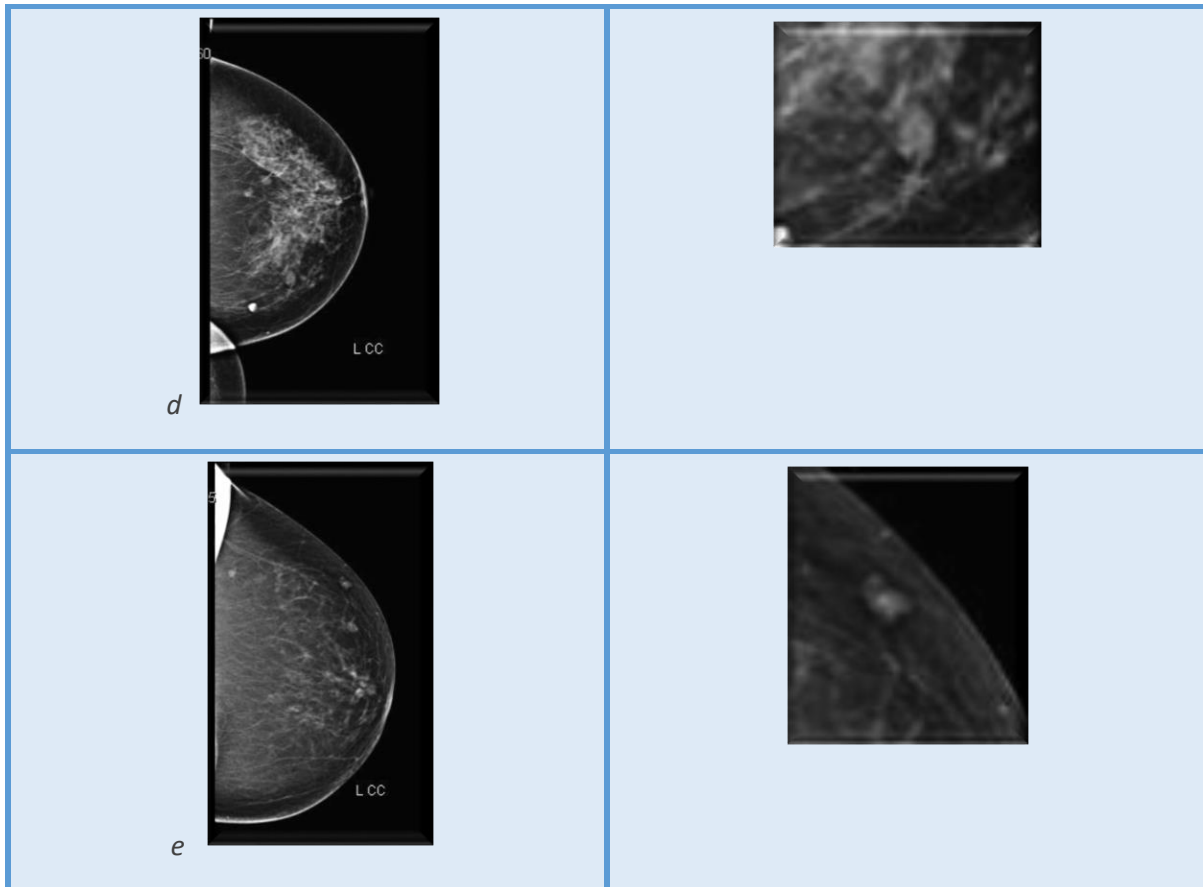


Figure18: Mass with different BIRAD scores: a) Irregular, spiky and ill define mass, b) High density and irregular, c) Lobulated and ISO dense mass, d) Well define, regular and smooth, e) Well define, regular, low density and smooth

All breast cancerous abnormalities and the way they are diagnosed were explained. As mass is the most common abnormality and has caused most cancers in women, this project will mostly focus on mass. [Metz 1998] [Arevalo2016,A.C.S 2010,Dhungel 2015,Wang2016,Elter 2009].

Next section, mass diagnosing procedure through BIRAD scoring is explained.

1.3 Abnormalities quantification using BIRADS

The radiologist's opinion regarding absence or presence of breast cancer and its type is given by a score called Breast Imaging Reporting and Data System (BIRADS) score [Orel1999]. Specialist will read the patient's Mammogram and extract clinically-relevant features (mentioned in section 1.2.6) from the detected mass. According to their location, family history, patient's medical history, age and extracted feature they will assign a BIRAD score to the mass. According to the given BIRAD (0-6), the patient will continue her treatment [Baker1996]. BIRAD Scoring system has been summarized as below [Carneiro 2015, DeSantis 2014, Altekruse 2007, Duda 2012]:

BIRAD 0	Incomplete study, there was not sufficient study or information for any diagnose, hence there is a need for more medical images.
BIRAD 1	Normal
BIRAD 2	Benign, will recommend another follow up in next 6 months or one year.
BIRAD 3	Probably benign, they will recommend another type of medical image and follow up in next 6 months or one year.
BIRAD 4	a) Low possibility of malignancy, another type of medical image and biopsy is required.
	b) Moderate possibility of malignancy, another type of medical image and biopsy is required.
	c) High possibility of malignancy, another type of medical image and biopsy is required
BIRAD 5	Highly suspicious of malignancy. Mass with BIRAD score 5 is a mass with BIRAD score 4 in mammogram that has been double checked in ultrasound and there they give it BIRAD score of 5 and biopsy is required.
BIRAD 6	Malignant mass. Diagnosed through pathology

Table 1: BIRAD scoring system

From radiologists point of view, a BIRAD score of 0 means it is an incomplete study and another modality needs to be applied to confirm a diagnosis. BIRAD score of 0 is given to a heterogeneously or extremely fatty breast (grade 3 and 4 as mentioned in section 1.1.2). In this type of breast, its highly probable that there are over lapping layers of breast tissue. So diagnosing mass will be difficult and not precise(Fig.19-a). A BIRAD score of 1 means that no mass was detected in the mammogram. A BIRAD score of 2 means that it is a benign mass(Fig.19-b). A mass that was previously checked and the density, size and other features are stable comparing to the previous checkup. A mass with BIRAD score of 3 has specifications of a mass with BIRAD score of 2(round or oval, well defined mass with smooth margin), but as they don't have any clinical or family history about it so they diagnose BIRAD 3. A round or oval mass with smooth or a little lobulated margin and less well defined comparing with mass with BIRAD score of 2, also will have BIRA score of 3. In this case mass should be checked in another modality and patient should do another follow up in next six months (Fig.19-c). A BIRAD score of four (4a,4b and 4c) means that it is suspiciously a malignant mass. BIRAD 4a is a Mass with irregular shape, lobulated or a little spikey and ill-defined margin. In this case, it is important to exactly position the mass and perform a biopsy (Fig 19-d). Mass with BIRAD score 4b means a mass with more speculated margin comparing with mass with BIRAD score 4a (Fig.19-

e). A lobulated mass with micro calcification also is diagnosed as BIRAD score of 4b. Mass with BIRAD score 4c has higher level of speculation comparing mass with BIRAD score 4b and it has micro calcification or it is close to nipple, there is also dermal or facial retraction (Fig.19-f). A mass with BIRAD score 5 means that it was check in ultrasound and there they found more speculation or more dermal and facial retraction in ultra sound or patient has family or medical history of breast cancer. A BIRAD score of 6 means specialist have the pathology report of the biopsy of mass, approving cancer and its type and level of cancer [DeSantis 2014] [Heath2000, Lincot Williams & Wilkins Publishers2010, Kopans 2007].

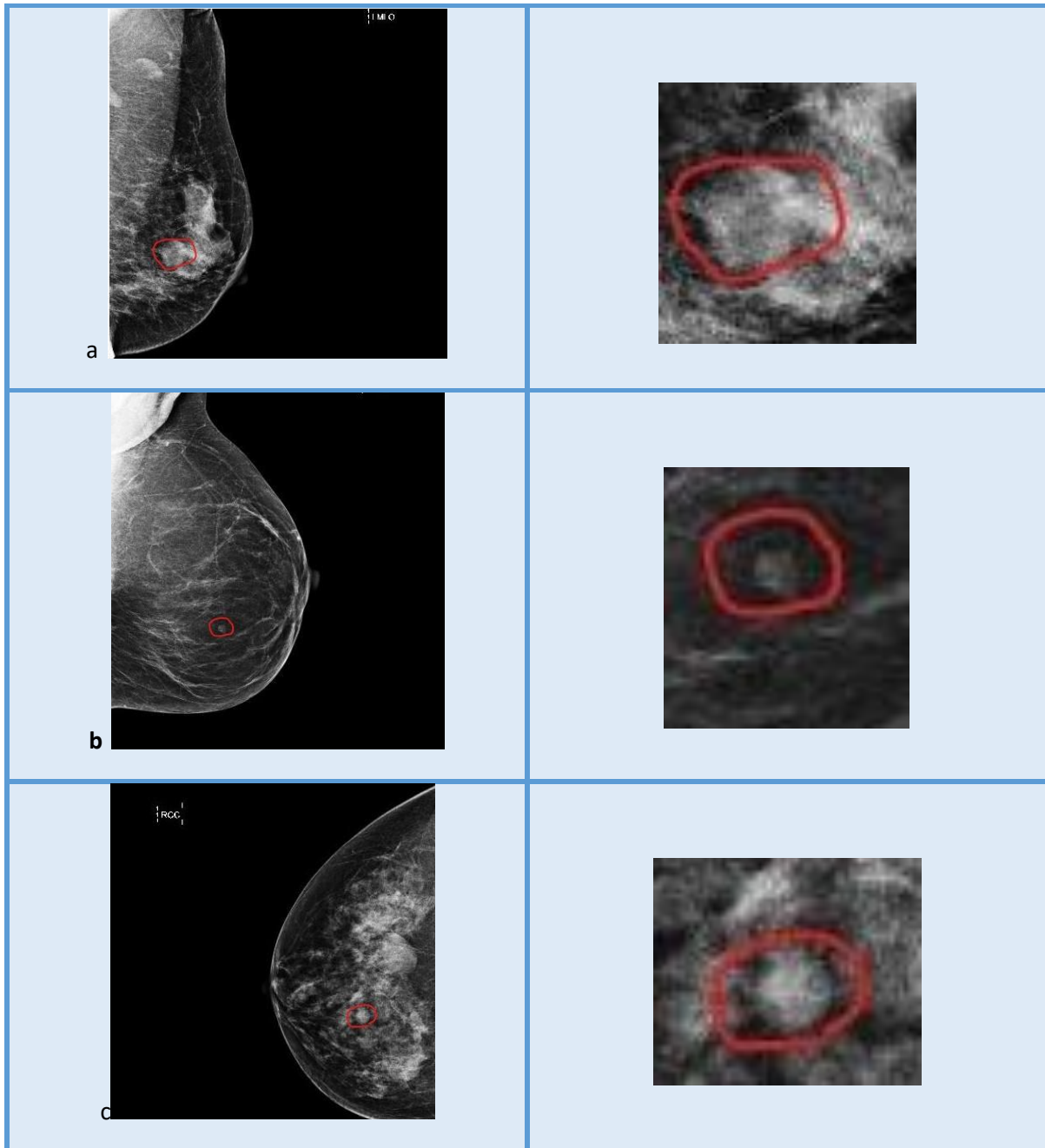


Figure 19: Mass with different BIRAD: B0) Mentioning incomplete study as it may be overlapping layer of breast tissue and it should be checked in another modality, B2) Benign mass as patient has medical history and the mass is stable comparing previous Mammography, B3) There is no previous Mammography(no medical history) to compare this mass with that one, B4a)Speculated mass, B4b)Mass with more speculation comparing to B4a, B4c)Mass with sever speculations, B5)Speculated mass that has been checked in ultrasound and in ultrasound specialist still found more speculations also there are some micro calcifications on the mass .

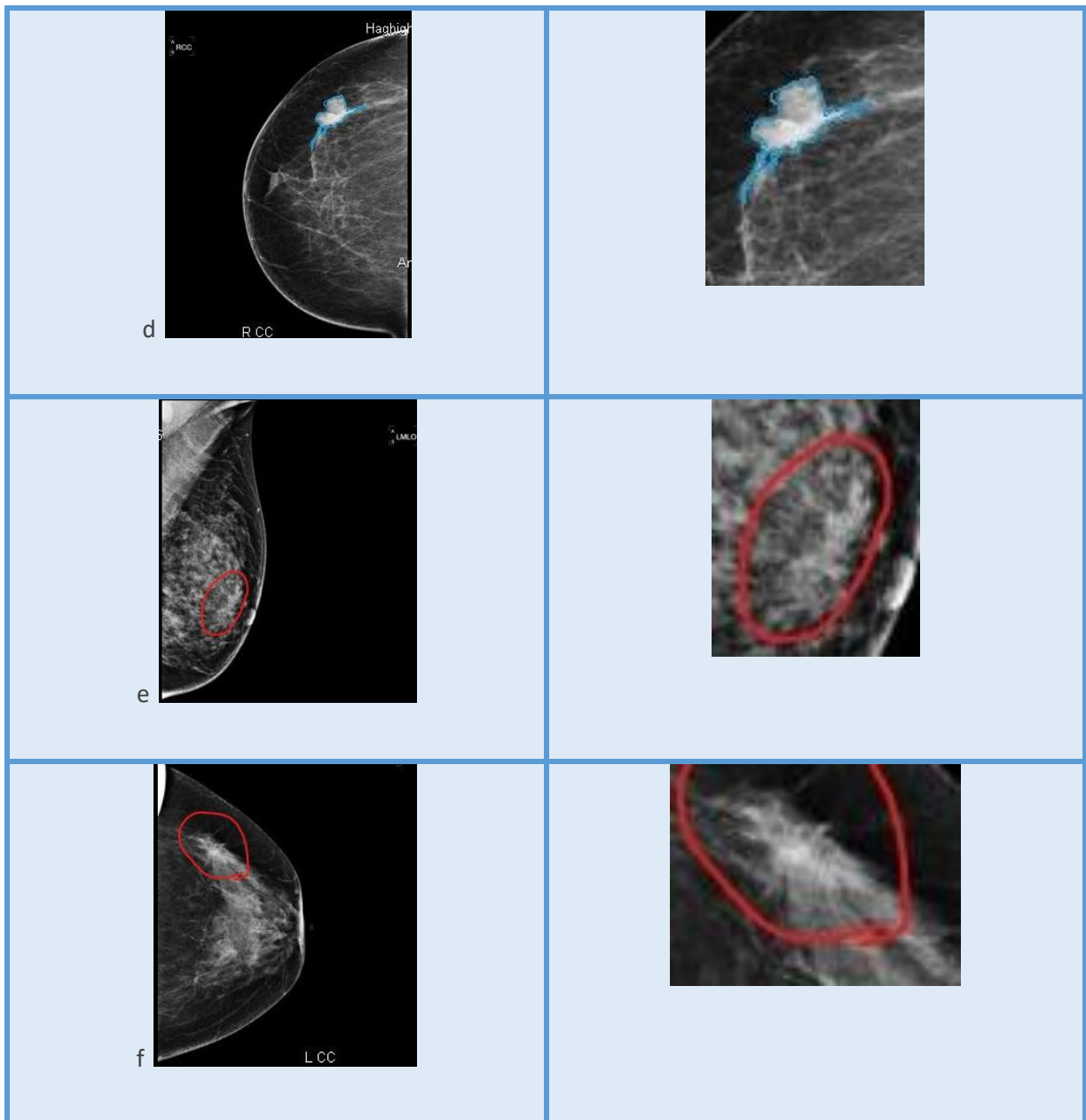


Figure 19: Mass with different BIRAD: B0) Mentioning incomplete study as it may be overlapping layer of breast tissue and it should be checked in another modality, B2) Benign mass as patient has medical history and the mass is stable comparing previous Mammography, B3) There is no previous Mammography(no medical history) to compare this mass with that one, B4a)Speculated mass, B4b)Mass with more speculation comparing to B4a, B4c)Mass with sever speculations, B5)Speculated mass that has been checked in ultrasound and in ultrasound specialist still found more speculations also there are some micro calcifications on the mass .

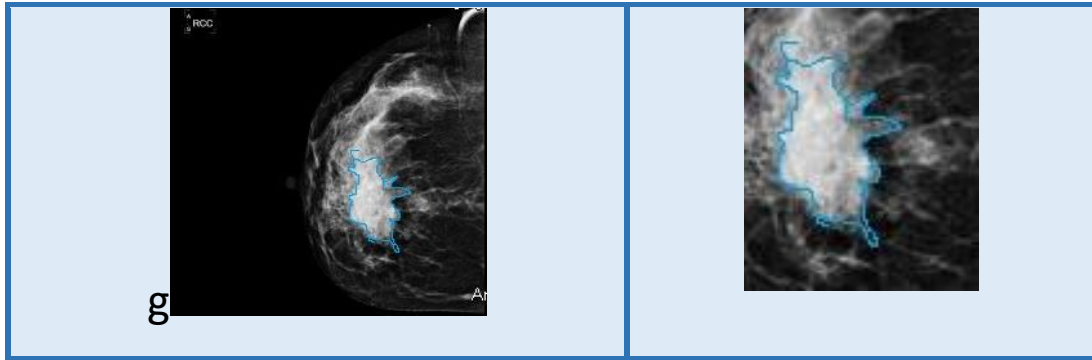


Figure 19: Mass with different BIRAD: B0) Mentioning incomplete study as it may be overlapping layer of breast tissue and it should be checked in another modality, B2) Benign mass as patient has medical history and the mass is stable comparing previous Mammography, B3) There is no previous Mammography(no medical history) to compare this mass with that one, B4a)Speculated mass, B4b)Mass with more speculation comparing to B4a, B4c)Mass with sever speculations, B5)Speculated mass that has been checked in ultrasound and in ultrasound specialist still found more speculations also there are some micro calcifications on the mass .

According point mentioned above BIRAD score 2 as benign mass and BIRAD score over 4(4a, 4b and 4c) as malignant mass are the most critical ones.

1.4 Computer Aided Diagnosis(CAD)

A CAD system refers to computer aided diagnosis and computer aided detection.

These systems are playing important role in helping specialists to improve their accuracy. Specialist has very complex and difficult task for diagnosing, as they should read the image, find the abnormality and then diagnose it. Many researches have reported how specialist read images, how they diagnose it, and why some abnormalities are sometimes missed or misdiagnosed [Birdwell 2004]. CAD are mainly used for detection of cancers [Warren Burhenne 2000] [Ball 2007]. CAD is combined of different technologies such as: image

processing, pattern recognition, artificial intelligence, medical imaging and etc. CAD has remarkably advanced since 1980 where the first CAD system has been developed at Kurt Rossman Laboratories for Radiologic Image Research in the department of Radiology at the university of Chicago and is widely then used in medical field since [Karssemeijer 2004, Freer 2001, Samulski,2010]. To accept the efficiency and usability of CAD, its performance should be equal or higher than specialist's performance. In other words, its sensitivity and specificity should be high in order to make sure no abnormality will be missed or not so many healthy tissue incorrectly detected as abnormality. Despite of the current advances in developing CAD, it only should be used as a second opinion. This means that CAD takes into account the role of specialist and computer equally. So the performance of CAD should be complementary to that by specialist [Carneiro 2015, Karella 2015, Dehmeshki 2015, Salama 2018]. Researchers studied effect of using CAD in cancer detection via mammograms [Salama 2018, Elter 2009]. [Carneiro 2015] showed that when CAD is used in Mammography, detecting of small but invasive cancer increased by 164% and 5.3 years' reduction in mean age of patient (survival rate) at the time of detection. The reason for this improvement is that specialist may miss some abnormality in first glance but when he or she use CAD as second opinion, he can correct his mistake [Elter 2009]. CADs false positives are different from

specialist false positives so when CAD have False positive specialist easily ignore those mistakes. In cases of non-expert specialist, CAD could increase their performances. As they are young resident or not enough experts, hence they may do more misses or miss diagnosis so CAD could be even more useful [Karella 2015, Dehmeshki 2015, Salama 2018]. As mentioned before specialists have difficult or heavy duty in their working hours, from reading medical image to diagnosing abnormality. CAD won't get distracted while human may, CADs won't get fatigue and they have constancy in their performance. All mentioned above could be concluded that using CAD will have positive affect in specialist performance.

1.5 Summery

Breast abnormalities and BIRAD system were explained in section 1.2 and 1.3. Among those abnormality, it was mentioned that mass is the most common cancerous abnormality [Lehman 2002, Mendez 2004, Baker 1996, Ruschin 2007, Hilleren 1991, Lewin 2001, Metz 1998, Arevalo 2016, A.C.S. 2010, Dhungel 2015, Wang 2016, Elter 2009]. It also mentioned different factors affect mass diagnosis (explained in section 1.3). Therefore, this project focus on diagnosing mass as the most complicated one.

Breast medical imaging techniques and their specifications are explained in section 1.1. Mammography is the most common technique used for high risk

patients (patient over 40 and patients that use Mammography as their follow up). Cyst, Mass and micro calcification are detected in mammograms. It's less expensive comparing MRI but it uses X-Ray radiation. Reading MRI information require more time and its more complicated comparing other techniques. Ultrasound is used frequently for diagnosing but in cases that specialist finds any probable cancerous abnormality it also should be checked in another modality for further approve. DICOMs could not be exported from ultrasound machines. A breast ultrasound examination is not considered a screening test, but an investigative technology used for taking a closer look at areas of the breast that specialist still has some doubt after doing a mammogram and clinical breast exam. So all together this project focused on diagnosing mass in mammograms. Moreover, a CAD has been developed to assist radiology to diagnose the mass by predicating BIRAD scoring. Benefits of using CAD as specialist second opinion are mentioned in section 1.4. Prior to describe our proposed CAD system, in next section the current trend on different CAD system for diagnosing mass using mammograms are reviewed.

1.6 Current CAD systems

In this section, current research in development of CAD related to diagnosis of mass using mammogram images will be explained in order to compare their

results and introduce some of their deficiencies. For this the major components of CAD which includes region of interest (ROI) segmentation, feature extraction, selection and classifier will be reviewed.

Sun et al in 2005 used mammograms data from digital data base for screening Mammography (DDSM)⁵ distributed by university of California to implement their algorithm. Normal regions were semi manually extracted and cancer regions were extracted from cancer cases with known cancer in the center of region, with region size $512 * 512$. Totally 460 regions were extracted, 164 cancer region and 296 normal regions. Cancer region consist of 53 mass, 56 speculated mass and 55 micro calcifications. 18 curvilinear features, 16 gray level co-occurrence texture features, 20 multi resolution statistical and 32 Gabor features were extracted for each $512 * 512$ region. Followed by adaptive sequential floating forward feature selection algorithm (ASFFS), simple genetic algorithm (SGA) and nontraditional genetic algorithm (CHC). The extracted features were incorporated to the supervised method called Linear discrimination analysis algorithm (LDA) for training and testing the algorithm.

⁵ The Digital Database for Screening Mammography(DDSM) is a collaborative effort between Massachuset general hospital, Sandia national laboratories and university of south Florida. The data base contains 2500 studies. Each study contains two images of each breast, some information about the patient and image information. Location of abnormality is marked. <http://www.mammoimage.org/databases/>

Considering all extracted features, which is quite large number of features, and adaptive sequential forward floating method and LDA the area under ROC curve was 96%. Also it should be considered that there are four types of breast densities. Ductal structure and density vary in each type. Even in each type of breast density, ductal structure varies in each person.

Liu et al in 2014 also used 804 mammograms from DDSM library distributed by university of California. 826 ROI were extracted, 418 benign and 408 malignant. level set based segmentation with fuzzy c- mean(FCM) initialization were used for segmenting mass. Just 40 of system's mass segmentations were checked with the ones done manually by specialist. 12 geometry and 19 texture features were extracted from each region of interest. Those features were feed into support vector machine recursive feature elimination (SVM-RFE), SVM-RFE algorithm followed by minimum redundancy and maximum relevance(MRMR) normalized mutual information feature selection(NMIFS). Selected features were incorporated into LDA+KNN and LDA+ SVM classifiers. The accuracy (with 10-fold cross validation) was for NMIFS feature selection and SVM classification, 94%.

Aswini et al in 2013 also used mammograms from DDSM distributed by university of California. Using a contour supplied by DDSM, the ROI was extracted. (50 * 50-pixel box with the mass in center of it). Only 50 malignant

mass and 120 benign mass were selected. They used grey level co-occurrence matrix (GLCM), and grey level run length matrix (GLRLM) for feature extraction and C5.0 decision tree for classification. The model is trained using several methods boosting, winnowing and pruning. Area under ROC curve is used for performance evaluation accuracy. The level of prediction was 0.995. the result was good but the number and how the data selected from database were under question.

Cascio et al in 2006 used 608 mammograms from MAGIC_5 (several hospitals belonging to the MAGIC-5 collaboration (Medical Applications on a Grid Infrastructure Connection) data base. Radiologist draw a circle around the mass area and they used edge base segmentation to segment the mass. 12 shape and texture features were extracted and neural network was used as a classifier. The system efficiency and sensitivity are 93% and 82%. Their focus was mostly on performance of segmentation. Not enough feature extraction was explored.

Eltoukhya et al in 2010 used mammographic image analysis society images (MIAS)⁶. 51 malignant, 64 benign and 207 normal breast were selected. Centre of each ROI was positioned by radiologist. The ROI was manually cropped from the mammogram. (128*128 ROI cropped) They did a comparative study

⁶ The Mammographic Image Society (MIAS) is an organization of U.K. research group. It's a data base of digital mammograms. The data base contains 322 digitized films. It includes radiologists "truth" markings on the location of any abnormality. <http://www.mammoimage.org/databases/>

between wavelet and curve let transform as feature extraction methods. Curvelet transform outperforms wavelet transform and the difference is statistically significant. Wavelet transform, two dimensional Wavelet transform and curve let transform was studied. Largest 100 co-efficient from each decomposition level was used for segmentation. Neural neighbor classifier based on Euclidian distance was used for classification. Classification accuracy for 2 * 5-fold cross validation was 94.07 for curve let and 90.05% for wavelet transform.

Micro classification, lobulated mass, speculated mass, ill define mass, architectural distortion and asymmetry were considered as cancerous abnormalities. While lobulated, ill define and spikey are specification of one cancerous abnormality; mass. Asymmetry can be diagnosed by considering four image of mammogram in its four direction, it could not be diagnosed just considering one direction of mammogram. Micro classification, mass and architectural distortion are quite different in texture.

Salabat Khan et al in 2017 collect 109 data from mammographic image analysis society images (MIAS), 20 malignant, 54 normal and 35 benign. Centre of each ROI was positioned by radiologist. The ROI was manually cropped from the mammogram. (128*128 ROI cropped). They used four type of feature extraction: magnetic Gabor response-LDA(MGR-LDA), MGR-PCA, statistic

MGR(SMGR), window based first order SMGR(WSMGR). MGR feature extraction has huge feature dimension so the LDA and PCR were used for dimension reduction. The selected feature was incorporated into a weighted SVM and successive enhancement learning based weighted SVM (SELwSVM) classifier. Ten-fold cross validation was used to evaluate the proposed method. For comparing all results Friedman statistic test was used. Best performance of feature extraction strategies and all datasets, using tenfold cross validation based on sensitivity and specificity, was 97.22% and 98.5% for LDA-WSMGR. As specificity SMGR has good performance but as mentioned, it's feature dimension is huge. Also accuracy is dependent on the size of windows chosen for feature extraction. In smaller windows the accuracy is higher but computation is time consuming.

Elmanna et al in 2015 also used mammograms from digital data base for screening Mammography (DDSM) distributed by university of California. Centre of each ROI was positioned by radiologist. The ROI was manually cropped (box).59 texture and statistical features were extracted. SFFA and SFS were used as feature selection while KNN, SVM, LDA, Quadric discriminant analysis(QDA)were used as classifiers. Performance of their method using leave one out cross validation based on sensitivity and specificity for SFS+KNN and

SFFS+KNN was 96% and 98% respectively. SFS and SFFS are both approximately the same methods.

Adi Nugroho et al in 2014 collect data from private data base of oncology clinic kota baru Yogyakarta. It consists of 14 benign, 6 malignant and 20 normal mammograms. They used 7 GLCM and 5 co-occurrence matrix features followed by Correlation based feature selection in that 6 features were selected. Despite of shortage of data (only 40 mass), sensitivity and specificity were 93.75% and 96.7% respectively while no indication of how segmentation of mass carried out.

Virmani et al in 2016 used MIAS data base. They used 322 mammograms. ROI has been selected from densest region of each mammogram (ignoring pectoral muscles). They extract different texture features with different length of law's texture energy masks from ROI. They used PCA for dimension reduction and from which top 4 features were chosen. They used SVM and probabilistic NN classifiers and their best accuracy was 94.45 AND 92.5% respectively.

Muramatsu et al in 2016 used 376 mammograms from Nagoya medical centre, 195 benign mass and 181 malignant mass. They used Radial Local Ternary Pattern (RLTP) as an ROI based feature technique. That technique present orientation of edge pattern from the center of mass. These features were compared with rotation invariant uniform, ordinary local temporary pattern (LTP), wavelet features and texture features (GLCM). For classification

they used artificial neural network(ANN), SVM and random forest. Combination of RLTP and ANN had higher accuracy in terms of area under the operating characteristic curves,90%.

Li et al in 2017 used digital data base of screening Mammography(DDSM) .383 mammogram were gathered, 143 benign mass and 180 malignant mass. Centre of each ROI was positioned by radiologist. The ROI was manually cropped from the mammogram. (128*128 ROI cropped) They extracted local contour features. They converted 2D contour of mass in to 1D vector of features. They segmented the 1D signature into four different subsections. They extracted roughness of contour by Mean square slope (RMS), beside fractional dimensional, mean and STD ratio features. Their highest classification accuracy was 99.6%.

Mohammed et al in 2018 proposed a ROI based convolutional neural network called YOLO. It can handle detection and classification in one framework consists of four classic stages: preprocessing, feature extraction employing convolutional NN, mass detection and classification with fully connected NN. They used 600 mammograms from DDSM. The CAD detect the mass and classify it as malignant or benign. The CAD accuracy with 5-fold cross validation for mass detection was 99.7% and 97% for mass classification.

Wei et al in 1995 used multi resolution texture analysis to differentiate mass from breast tissue. They used wavelet transform to decompose ROI. Extracted

texture features from wavelet coefficient LDA were used for classification purpose. The method was tested with 168 malignant and 504 benign masses from DDSM. Their best accuracy was 89% and 86% for training and test groups.

Liado et al in 2009 used a texture descriptor, spatially enhanced local binary pattern (LBP), for extracting texture features. They used 256 normal and 256 malignant masses from DDSM. LBP descriptor makes statistics on local micro pattern such as dark or bright spots, flat areas and etc. It's not robust against noise. LBP features were classified by SVM classifier. The area under the ROC curve was 94%.

Zhang et al in 2009 used Gabor filters to create Gabor images. They were used to extract a set of edge histogram descriptors. A fuzzy c-mean and KNN were used as classifier. 159 normal and 272 malignant mass from DDSM were used to implement their method. The true positive reported as 90% with 1.2% false positive. This method used edge histogram which is holistic descriptor and doesn't present mass's local texture feature.

Lahmiri et al in 2011 used discrete wavelet transform (DWT) and Gabor filters for their proposed CAD. They applied Gabor filter with different frequencies and orientation on high frequency sub images derived from DWT. This followed by extracting statistical features from Gabor images. SVM were then used for classification. The method was tested with 189 mammograms from DDSM with

tenfold cross validation. The best accuracy was 98%. They didn't mention their segmentation method.

Costa et al in 2011 used Gabor wavelet, Principal component and efficient coding based on independent analysis (ICA) for feature extraction and LDA for classification. They used 5090 ROI from DDSM data base (they didn't mention ratio between benign and malignant mass).

Their best accuracy was using ICA extracted features and LDA classifier, 90.07% sensitivity and 93.83% specificity and 85.8% specificity. They didn't mention anything about segmentation.

Duda et al in 2012 used Moran's index and Geary's coefficient as features. SVM were also proposed as classifier. Accuracy of CAD were evaluated using both group of features. Moran index accuracy reported as 89% and Geary's accuracy reported as 87.7% using 1394 masses from DDSM, they didn't mention ratio between benign and malignant masses. Also segmentation technique was not mentioned.

Loan et al in 2011 raw magnitude response from 2 D Gabor wavelet were used as features. Similar to others, PCA were used as a dimension reduction process followed by SVM as classifier. 322 mass from MIAS were used, consisting of 208

normal breast, 63 benign and 51 malignant mass. The best performance was for Gabor wavelet+ PCA was 97.56% sensitivity and 60.86% specificity.

Salabat khan et al in 2017 Considered ROI center and size and then they make a box around the mass. For segmenting the ROI. They used bank of Gabor filter to extract the directional features related to mass. These features represent structural specification of mass and normal tissue in different orientation and frequencies. Their focus was on micro classification and mass detection. They used successive enhancement learning based weighted support vector machine. 109 data were gathered from MIAS data base, 20 malignant mass, 35 benign mass and 54 normal but suspicious tissue. Their average accuracy was from 68% to 100% based on applying different techniques.

Mohammed et al in 2010 used MIAS data, 42 malignant mass, 174 benign mass and 106 normal mammograms. Considering ROI center and size they make a box around the mass. They present a CAD using ANN for mass classification. They mostly focus on texture features for characterizing mass and other features such as mean, STD entropy, skewness, kurtosis and uniformity. Their effort was on reducing false positive results. They used those features with layer artificial NN and 90.1% were reported for sensitivity while 83.87 were reported for specificity.

Daniel Levy et al in 2017 used combination of transfer learning, preprocessing and data augmentation followed by convolutional NN. DDSM also was used at this paper, 1820 masses (they didn't mention ratio between benign and malignant masses). Their accuracy reported as 96.5% while only mathematical features were considered.

Surendiran et al in 2011 used 300 data from DDSM database. They used 17 shape based features including size and the direction of ROI. Analysis of Variance ANOVA Discriminant Analysis (ANOVA DA) was then used as a feature selection process. It provides Wilk's Lambda statistics for each feature and compute its level of importance based on grouping class variables. PCA was also used for feature extraction. The accuracy with stepwise ANOVA DA was reported as 87.3% with 5-fold cross validation and 82% using PCA features.

Markey et al in 2002, 2003 used 2258 mass from DDSM. They used statistical and shape based features for characterizing a mass. Statistical features derived from frequency histogram or gray value of pixels within a mass. As noise and over enhancement may influence value in histogram, the method is highly sensitive to noise which may lead to less accurate result. They used feed forward back propagation artificial neural network (BP-ANN) for mass classification. They had 25% specificity and 98% sensitivity.

Bovis et al in 2000 used 144 mass from MIAS. They also used texture features such as: angle of second moment, correlation, difference entropy, sum average, inverse, sum entropy and etc. Followed by PCA. But similar to previous works noise has strong influence on these features and they couldn't be considered reliable. ANN were used as a classifier. Their best accuracy was 77%.

Markey et al in 2003 used shape based encoded descriptor from DDSM data base. They considered 4 groups for masses, based on their shape (round, oval, lobulated and irregular). They used 25 mass from each group. They used KNN for mass classification. They didn't mention anything about segmentation technique. They obtained a classification accuracy of 78% for classifying masses as oval or round and an accuracy of 72% for classifying masses as lobulated or irregular.

Zaiane et al in 2002 used 322 mammograms from MIAS. 208 normal mammograms, 63 benign mass and 51 malignant mass. The center of mass and its size was considered to draw a box around the mass. They used statistical features for classification of mass. They used association rule based classification by category (ARC-BC). Their method's accuracy was 80.33% while noise could have strong influence on their results.

Vadive et al in 2013 collect 224 mammograms from DDSM, 117 benign and 109 malignant mass. For segmentation they used threshold, considering center of mass and its size which was mentioned in DDSM. They extract 17 shape based features from each mass They could classify round, oval, lobulated and irregular mass. The mass classifier showed certain amount of impreciseness using fuzzy membership function and some set of rules. The Fuzzy function and rules were based on C5.0 algorithm. Their classification accuracy was 97.22% Accuracy of segmentation which is important for shape based features, was not mentioned.

Bovis et al in 2000 extract 280 texture features from Spatial Gray Level Dependency(SGLD). They used 40 malignant samples and 104 normal mammograms from DDSM Considering center of mass and its size which was mentioned in DDSM they cropped ROI from mammogram. NN classification sensitivity of the method was reported as 77%.

Eltoukhy et al in 2010, gathered 142 mammograms from MIAS data centre. 51 malignant, 64 benign and 207 normal mammograms. ROI 128*128 is cropped by considering center and size of abnormality. Comparative study between wavelets and Curvelet co-efficient features was done. Neural network with Euclidian distance metric was used for classification. Best accuracy was for curvelet transform and NN, 94.07%

Casio et al in 2006 used 3762 digital images acquired in several hospitals belonging to the MAGIC-5 collaboration (Medical Applications on a Grid Infrastructure Connection). Segmentation was done by means of a ROI Hunter algorithm. Geometrical and shape based features were extracted. They used ANN classification. The area under ROC curve was 86.2%

Rosin et al in 2009 used 144 mass from MIAS data set. Centre of mass and size of it was used in order to consider a box around the mass. They used shape based features and convexity based measures to classify mass. Convexity measures are classified by KNN classifier and their accuracy was 74.1%.

Another technique for object detection and classification is deep learning. Other machine learning methods need feature extraction but deep learning method adaptively learn the appropriate feature extraction process from the input data, considering the target output. The most common deep learning method is convolutional neural network(CNN).

Kim et al in 2018 introduced a system which was trained with malignant and benign cases. They gathered 29107 mammograms from five institutes. They assessed the feasibility of a data-driven imaging biomarker based on weakly supervised learning (DIB; an imaging biomarker derived from large-scale medical image data with deep learning technology) in mammography (DIB-MG). For

each case in a training set, the cancer probability inferred from DIB-MG is compared with the per-case ground-truth label. Then the model parameters in DIB-MG were updated based on the error between the prediction and the ground-truth. Their system accuracy was 90.2%

Arevalo et al in 2016 tested various CNNs and compared them with two hand-crafted descriptors mass diagnosis procedure. They used BCDR-FM dataset. They reported performance improvement with the combination of both learned and hand-crafted representations. They didn't test the performance of pre-trained networks and used simpler CNN architectures. Their system best accuracy was 82.6%.

Huynh et al in 2016 used pre-trained system. They analyzed the performance of classification using features from various intermediate layers of the network using SVM for the classification. They compared their results to two approaches: a classifier operating on hand-crafted features and an ensemble of both, using soft voting. Best achieved average accuracy was 86%.

Jiao et al in 2016 proposed a system in which a pre-trained CNN was fine-tuned on a subset of DDSM database. Then, features of masses were extracted from different layers of this model. In this way 'high-level' and 'middle-level' features were obtained. Two linear SVM classifiers are trained for the decision

procedure, one for each feature group, and their predictions are fused. Their system best accuracy was 86.5%.

Ting et al in 2019 created and trained their network for breast mass classification. Their network comprises 28 convolutional and fully-connected layers and it is fed by proposal ROIs detected by a one-shot detector. They used MIAS database. Their system best accuracy was 90.7%.

Summary of reviewed CADs is presented in table 2.

Author/year	Segmentation method	Feature extraction	Classification method	Accuracy
Sun/2005	box	Texture feature	LDA	96%
Liu/2014	Level set and fuzzy c-mean	Geometry and Texture feature	LDA+KNN	94%
Aswini/2013	box	Texture feature	Decision tree	99.5%
Cascio/2006	Edge based	Texture feature	NN	93%
Eltoukhya/2010	box	Wavelet transform	NN	94.07%
Salabtkhan/2017	box	SMGR and WSMGR	SELWSVM	98.5%
Elmana/2015	box	Texture feature	QDA	98%
Adinugroho/2014	N/A	GLCM	SVM	96.75%
Virmani/2016	N/A	Texture feature	SVM	94.45%

Table 2: Presents summary of all reviewed CAD systems

Muramatsu/2016	N/A	GLCM an Radial local ternary pattern	ANN	90%
Li/2017	box	Extract roughness of contour by mean square slope	SVM	99.6%
Wei/1995	Wavelet transform	Texture feature	LDA	89%
Lidao/2009	N/A	Texture feature	SVM	94%
Zhang/2009	N/A	Texture feature	Fuzzy C-Mean	90%
Lahmiri/2011	N/A	Wavelet transform and Gabor filter	SVM	98%
Costa/2011	N/A	Independent coding analysis	LDA	93.83%
Duda/2012	N/A	Geary Coefficient	SVM	89%
Loan/2011	N/A	2D Gabor filter	SVM	97.56%
Salabatkhan/2017	box	2D Gabor filter	SVM	68%
Mohammed/2010	box	Texture and shape feature	NN	90.1%
Daniellevy/2017	N/A	Texture feature	NN	96.5%
Surendiran/2011	N/A	Shape features	ANOVA DA	87.3%
Markey/2002	N/A	Shape features and statistical features	BP-ANN	98%
Bovis/2000	N/A	Texture feature	ANN	77%
Markey/2003	N/A	Shape features	KNN	78%
Ziane/2002	box	Statistical feature	ARC-BC	80.33%

Table 2: Presents summary of all reviewed CAD systems

Vadive/2013	Threshold	Shape feature	Fuzzy C-mean	97.22%
Bovic/2000	Box	Texture feature	NN	77%
Eltoukhy/2010	Box	Wavelet and Curve let feature	NN	94.07%
Casio/2006	Mean of ROI Hunter algorithm	Shape and geometric feature	ANN	86.2%
Rosin/2009	Box	Shape and Convexity feature	KNN	74.1%
Rim/2018	Deep learning		ANN	90.2%
Avelao/2016	Deep learning		ANN	82.6%
Huynh/2016	Deep learning		SVM	86%
Jiao/2016	Deep learning		SVM	86.5%
Ting/2016	Deep learning		SVM	90.7%

Table 2: presents summary of all reviewed CAD systems

As presented in table 2, some authors extracted just texture features for diagnosing a mass and they didn't mention any thing about noise removal and pre-processing, despite the fact that noise can affect classification results (Hanmandlu in 2004).

Automatic segmentation of ROI and its accuracy is important in diagnosing system. Most of reviewed CADs didn't mention any thing about detecting margin/border of mass and thus no report was given about the accuracy of segmentation. Shape of ROI's margin is important in diagnosing system. Despite such important fact, Many researcher used a box for approximate segmentation

of ROI without applying a segmentation method to segment the exact ROI. In addition, the size of the box can affect the result and complexity of computation.

A CAD diagnosing system beside accurate segmentation needs feature extraction, feature selection and classification.

For complete analysis of ROI, different types of features such as shape based, intensity based and texture based features were examined in this project. Most of authors just used either shape based features, texture features or intensity based features and none of the reviewed papers, used all types of features together to fully characterize ROI.

In addition to segmentation and feature extraction components of CAD, classifications algorithms also play important role in achieving a good result in terms of sensitivity and specificity. Despite this fact, most of the current CADs had either lower accuracy comparing to our proposed CAD, or the feature dimension was huge which add extra overhead. Furthermore, accuracy of such works depends on the size of windows chosen for feature extraction. Last and most important one, none of the reviewed CADs assigned a BIRAD to a mass where it was ultimate goal of our proposed CAD.

1.7 Aim and Objectives

Aim of this research was to develop a fully automatic CAD system to predict and assign a BIRAD scoring to a given mass.

To validate such proposed CAD system, 160 masses were collected from “Haghighat imaging center” and “Farokhi imaging center”.

As mentioned in section 1.4, BIRAD score of 2 and over 4 are the most critical ones so, 80 masses with BIRAD score 2, 30 with BIRAD score 4a, 30 with BIRAD score 4b and 20 with BIRAD score 4c.

To achieve such goal, the project has the following objectives.

1. Developing image enhancement techniques to reduce the effect of noise, on extracted texture or intensity based features. To develop an image segmentation method to automatically segment the region of interest(ROI).
2. To develop features extraction techniques to fully characterize ROI. All medical and mathematical features (ROI shape, margin shape, ROI intensity, margin intensity, ROI texture) beside some meta data should be extracted.
3. To develop a feature selection scheme to reduce the dimensionality of the extracted feature.

4. Developing a classifier to classify mass with BIRAD score 2 and over 4.
5. Developing a classifier to sub classify mass with BIRAD score 4 to three groups 4a, 4b and 4c.
6. Developing a validation scheme to evaluated accuracy of automatic segmentation of ROI as well as the performance of CAD in terms of its sensitivity and specificity

CHAPTER 2

2. Image enhancement

2.1 Methodology

The aim of this project was to automatically analyze and diagnose a detected mass in a mammogram and to assign a BIRAD to the given mass. Automated analysis of the mammographic images was divided into three main steps: (1) automatic segmentation of each mass, (2) feature extraction, feature selection and finally (3) classification. Prior to any analysis, in order to reduce effect of noise on classification result, pre-processing should be done. Diagram 1, present all these steps. In this chapter, first a short review of each step of the proposed methodology is explained and after that image processing (noise- removal, segmentation and feature extraction) will be fully explained. In next chapter pattern recognition (feature selection and classification) and meta data extraction will be explained in details.

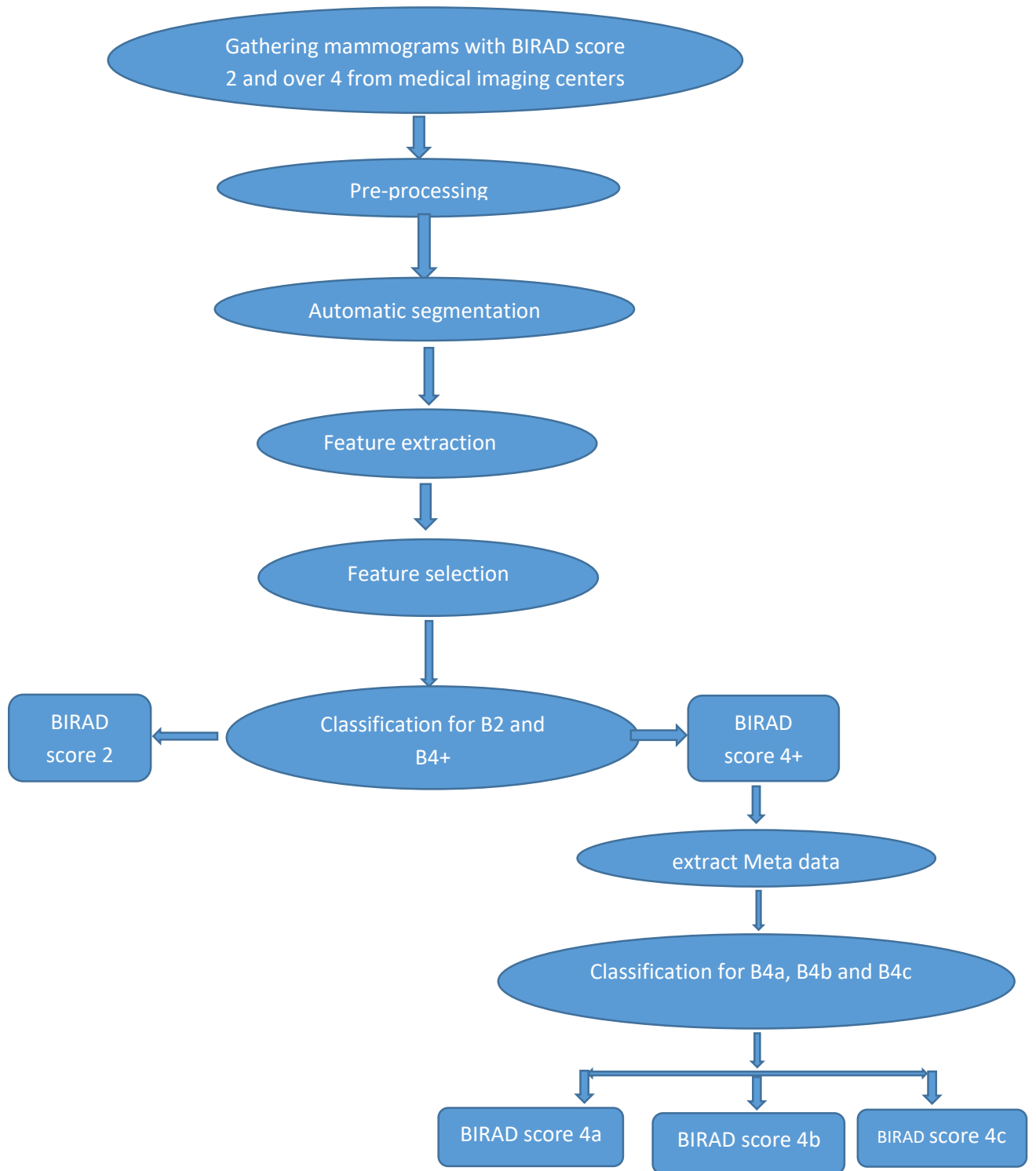


Diagram 1: Steps for automatic diagnosis of detected mass in mammogram (CAD).

- **Patient information, data and image acquisition**

173 digital mammograms and their reports were gathered from “Haghighat medical imaging center” and “Farokhi medical imaging center” from Theran, Iran under supervision of Dr. Parto Hasanizadeh and Dr. Neda Nasser, respectively.

All figures presented in this thesis are the data that I have collected from these medical imaging centers, except the ones that is mentioned under the images. I have the ethical approve for using these data from both medical imaging centers.

As mentioned in section 1.3, BIRAD score of 2 is for benign mass and BIRAD score of over 4 is for malignant mass. From oncologist point of view, differentiating between BIRAD score of 4a, 4b and 4c is critical as their treatment is different. So from those mammograms, 80 of them have BIRAD score 2, 30 with BIRAD score 4a, 30 with BIRAD score 4b and 20 of them have BIRAD score 4c. Also 13 mammograms with BIRAD score 3 were gathered in order to test the system with those mammograms.

According the information gathered from patient’s reports it could be concluded that mass with BIRAD score 2 belongs to patients within the age range of 35-55years old and masses with BIRAD score over 4 belongs to patients with age range of 45-65 years old.80% of mass with BIRAD

score 4c has retraction.35% of them has micro calcifications (lobulated mass with micro calcification inside.

- **Pre-processing (image enhancement)**

Some of probable problems in mammogram could be the presentation of noise, poor image contrast and weak boundaries. New mammography devices have high resolution and contrast. The mammograms that were collected from “Haghighat medical imaging center” and “Farokhi medical imaging center” didn’t have problem in image contrast and boundaries, as data has been captured using advance image devices. Technical specification is presented in section 1.1.2.

Hanmandlu et al in [Hanmandlu 2004] demonstrated that noise on mammograms could affect texture features values and consequently classification results. The mammograms that are used in our research are not very noisy. But in order to have a generic CAD system and check the effect of noise on final results of CAD, some of most frequent noises on mammogram are employed on small amount of our mammograms (50 mammograms). Then the CAD was evaluated with those noisy mammograms as inputs. Accuracy of CAD is assessed and presented in section 4.2.1.

Most frequent noise and de-noising filters for mammograms are presented in detail in section 2.1.

- **Automatic segmentation of ROI (mass)**

After preprocessing steps, masses were annotated in Haghghat medical imaging center and Farokhi medical imaging center by 3 senior specialists to highlight the location and the border of the mass. They highlighted the location of mass by a pen on a printed Image. All of 160+13 masses were automatically segmented by a proposed active contour segmentation method. Most common Mammography segmentation techniques and the proposed segmentation are described in section 2.2 in order to compare them together and find the deficiency and efficiency of each technique. Results of segmentation method as the selected one is presented in section 4.2.2.

- **Feature extraction**

Once the ROIs were segmented, it's time to characterize each ROI. To fully characterize each ROI, specialist's visual aspect beside mathematical features for each mass should be considered. Specialist's Visual characteristics of a mass are mentioned in section 1.2.6. To reflect visual characteristics used by specialist for diagnosing mass, corresponding features should be extracted. Beside medical/clinical features

mathematical features also were extracted to fully characterize a mass for the next step. Several feature extraction methods are presented in section 2.3. Results of extracted features, which are presented in section 4.2.5, were manually compared and some irrelevant ones are omitted to reduce the dimensionality of the feature space. Finally, the extracted features, which are presented in section 4.2.5, are send to feature section algorithms for further reduction of feature space dimension.

- **Feature selection**

Prior of applying any classification techniques, for removing redundant and irrelevant ones some feature selection and dimension reduction techniques were applied. Feature selection methods are either filter, Wrapper or embedded method. It was tried to employ one algorithm from each method to find out the best algorithm for increasing the performance of the proposed CAD. Feature selection methods that were used for the proposed CAD are described in section 3.1 and Feature selection results, feature subsets, are presented in section 4.2.6. Selected features and their weights were send to classification techniques in order to classify each mass to different BIRADS.

Classification of BIRAD score 2 and over 4

For having more accurate classifier, number of data in different groups (B2, B4a, B4b and B4c) should be approximately the same. So it was decided to classify in two levels. First classify B2 and B4+ and in next level classify B4a, b4b and B4c.

The first step of classification of BIRAD is to classify mass with BIRAD score 2 from mass with BIRAD score over 4(4+).

BIRAD scoring system and how being assigned to a giving mass are presented in section 1.3.

It is worthy to mention that BIRAD 0 means that the mammogram is not reliable. BIRAD 1 means there was no abnormality detected and BIRAD 2 means it's a benign mass. BIRAD 3 means it's probably benign mass but needs another follow up in next six months to confirm it. BIRAD 4(4a,4b and 4c) means its suspicious of malignancy. BIRAD 5 means its highly suspicious of malignancy and BIRAD 6 means malignancy is approved with biopsy result.

From points mentioned above it was proposed to first distinguish mass with BIRAD score 2 from mass with BIRAD score over 4 as benign and malignant mass. For this, weighted Features chosen by different feature selection schemes presented in 3.1 were fed to some classification

methods in order to classify a mass into BIRAD score 2 and over 4. For this, the most frequent classification methods are described in section 2.5. Accuracy of combinations of feature selection and classification methods were accessed and compared. The results are presented and explained in section 4.2.7. The system was able to diagnose mass with BIRAD score 2 and over 4.

- **Sub classification of BIRAD 4 and use of Meta data**

Up to this point, the proposed CAD classified mass with BIRAD score 2 over 4. It was aimed also to sub classify mass with BIRAD score over 4 to three groups (4a, 4b and 4c). Mass with BIRAD score over 4 (4a, 4b and 4c) is described as below:

4a: low possibility of malignancy, another type of medical image and biopsy is required.

4b: moderate possibility of malignancy, another type of medical image and biopsy is required.

4c: high possibility of malignancy, another type of medical image and biopsy is required.

From points mentioned in section 1.3, it could be concluded that its difficult and time consuming to diagnose and differentiate mass with

BIRAD score 4a, 4b and 4c as there are lots of facts that may affect diagnosing procedure and lots of times it leads to miss diagnosis. An example of those facts could be: length of speculation and existence of retractions. Young specialists or interns, may not differentiate between retraction and spikes. If specialists diagnose spikes as retraction, they will give higher BIRAD score to the mass and if they don't detect retractions they will diagnose lower BIRAD score for the mass. So retraction and specular index beside existence of micro calcification are meta data that have high influence on diagnosing mass with BIRAD score over 4.

Patient with BIRAD score over 4 after doing medical screening will go to oncologist. Oncology treatment vary for each BIRAD score. For patient with BIRAD score 4a, oncology may just evacuate the mass, for mass with BIRAD score 4b specialists may evacuate the mass and continue treatment with chemio therapy or radio therapy and for mass with BIRAD score 4c they will evacuate that breast and continue the treatment with chemio therapy or radio therapy.

So it's highly important to differentiate masses with BIRAD score of 4a,4b and 4c. So, some meta data such as micro calcification, Specularity index (shape of margin) were extracted automatically. Retraction was extracted manually. The technique for extracting those meta data are fully

explained in section 4.2.8. These meta data are extracted as features and used in different classification methods to classifying mass with BIRAD score 4a, 4b and 4c.

- **Classification for BIRAD score 4a, 4b and 4c**

Meta data considered as features were extracted from 80 masses with BIRAD score over 4. From which, 30 masses are with BIRAD score 4a, 30 with BIRAD score 4b and 20 with BIRAD score 4c. Feature values were sent to different classification methods with different parameters. The accuracy of different classifications were considered and the values are presented in section 4.2.9. The system can diagnose mass with BIRAD score 4a, 4b and 4c.

In this section a short review of all steps undertaken for the CAD is presented. Remaining part of this chapter introduce several image preprocessing steps for image enhancement as well as image segmentation methods to segment a given mass. Next chapter pattern recognition steps will be fully explained.

2.2 Image enhancement

Some of the problems with medical images are: noises, weak boundaries, poor image contrast and unrelated parts. These problems could be solved in pre-processing stage. Pre-processing is a fundamental step as it could improve image

segmentation and feature extraction steps. In pre-processing, some filters will be used to remove noise, preserve edge, smooth image and do image contrast adjustment [Bochud 1999,Burgess 1999]. As mentioned in methodology, mammogram used in this project don't have noise but for evaluating effect of noise on texture and intensity based features and consequently classification results, some of most frequent noise on mammograms and their de-noising filters are applied on the mammograms, and are described in next section.

2.2.1 Noise removal

In breast mammograms the most noises that accurse on images are: salt &pepper, Gaussian, Speckle. In (Fig. 20) the original image and images with these three types of noises are presented [Bochud 1999, Burgess 1999].

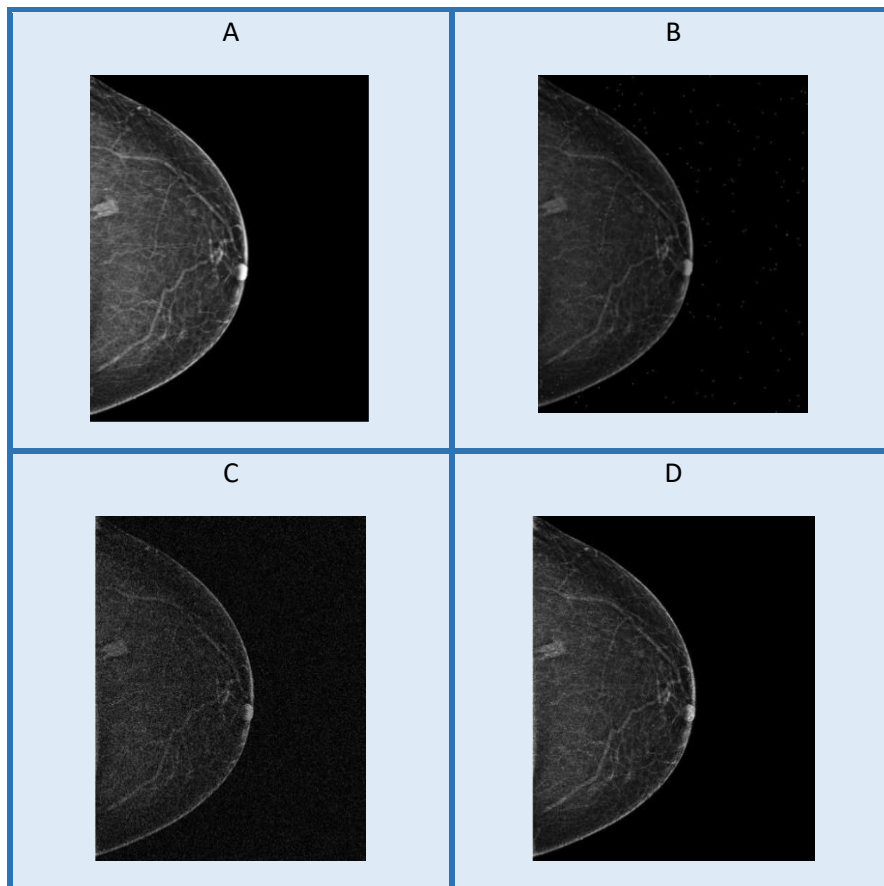


Figure 20: a-Original image, b- Salt &pepper noise, c-Gaussian noise, d- Speckle noise

As its presented in (Fig.20-b), salt & pepper noise makes white spots on the whole image. While classifying mass with BIRAD score over 4, these white spots could be miss diagnosed as micro-calcification in the mass. As its explained in section 3.3 mass with micro calcification will have higher BIRAD score. So it's essential to remove any possible sat & pepper noise from mammogram.

(Fig.20-c) presents the effect of Gaussian noise on mammogram, it brings more white spots on the image but the spots have lower density comparing to salt &pepper spots. But as it could be seen, it weakens the boundaries and lowers

image contrast, which may affect classification results. So it's essential to remove any possible Gaussian noise from the mammogram.

(Fig.20-d) presents the effect of Speckle noise on mammogram. This one same as Gaussian noise, add some white spots on the image. But the spots are not on the whole image as it was on Gaussian noise (it will not appear on the black area). It could weaken boundaries and lower image contrast hence lowering the accuracy of classification.

In this project for removing these noises some frequent filters were employed such as: median filter, adaptive median filter, Wiener filter which are explained below.

-Median filter:

Median filter is a nonlinear filter that removes noises whilst keeping the sharpness of edge in the image [Rafferty 2003, Makandar 2015] and that's the reason it's widely used. Median filter beside removing noise, it's also kind of smoothing technique that removes noise in smooth patches and adversely affect edges. As mentioned before sharp edges will lead to more accurate segmentation results.

Median filter moves through image pixel by pixel. It replaces each value by the median value of the neighboring pixels. The pattern of neighbor is called

“window”, it slides pixel by pixel over the entire image. First all the pixel values in the “window” are sorted into numerical order and then its replaced with the pixel considering the middle pixel value(median).

The algorithm for median filter is as below:

```
Allocate output pixel value [image width] [image height]

Allocate window [window width* window height]

Edge X = (window width/2) rounded down

Edge Y = (window width/2) rounded down

For X from edge X to image width-edge X

    For Y from edge Y to image height- edge Y

        i = 0

        For FX from 0 to window width

            For FY from 0 to window height

                Window[i]=input pixel value[X + FX -
                edge X][Y + FY - edge Y]

                i = i + 1

            Sort entries in window []

        Output pixel value [X][Y]=
        Window [[window width* window height]/2]
```

In figure 20 different noise are applied on mammograms now in Figure 21, image with noise (a: salt& pepper, b: Gaussian, c: speckle) after employing the median filter are presented.

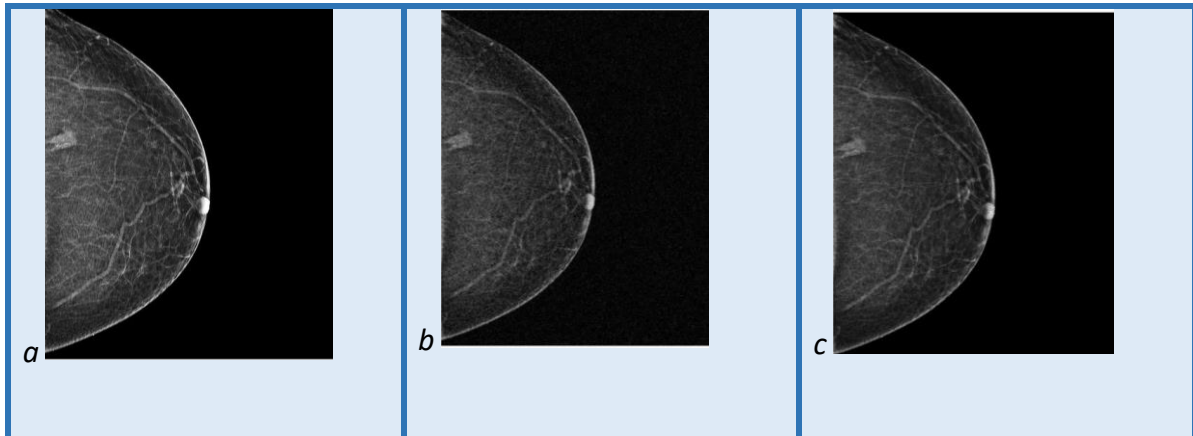


Figure 21: Median filter was employed on images with different noises. a-Image with salt &pepper noise, b-Gaussian noise , c-Speckle noise

As it could be seen, median filter removed salt & pepper and speckle noise properly, but as for Gaussian noise it didn't make that much difference on the image.

-adaptive median filter:

Adaptive filters were used in image processing in order to restore data by removing noise while it didn't blur the structure of image and it preserved details of an image. It's an advance method of median filter. Adaptive median filter can detect which pixel in the image is affected by noise and replaces value of effected pixel with the median value.

For classifying pixels as noise it compares each pixel in the image by its surrounding neighbor pixels. The size of neighborhood and the threshold for comparison are adjustable. A pixel that's its value is different with value of its neighbors is labeled as noise. Then value of that pixel is replace by median pixel value of the pixels in its neighborhood [Rafferty 2003, Makandar 2015]. For generating adaptive median filter thus algorithm is used.

First consider these notations:

Z_{max} = maximum pixel value in S_{xy}

Z_{med} = median pixel value in S_{xy}

Z_{min} = mminimum pixel value in S_{xy}

Z_{xy} = pixel value at coordinates (x, y)

S_{max} = maximum allowed size of S_{xy} , Where S_{xy} is the window

The algorithm works as below:

Level A: $A1 = Z_{med} - Z_{min}$

$A2 = Z_{med} - Z_{max}$

If $A1 > 0$ and $A2 < 0$ then go to level B

Else increase window size

If window size $< S_{max}$ repeat level A

else output Z_{xy}

Level B: $B1 = Z_{xy} - Z_{min}$

$B2 = Z_{xy} - Z_{max}$

If $B1 > 0$ and $B2 < 0$ then output Z_{xy}

Else output Z_{med}

Explanation for the algorithm:

Level A: IF $Z_{min} < Z_{med} < Z_{max}$, then Z_{med} is not an impulse, so go

to level B to test if Z_{xy} is an impulse ..

ELSE Z_{med} is an impulse so the size of the window is increased

A is repeated until Z_{med} is not an impulse and go

to level B or S_{max} reached so then output is Z_{xy} .

Level B: IF $Z_{min} < Z_{xy} < Z_{max}$, then Z_{xy} is not an impulse and output is Z_{xy}

ELSE either $Z_{xy} = Z_{min}$ or $Z_{xy} = Z_{max}$ so output is Z_{med}

(standard median filter) and Z_{med} is not an impulse

In figure 22, image with noises (a: salt& pepper, b: Gaussian, c: speckle) after employing the adaptive median filter are presented. As it could be seen, it removed salt & pepper noise properly while keeping sharp edges and enhanced contrast, but it didn't have severe impact on Gaussian noise.

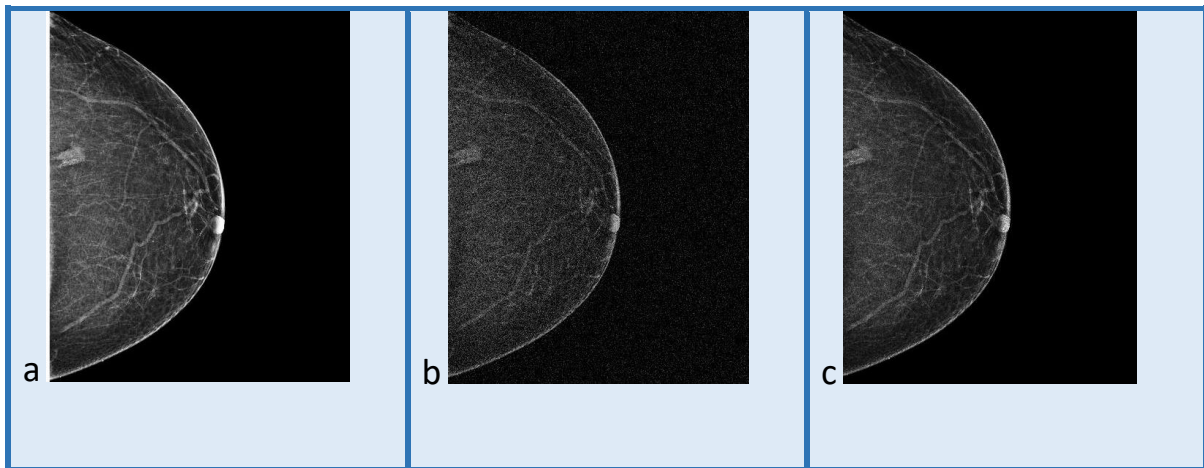


Figure 22: Image with salt & pepper (a), Gaussian (b) Speckle (c) noise after employing adaptive median filter

-wiener filter:

Wiener filter is based in statistical approach and beside removing noise it also can invert any possible blurring in the image. Wiener filter approaches filtering from different angles. It has knowledge of the original image and the noise, one seeks the linear time-invariant(LTI) filter, the filter that output would be as close as possible to the original signal. It can minimize the mean square error(MSE) between desired process and random process [Vijikala 2016, Manas 2015].

Wiener filter are characterized as below:

- 1- Assumption: signal and noise are stationary linear random processes with known spectral characteristics.
- 2- Requirement: filter should be physically realizable.
- 3- Performance criteria: Minimum MSE (mean square error).

Wiener filter are mostly solution for these two cases:

- 1- Cases that non causal filter are acceptable.
- 2- Cases that causal filter are acceptable.

Non causal solution is as below:

$$G(s) = \frac{S_{x,s}(s)}{S_x(s)} e^{\alpha s}$$

Where S are spectral densities

If $g(t)$ is optimal so minimum MSE equation reduces to

$$E(e^2) = R_s(0) - \int_{-\infty}^{\infty} g(t) R_{xs}(t - \alpha) dt$$

The solution $g(t)$ is inverse two-sided Laplace transform of $G(s)$

Causal solution is as below:

$$G(s) = \frac{H(s)}{S_x^-(s)}$$

Where:

$H(s)$ is the causal part $\frac{S_{x,s}(s)}{S_x^-(s)} e^{\alpha s}$

$S_x^+(s)$ is the causal component of $S_x(s)$

$S_x^-(s)$ is the anti causal component of $S_x(s)$

In figure23, image with noise (a: salt& pepper, b: Gaussian, c: speckle) after employing the Wiener filter are presented. As it could be seen in (Fig.23-a) it

removed salt & pepper noise properly but comparing two other mentioned filter it removed Gaussian noise much better(Fig.23-b) and as for speckle noise(Fig.23-c) it reduced the effect of noise.

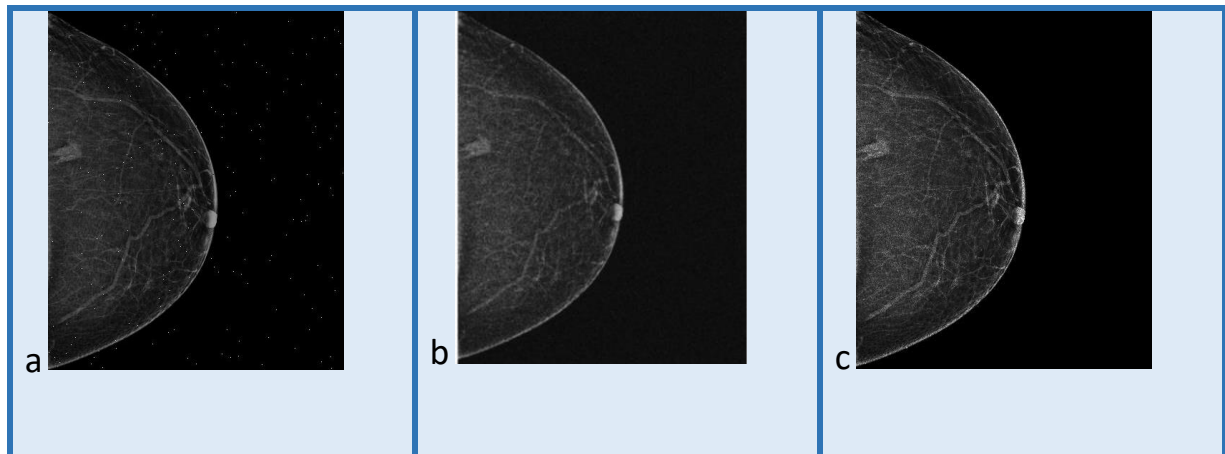


Figure23: Image with salt &pepper (a), Gaussian (b) Speckle (c) Noise after employing adaptive Wiener filter

Most frequent noises on mammogram and their impact on mammogram are explained. Some of noise removal filters are explained and their impact on images are discussed.

After employing these noise removal filters (median, adaptive median, wiener) it was tried to compare the output images. To evaluate the effect of these filter on removing mentioned noises, mean square error (MSE) was computed. The outcomes of MSE values, selected filter for each noise, evaluation of noise on the CAD are presented in section 4.2.1.

After removing noise from mammograms, segmentation could be done which is explained in details in next section.

2.3 Segmentation of ROI

Segmentation is dividing a digital image into regions. pixel in same regions has similar specifications. The goal of segmentation is to transfer an image into something that is more meaningful and much easier to analyze. Image segregation can locate objects and define boundaries [Domínguez 2009]. The result of image segmentation is a contour extracted from image. In that contour each pixel is similar with respect to some characteristics such as color, intensity or texture. The most general segmentation methods are: thresholding, edge base and region based segmentation, which are explained below [Song2017,Li 2001,Kopans 2007].

2.3.1 Thresholding

Thresholding of an image is the simplest and most common segmentation method. It divides the image based on pixels' grey level values. Pixels with values less than "x" will be categorized in one category and rest will be in other category [Domínguez 2009]. The threshold function is defined as below:

$$g(v) = \begin{cases} 0 & \text{if } v < t \\ 1 & \text{if } v \geq t \end{cases}$$

Where t is threshold value and v is grey level value.

There are two types of thresholding: **local thresholding** and **global thresholding**.

Local thresholding using multiple segmentations divides the image into multiple target regions and backgrounds. **Global threshold** method using a single threshold divides the image into two regions: the target and the background [Domínguez 2009].

Threshold segmentation is simple and fast in computation and is the best method when the target and background have high contrast. In case that there is no big difference in gray scale values, such as masses that are ill defined, it couldn't have accurate results and needs to be combined with other methods. In threshold method special details are not considered, so ROI margin which important for giving BIRAD score, couldn't be accurately segmented in this method. (Fig.24)

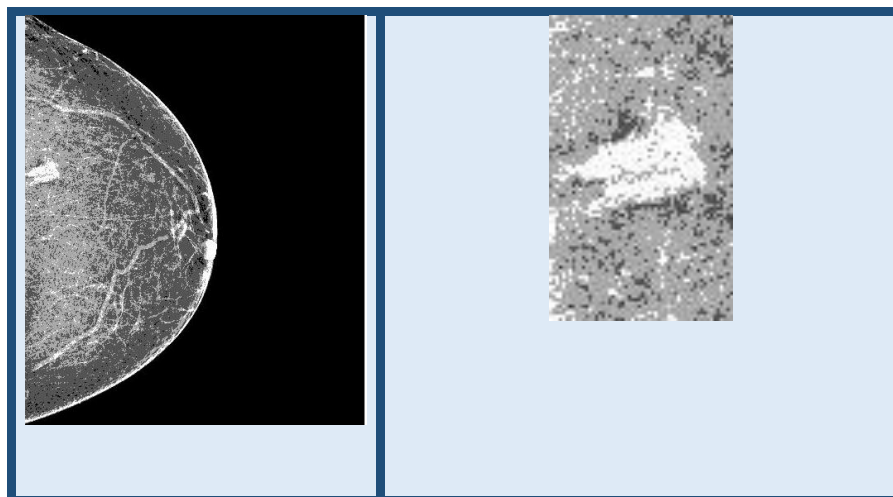


Figure 24: Threshold segmentation of a mammogram. As there is no big difference in density of background and mass, so mass is not fully segmented and margin of mass also is not fully detected.

2.3.2 Edge based segmentation

In this method, in case there is rapid change in intensity value of a pixel an edge is detected. In edge based segmentation first, edges are detected then they are connected together to make boundaries. To detect rapid change in intensity values, first derivative and second derivative operators are used.

Some of most common first derivative operators are: Prewitt operator, Roberts operator and Sobel operator.

Most common second order operators which are nonlinear are: Laplacian, Kirsch operator and Wallis operator.

Where the first derivative of intensity value is greater than a number (for example 'x') or second order derivative is zero then an edge is detected [Song2010, Li 2001].

In edge based algorithm gradient is considered as criteria to stop the evolving contour of the desired object. So gradient of image can't be zero on its boundaries and it may cross the boundaries where the edge is weak. In this model a positive and decreasing function g could define edge detector.

$$\lim_{x \rightarrow \infty} g(x) = 0, g(0) = 1$$

So it can be as below:

$$g(|\nabla u_0(x, y, z)|) = \frac{1}{1 + |\nabla G(x, y, z) u_0(x, y, z)|^2}$$

Where $G * u_0$ is convolution of image u_0 with Gaussian filter.

$$G(x, y, z) = \frac{1}{\sqrt{\sigma}} e^{-|x^2 + y^2 + z^2|/4\sigma}$$

It could be concluded that edge based segmentation is good for images with better contrast between objects and background. Also it's not good for images with lots of edges. (Fig.25-a) shows initialization of mass in a mammogram and (fig.25-b) is the edge base segmentation result. As it could be seen, it didn't segment the margin properly and size of mass is bigger than the one that edge base segmentation has detected.

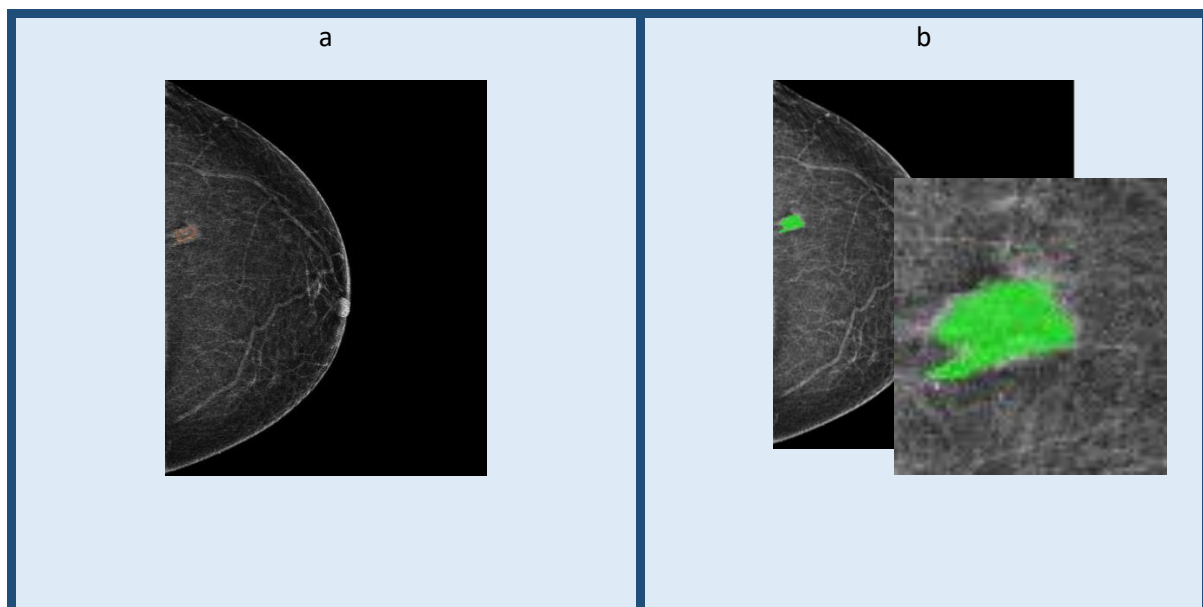


Figure 25: a- First initialization of mass in the mammogram, b- Edge based segmentation of mass.

2.3.3 Region based segmentation

It segments the image into various regions with similar characteristics. There are two methods: ***region growing method, splitting and merging.***

1-Region growing method: this method segment the image based on a seed (initial pixel) which can be selected manually (with prior knowledge) or automatically (with particular application). This method takes the seed as an

input. All pixel in image which are not yet categorized are considered. If it is similar to the seed it will be selected in same region as seed. The region grows till all the pixels in the image are considered. The similarity measurement is the difference between pixels' intensity value and region's mean value. Pixels with smallest difference will be in same region as the chosen seed. In this method accuracy depend on selected initial seed so noise can affect the results and also seed is an additional input. [Cheng 2003,Li 2001]

2-Region splitting and merging: in this method segmentation decision should be based on a value such as variance. Algorithm consist of two phase and need a limit to be considered as max variance in pixel values in a region.

- Splitting phase: in this phase the whole image's variance, as root of a tree, is computed and if it is greater than the limit variance, then the image will be divided into four images (4 quadrants as children of the root). This continues till all the image is divided into squares that their variance value is less than the limit. This system is also known as quad tree segmentation.
- Merging phase: in this phase squares which has common edge will be amalgamated and if the new region's variance value is less than the limit then it will be considered as new region. This continues until no other amalgamation could be done.

In this method, unlike the second phase, first phase result is unique. Second phase results depend on the order of which square are considered. This method is more immune regarding noise but it's an expensive method regarding time and memory usage.

Active contour as most common segmentation method was chosen and is explained below (Fig. 27-c).

2.3.4 Active contour segmentation

Active contour could be categorized into two models: parametric and geometric active contours [Yue 2013] [Beatriz 2004].

1- parametric active contour:

First parametric active contour(snakes), was introduced by Kass et al[Kass 1998] for detecting an object from an image. The idea is that the user initializes the object then it will be moved by image forces to the desired region of interest(ROI). Two types of forces are considered:

- 1) internal forces which are defined within the curve and are designed to keep the model smooth during process.
- 2) external forces are computed from underlying image and are defined to move the initial image toward the ROI.

In this model the contour moves while it was tried to minimize the energy and that's the reason for calling it "snake". This model has some drawbacks:

dependent to initial parametric, numerical instabilities and resampling problem, detecting non-convex objects, minimizing energy and speed [Kass1998] [Xu 1998]. Level set active contour introduced by Osher-Sethian solved most of the drawbacks [Osher 1998]. They introduced a model in which they overcome the problem associated with classical active contour, called level set active contour. Level set is an implicit representation based on partial differential equations (PDEs) for contour evolution. It's a numerical method which intend to solve initial value problem. By building high order advection scheme it can handle subtle topological changes of merger and breakage [Osher 2004]. level set makes an effective and stable algorithm for solving curve evolution equation. A fixed discrete grid in spatial domain and finite difference approximation for spatial derivatives will be used for numerical approximation. It can be extended to any dimension which is not straight forward with the energy minimization scheme [Osher 2004, Malladi 1995, Vese 2002].

This active contour could be categorized into two classes: **region based** and **edge based** model [Gao 2005, Mumford 1989].

- For attracting the contour toward object boundaries edge based model used gradient as control force. Edge based model is sensitive to noise and don't detect weak boundaries and the result is dependent to initial contour placement.

- Region based model for generating the force to attract constraints on the evolution curve, uses image region information. It can segment image with weak boundaries better and it is more robust towards noises.

2-Geometric active contour:

Geometric active contour uses the advantage of level set model and beside those, it is also based on designing speed term so that the evolving front gradually attains zero speed as it gets closer to the object boundaries and comes to stop. The speed makes use of the information from the region enclosed by the evolving front. Also it depends on the boundary of front [Caselles 1993,Caselles1997].

2.3.4.1 Active contour without edge(Chan-Vese)

Classical snake and active contours rely mostly on edge function. For stopping curve evolution, they depend on image gradient. So these models only can detect object with edge defined by gradient and sometimes the stopping function never gets zero and curve may pass through boundaries. Also in noisy images the isotropic smoothing Gaussian should be strong [Wang 2010, Cao 2008].

Chan Vese active contour introduce a model which is not based on gradient of image and it is based on Mumford-Shah segmentation technique stopping

criteria. So this model can detect contour with and without gradient (object with smooth boundaries or discontinuous boundaries.) [Vese 2002, Gao 2005, Mumford 1989].

In Chan-Vese model interior contours are detected and initial curve could be anywhere in the image. Length and area of the curve is important and has minimal partitioning due to use of level set formula. Use of level set contour and Dirac delta function [Dirac delta] causes that the interior contour doesn't grow excessively. Noise has no effect on their contour detection as they used automatic change of topology and second contour. As curves are automatically attracted toward edge and Velocity has global dependence so this model can detect different objects with different intensities and also blurred image. (fig.26) present initialized mass in the mammogram and it was segmented by Chan_vese active contour method. As it could be seen it segmented the ROI properly and edge were accurately detected.

Chan-Vese active contour energy function is as below:

$$F(c_1, c_2, C) = \mu \cdot \text{Lenght}(C) + v \cdot \text{Area}(\text{inside}(C)) \\ + \lambda_1 \int |u_0(x, y) - c_1|^2 dx dy + \lambda_2 \int |u_0(x, y) - c_2|^2 dx dy$$

Where $\lambda_1 > 0, \lambda_2 > 0, \mu \geq 0, v \geq 0$ are constant.

First and second integrals are taken inside and outside the closed curve C. the function is minimized and become as below:

$$\frac{\partial \varphi}{\partial t} = \partial_\epsilon(\varphi) \left(\text{div} \left(\frac{\nabla \varphi}{|\nabla \varphi|} \right) - v - \lambda_1 (u_0 - c_1)^2 - \lambda_2 (u_0 - c_2)^2 \right) \\ \varphi(0, x, y) = \varphi_0(x, y)$$

Where c_1, c_2 are mean intensity values inside and outside of the curve and $\varphi_0(x, y)$ in the initial contour.

Instead of using c_1, c_2 , two fitting functions (e_1, e_2) could be used to locally approximated the image intensities on two side of the contour.

$$e_i(x) = \int k_\sigma(y - x) |I(x) - f_i(y)|^2 dy \quad i = 1, 2$$

$$f_i(x) = \frac{k_\sigma(x) * (\mu_i^\epsilon(\varphi(x)) t(x))}{k_\sigma(x) * (\mu_i^\epsilon(\varphi(x)))}$$

$$k_\sigma(x) = \frac{1}{(2\pi)^{\frac{1}{2}} \sigma^2} e^{-|x|^2/2\sigma^2}$$

$$M_1^\epsilon(\varphi) = H_\epsilon(\varphi) \quad M_2^\epsilon(\varphi) = 1 - H_\epsilon(\varphi)$$

$$H_\epsilon(\varphi) = 1/2 \left(1 + \frac{2}{\pi} \arctan\left(\frac{\varphi}{\epsilon}\right) \right)$$

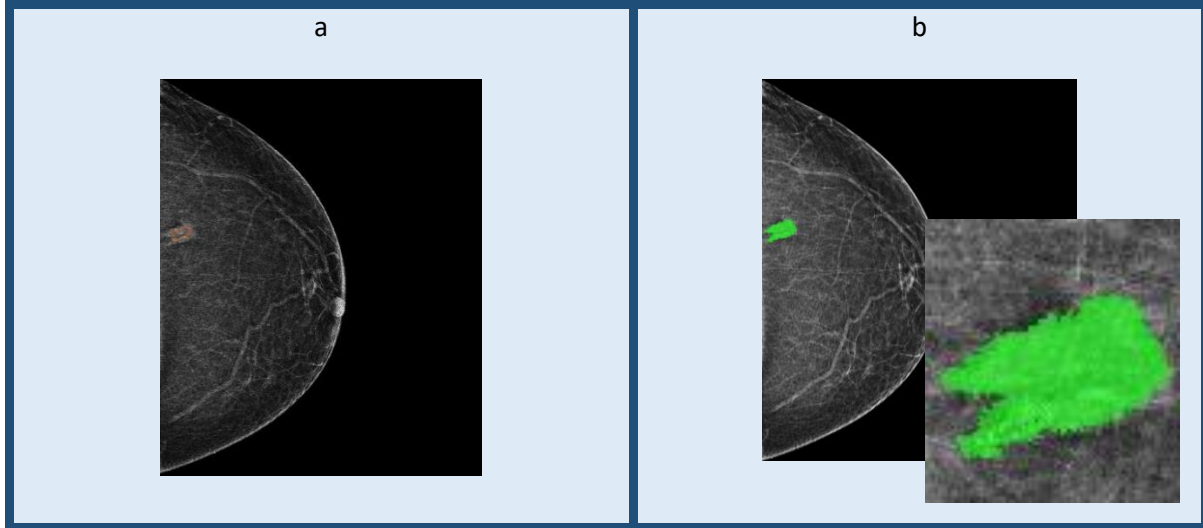


Figure 26: a-ROI initialization, b- Segmented ROI by Chan Vese active contour

2.4 Feature extraction

In previous section an algorithm to segment the ROI region (ROI) were proposed.

In this section, a feature extraction scheme for characterizing the segmented ROI (mass) is discussed.

According to specialist point of view which is a visual aspect, features are categorized into five groups (table 3): intensity of region-of-interest (ROI), shape of ROI, intensity of ROI margin, shape of margin and texture of ROI. In order to fully characterize the mass beside those visual features, other mathematical and texture features also were extracted. In each category, both clinical and mathematical features were extracted [Nagel Rufus 1998,Arevalo 2016,Wang 2016,Alam 2014,Elter 2009].

First order statistics Intensity features	
ROI intensity	Dense, Low density, ISO dense or High density
Margin intensity	Well define or Partially obscured with adjacent area(ill defined)
Shape features	
ROI shape	Round, oval or irregular
Margin shape	Spikey, lobulated or smooth
Second order statistics texture features	
Texture	Homogenous , heterogeneous

Table 3: Medical features

2.4.1 Intensity based features

A) first order statistics:

Features based on the individual gray level (pixel) value of the ROI and its surrounding region are first order statistics. Specialists for diagnosing the mass, mentioned in section 1.2.6, consider the density of a mass by defining it as fatty, low dense (figure 27-e), Iso-dense (figure 27-c) or high dense (figure 27-b). High dense masses are more white while fatty ones are less white [Dhungel 2015,Wang 2016,Elter 2009,Sun 2002].

68 intensity based features are extracted for this reason and they are presented in table 2. Among all those features some were manually selected based on a visual judgment (described in section 4.2.3).

The manually selected features for characterizing a mass based on its density are: STD- Mean, Sum Histogram ROI, Max STD ROI- Max STD Surrounding area, STD surrounding area –STD ROI, Mean Global, Mean ROI, Contrast ROI, contrast surrounding area (table 4).

Beside the density of mass, specialist also use the density of the surrounding area in order to define a mass either as a well-defined or partially obscured by adjacent tissue mass. If there is a big difference between mass intensity and its surrounding area's density, it is a well-defined mass (Fig. 27-d) otherwise its mostly obscured with its surrounding area (Fig. 27-a) [Lehman 2002, Mendez 2004, Baker 1996] [A.C.S 2010] [Nagel Rufus 1998][Bhangale 2015] [Zhang 2009, Soltanian-Zadeh 2004].

For evaluating density of surrounding area, acutance and diff histogram were calculated (Table 4).

Beside those visual and mathematical first order features, some second order statistic such as HARALICK features were also extracted to fully characterize the texture within a mass (Table 4).

B) Second order statistics:

Texture features, as second order statistics, characterize the grey level variation between adjacent pixels in an image. It provides measures of properties such as smoothness, coarseness and regularity of the intensity values in a region. The

Haralick texture features and the grey level co-occurrence matrix (GLCM) are used to characterize textures of mass [Conners 1980, Arivazhagan 2017, Hiremath 2006, Hall-Beyer 2017].

Twenty-two Haralick and co-occurrence matrix features were extracted from which some of the most irrelevant ones were manually/visually deleted. The chosen Haralick texture features were: Cluster Prominence(CPROM), Sum variance(SVARH), Cluster Shade(CSHAD), Sum of squares(SOSVH), Difference variance(DVARH), Sum average(SAVGH), Dis- similarity, maximum probability(MAXPR), Entropy, Sum entropy(SENTH), Autocorrelation(AUTO), Information measure of correlation2(INF2H), Information measure of correlation1(INF1H), Correlation(CORRM), Diff correlation, Difference entropy(DENTH). Beside those Haralick features some other texture features also were detected such as: Solidity, uniformity, Diff moment 2, energy, homogeneity, diff variance of ROI and its surrounding area, also difference of smoothness in ROI and its surrounding area was calculated (table 2) [Papadopoulos 2008, Yang 2005, Cheng 2006, Sheshadri 2007, Blot 2000, Haralick 1973, Haralick 1992, Bovis 2000, Wei 2005].

2.4.2 Shape-based features

The shape of a mass is of great importance for diagnosing a mass, from a specialist's point of view. The shape of a mass is classified as either round (figure

27-e), oval or irregular (Fig. 27-a) [Tzikopoulos 2010, Sampat 2005, Ertas 2001, Ertas 2001, Soltanian-Zadeh 2004]. Shape-based features, also called morphological or geometrical features, are based on shape of ROI (mass) could be used to represent such morphology.

For evaluating shape of mass, features such as compactness and kurtosis were computed (table 2) to show how round or oval is a mass.

Beside shape of a mass, shape of its margin also is of great importance. Shape of margin is described as smooth (Fig. 27-e), lobulated (Fig. 27-c) or needle-like (spikey) (Fig. 27-a).

For evaluating margin's shape, feature called specular index (SI) was calculated. (section 3.2.7 [Boser 1992, Liu 2011]).


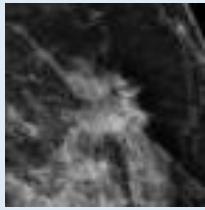
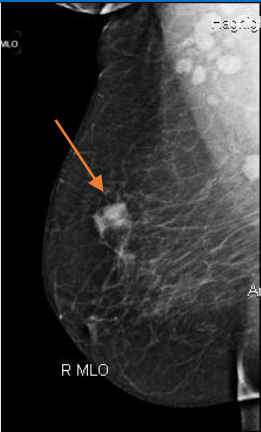
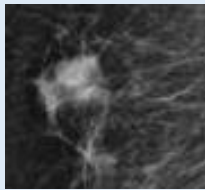
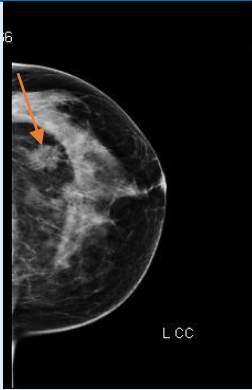
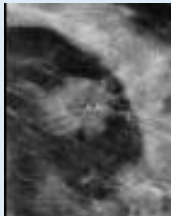
<p>a) Irregular, spiky and ill define mass</p>		
<p>b) High density and irregular</p>		
<p>c) lobulated and ISO dense mass</p>		

Figure 27: a) Irregular, spiky, ill-defined and low dense mass, b) High dense and irregular mass, c) Lobulated and Iso dense mass, d) Well defined, irregular and smooth mass, e) Well defined, irregular, smooth and low dense mass

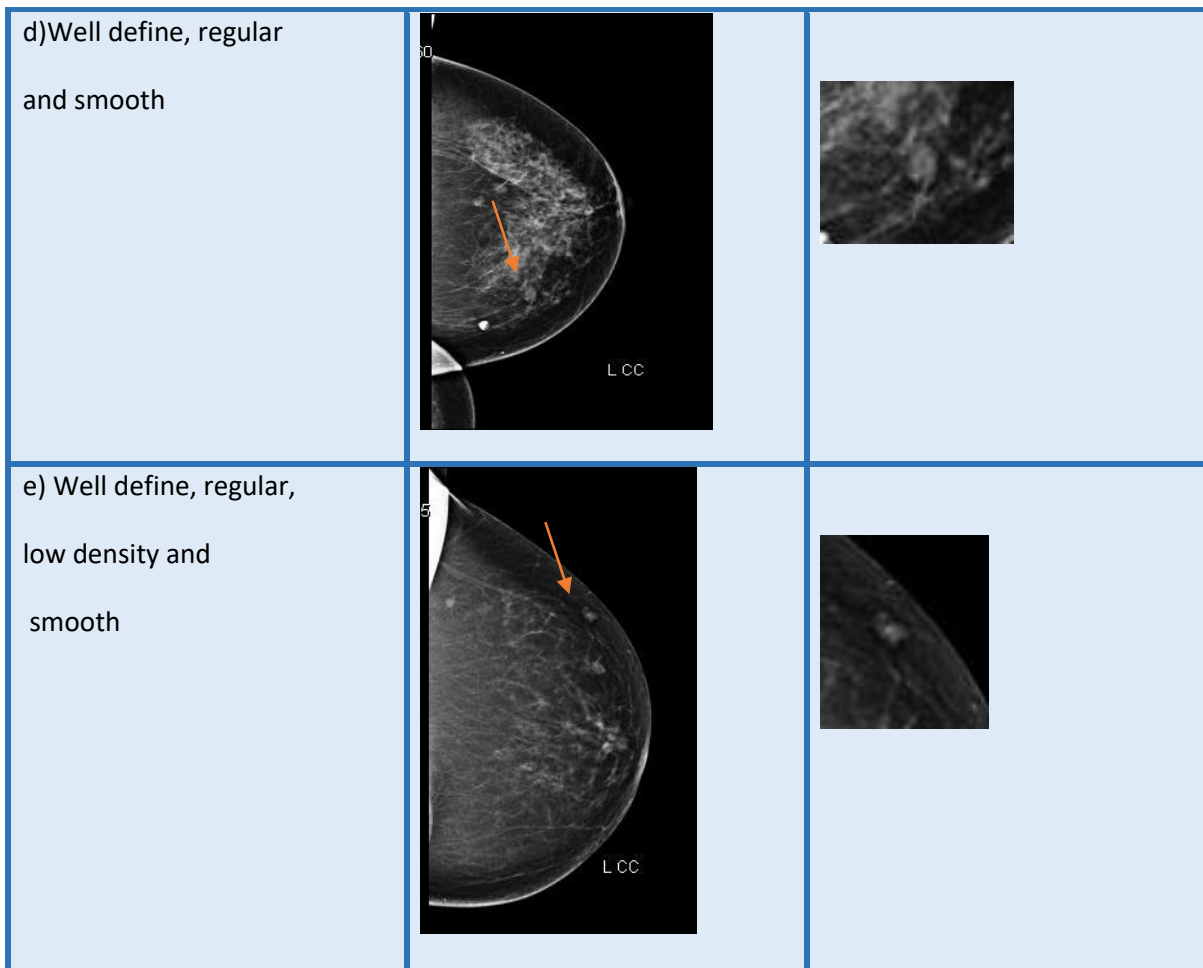


Figure 27: a) Irregular, spiky, ill-defined and low dense mass, b) High dense and irregular mass, c) Lobulated and Iso dense mass, d) Well defined, irregular and smooth mass, e) Well defined, irregular, smooth and low dense mass

<p style="text-align: center;"><i>First order statistics Intensity features</i></p>	
<p>Mean</p>	<p>Mean is average value of pixels' intensity. For a random variable, vector A made up of N scalar observations, the mean is defined as</p> $\mu = 1/n \sum_{i=0}^n A_i$ <p>Mean of surrounding area of the ROI also could be calculated.</p> <p>For obtaining the surrounding area, first the ROI will be dilated 8 times and the resulting image (Dilate-ROI) will be co-ordinate with the Image and the result will be deducted from the original Image. For computing the enclosed are of the ROI the same procedure will be repeated but instead of dilating it will be eroded.</p> <p>Mean global can be used to determine the ratio of contrast between the region and the Mean ROI. Mean-global=Mean (ROI)/Mean (Image)</p> <p>Mean ROI, Mean surrounding area, Mean-global, min ROI, max ROI also were computed.</p>

Table 4: Features and their formula

<i>First order statistics Intensity features</i>	
STD	<p>For measuring how widely the values are spread in the image, standard deviation is calculated. It shows average distance of pixel from the mean. For a random variable vector, A made up of N scalar observations, the standard deviation is defined as</p> $S = \sqrt{\left(\frac{1}{n-1}\right) \sum_{i=1}^n A_i - \mu ^2}$ <p>SD, SD of mage, SD of surrounding area, SD –global=SD (ROI)/SD (image), Min (SD), Max (SD), min SD (surround image), max SD (surround image) were computed.</p>
Histogram	<p>Histogram is graphical representation of intensity values. Normalized histogram (NH) re-assigns the intensity to new extend.</p> $NH = H(ROI) / \text{Area}$ <p>min (histogram ROI), max (histogram ROI), sum (histogram ROI), sum (histogram surrounding image) were computed.</p>
Smoothness	<p>It shows the smoothness of the region.</p> $\text{Smoothness} = 1 / (1 + \text{stdROI}^2);$ <p>smoothness ROI and smoothness of surrounding ROI were computed. shows the smoothness of the region.</p>
Skewness	<p>Skewness shows asymmetry of probability distribution of a real value in histogram.</p> $\text{Skewness} = \frac{E(x-\mu)^3}{s^3}$ <p>Skewness ROI and Skewness of surround image were computed.</p>
Contrast	<p>Contrast surrounding area, contrast ROI were computed.</p> $\text{contrast} = \text{contrast surround ROI} / \text{contrast ROI}$
Inverse	<p>If the high value of the entry of NH is near the main diagonal, inverse will produce high value. where “pt.” is the ROI vector</p> $\text{inverse} = \text{sum}(\text{pt} . (1:c));$ $\text{inverse} = 1 / \text{inverse};$ $\text{inverse ROI} = \text{inverse} * \text{NH};$
Acutance	<p>Acutance measures sharpness or change of density across the margin of mass.</p> $\text{Acutance ROI} = (\text{mean}(\text{Image}) - \text{mean}(\text{surround image}));$

Table 4: Features and their formula

<i>Shape based features</i>	
<i>Compactness</i>	<p>Compactness feature shows how much the ROI is round.</p> $\text{Compactness ROI} = 1 - (4 * 3.14 * \text{Num_pixel}) / \text{perimeterROI}^2$
<i>kurtosis</i>	<p>Kurtosis shows how tall and sharp the central peak of histogram is. Kurtosis is the fourth moment of ROI.</p> $\text{Kurtosis} = \frac{E(x-\mu)^4}{s^4}$ <p>$E(t)$ represents the expected value of the quantity t</p>
<i>Specularity Index</i>	<p>It shows how spikey or lobulated a mass is.</p> <p>Number and height of lobulation and spikes were calculated. (It is explained in section 3.2.7)</p>
<i>Second order statistics texture features</i>	
<i>GLCM and HARALICK features</i>	<p>Autocorrelation(AUTOC), contrast(CONTR), correlation(CORRM), cluster Promina(CPROM), cluster shade(CSHAD), dis similarity(DISSI), maximum probability(MAXPR), sum of squares(SOSVH), sum of average(SVAGH), sum of variance(SVARH), sum entropy(SENTH), difference variance(DVARH), difference entropy(DENTH), information measure of correlation(INF1h, INF2H)</p>
<i>Solidity</i>	$\text{Solidity} = \text{Area} / \text{Convex Area}$
<i>Entropy</i>	<p>It shows measure of disorder in the image's grey level.</p> <p>$E = \text{entropy}(I)$ returns E, a scalar value representing the entropy of intensity image I. Entropy is a statistical measure of randomness that can be used to</p> $\text{Entropy} = \sum_{i,j}^N -\ln(p_{i,j})p_{i,j}$ <p>where p contains the histogram counts returned from image hist.</p> <p>Entropy and entropy surrounding image were computed.</p>
<i>Correlation</i>	<p>If the ROI has large connected sub-component of constant gray level and large difference gray level is created between adjacent components, the Correlation produce a large value.</p> $\sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i,j)}{s_i s_j}$

Table 4: Features and their formula

<i>Second order statistics texture features</i>	
Variance	<p>Variance is the square root of standard deviation.</p> $V = S^2 = \frac{\sum(A_i - \mu)^2}{n - 1}$
Homogeneity	<p>It shows how homogeneity is the ROI.</p> $\sum_{i,j} \frac{p(i,j)}{1 + i - j }$
Energy	<p>Benign masses have low energy value.</p> $\sum_{i,j} p(i,j)^2$
Uniformity	<p>Uniformity is the uniformity of intensity in histogram of ROI.</p> <p style="text-align: center;">uniformity = NH^2</p>
Moment2	<p>Moment2 of ROI, its surrounding area and difference moment were computed.</p> $M2 = \frac{E(x-\mu)^2}{s^2}$
Contrast	<p>Contrast surrounding area, contrast ROI were computed.</p> <p style="text-align: center;">Contrast ROI = $\sum_{i,j} i - j ^2 p(i,j)$</p> <p>contrast = contrast surround ROI \ contrast ROI;</p>
Prominence	$P = Sgn(B) B ^{\frac{1}{4}}$ <p style="text-align: center;"><i>Sgn is sign of number(-1,0, +1)</i></p>
Shade	$Sh = Sgn(A) A ^{\frac{1}{3}}$ $A = \sum_{i,j=0}^{N-1} \frac{(i+j-2\mu)^3 p_{i,j}}{\sigma^3 (\sqrt{2(1+c)})^3}$ $B = \sum_{i,j=0}^{N-1} \frac{(i+j-2\mu)^4 p_{i,j}}{4\sigma^4 (1+c)^2}$ <p style="text-align: center;">C is correlation feature</p>

Table 4: Features and their formula

2.5 Discussion and conclusion

In this chapter most frequent noises on mammograms (Salt & Pepper, Gaussian and Speckle noise) were explained. Salt & Pepper noise makes white spots on the mammogram that may cause false micro calcification detection. As mentioned in section 3.3, it could cause miss-classification of mass. Speckle and Gaussian noises, lower the contrast of mammogram, which also may reduce classification accuracy. So it is essential to use noise reduction filters in order to reduce any possible noise on mammogram. Therefore, several noise filter such as median filter, adaptive median filter and Wiener filter were proposed. Experimental results of noise reduction methods are presented in section 4.2.1. Different segmentation methods to segment a mass are also presented, it could be concluded that threshold segmentation couldn't detect ROI margin in ill-defined masses accurately. Edge based segmentation also couldn't detect spatial details in mammograms when there is no big difference in background and target. Region based active contour segmentation presented in section 2.2.3, is the most proper segmentation technique, as it doesn't grow excessively, it detected edges properly and noise doesn't have effect on its edge detection. This chapter completed by introducing medical and mathematical features including intensity based, shape and texture features in order to fully characterize a mass.

In next chapter feature selection methods, classification algorithms, meta data that are used for classification mass with BIRAD score are presented.

Chapter 3

3. Pattern recognition

3.1 Feature selection and dimension reduction

All possible related features that may characterize mass are explained and extracted in previous chapter. There is a hesitation of extracting irrelevant and redundant features, so for selecting the most relevant ones that have dominant effect on characterizing mass, feature selection methods should be used. Feature selection is the process of selecting subset of relevant features. Feature selection is used for these reasons:

- simplify the model in order to make it easier for researchers to interpret it.
- make training time shorter.
- reduce curse of dimensionality.
- reduction of variance.

In feature selection redundant and irrelevant features will be removed. Removing these features doesn't incur loss of information [Hanmandlu 2004, Torheim 2001, Sohail 2011, Gibbs 2008].

- Relevance features: [Ding 2005]

If there are some examples in the instance space that twiddling value of x affects the classification given by the target, feature x is a relevant feature.

In other word, relevance is computed as below:

$$D(S, c) = \frac{1}{|S|} \sum_{f_i \in S} I(f_i, c)$$

average value of all mutual information value between each f and class c .

- Redundant features: Features that express more than once. Redundancy is the relation between two features. [Peng 2005]

$$R(s) = \frac{1}{|S|} \sum_{f_i, f_j \in S} I(f_i, f_j)$$

Average value of all mutual information values between feature f_i and f_j

Feature selection could be done using different approaches including: **filter**, **wrapper and embedded**. [Pudil 1994, Somol 1999, Jain 1997, Yang 1998, Kudo 2000]

- ***In filter method***, features are evaluated by employing an independent test feature. Features are highly correlated with the class label and uncorrelated with each other's. Filter method produce a feature set that

is tuned to any specific prediction model so its computationally less expensive comparing other models.

Filter method is fast but it doesn't use learning algorithms, and no iteration is considered between feature selection and classification, hence it has lower classification performance.

- ***wrapper method*** uses a machine learning algorithm to find an efficient subset and score it. Each new subset is used in training section and then its tested and scored on a hold out method. Wrapper method uses classification accuracy as a measure for feature selection. Average area under the ROC curve(AUC) of k-fold cross validation is used for computing classification accuracy. Wrapper method is computationally more expensive comparing filter method but have better classification performance. Also, in case that there are not enough observations, overfitting might happen.
- ***embedded method*** is a combination of both mentioned methods. It uses evaluation function and independent test both.

Performance of several feature selection algorithms, one algorithm from each method, are compared. The compared algorithms are: sequential forward floating feature selection(SFFS), KruskalWallis and minimum redundancy and maximum relevance (MRMR). As MRMR could remove redundant and irrelevant

features [Peng 2005, Liberman 1997], so it was decided to also execute SFFS+MRMR and KruskalWallis+ MRMR (apply MRMR algorithm on SFFS features subset and Kruskal Wallis feature subset). In addition, principal component analysis (PCA) was also investigated as a technique to reduce the dimensionality of the feature set.

3.1.1 Sequential forward floating feature selection(SFFS)

Considering filter and wrapper method's specification, mentioned in section 2.4, SFFS feature selection method as a bidirectional search was used. SFFS starts from null subset. Performs an iterative procedure for selecting most significant feature and add it to the subset. After each iteration, results are compared to those of previous steps. If the outcome has improved, the new subset = subset + most significant feature. Then repeatedly find and delete least significant feature. After each iteration, results are compared to those of previous steps. If the outcome hasn't improved the new subset = subset- least significant feature. The most and the least significant features are selected by applying wrapper algorithm and evaluation criterion [Mohanty 2013, Shirbani 2016].

SFFS aggregate the best feature and eliminate the worst feature iteratively. While in SFS as a filter method, if one feature is selected or deleted (not chosen) there is no way to remove it or add it to the subset, while in bidirectional SFFS inclusion and exclusion could be done. The algorithm is described below:

$$Y = \{y_1, y_2, y_3, \dots, y_d\}$$

$$X = \{x_j \mid j = 1, 2, \dots, k\}$$

$$k = (0, 1, 2, \dots, d)$$

$$d < k$$

The algorithm is initializing with an empty set.

1-start with empty set $k=0$ (k is size of subset) $y_0 = \{\Phi\}$

2-sequentially add next best feature x^+ that maximize function $J(Y_k + x^+)$
when added to feature Y_k , (already selected features)

$$x^+ = \arg \max J(Y_k + x) \quad x \neq Y_k$$

$$\text{update } Y_{k+1} = Y_k + x^+ \quad ; \quad k=k+1$$

3-sequentially remove the worst feature x^- that reduce value of function

$$J(Y_k - x^-)$$

$$x^- = \arg \max J(Y_k - X) \quad x \neq Y_k$$

4- If $J(Y_k - x^-) > J(Y_k)$ then

$$\text{update } Y_{k+1} = Y_k - x^- \quad ; \quad k=k+1; \text{ go to step 3}$$

else go to step 2

3.1.2 KruskalWallis feature selection

Kruskal Wallis is a wrapper method. It is a non-parametric version of Analysis of variance (ANOVA). Median of groups of data are compared in order to determine if the sample come from the same class or not [145] [152].

It orders the data from smallest to largest in order to find ranks and takes the numeric index of its ordering. It uses ranks of data instead of numeric values for computing test statistics. Rank for an observation is equal to rank of all

observations tied with it. It assumes that samples coming from populations with same continuous distribution and all observations are independent. Where ANOVA assumes that populations have normal distributions. Kruskal Wallis uses chi-square statistics and p-value instead of F-statistic in ANOVA.

Kruskal Wallis tests variance using population variance among groups. If two or more classes have equal median it will give a value of P. In case the P is close to 0 it means that feature contains discriminative information and it will be chosen. Kruskal Wallis comparing other wrapper methods is computationally less expensive and simple to use.

The algorithm works as below:

1- Rank all data from all groups

2- Test statistic is as below:

$$H = (N - 1) \frac{\sum_{i=1}^g n_i (\bar{r}_i - \bar{r})^2}{\sum_{i=1}^g \sum_{j=1}^{n_i} (r_{ij} - \bar{r})^2}$$

n_i is number of observation in group i

r_{ij} is Rank of observation j in group i

N is total number of observation in all groups

$$\bar{r}_i = \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i} \text{ Is average rank of observations in group } i$$

3- $\bar{r} = 1/2(N+1)$ Is Average of all r_{ij}

4- If there are no ties in data, the demonstrator is as $:(N-1) N(N+1)$

$$\text{And } \bar{r} = 1/2(N + 1)$$

So

$$H = \frac{12}{N(N+1)} \sum_{i=1}^g \frac{\bar{r}_i}{n_i} - 3(N + 1)$$

5- A correction for ties could be done by dividing H by $1 - \frac{\sum_{i=1}^G (t_i^3 - t_i)}{N^3 - N}$

G is number of grouping of different ties ranks

t_i is number of tied values in group *i*

P value is approximated by

$$Pr(\chi_{g-1}^2 \geq H)$$

3.1.3 MRMR feature selection

Sequential Forward (SF), Sequential Backward(SB) and floating selection all find feature subsets with max relevance (features that has strongest correlation to classification variable), but they also contain redundant features. MRMR as a filter method, select features mutually far away from each other while still having high correlation to classification variable. MRMR removes redundant features [Peng 2005, Liberman 1997].

In this algorithm relevance is calculated by f- statistic for continuous features and mutual information for discrete features. Redundancy is calculated by Pearson correlation coefficient for continuous features and mutual information for discrete features.

Features are selected one by one employing greedy search in order to maximize the objective function.

The objective function could be mutual information difference criterion (MID) or mutual information quotient criterion (MIQ) [Sheshadri 2007].

MRMR algorithms is as below:

The mutual information of two variable x and y is defined according their joint probabilistic distribution, p (x, y).

P(x) ad p(y) are respective marginal probabilities.

$$I(x, y) = \sum_{i,j} p(x_i, y_i) \log \frac{p(x_i, y_i)}{p(x_i)p(y_i)}$$

1-Minimize Redundancy: the idea is select features that they are mutually maximally dissimilar.

$$\text{Min } w_I, w_I = \frac{1}{|S|^2} \sum_{i,j \in S} I(i, j)$$

S is the set of features.

I(i, j) is mutual information between features i and j

2-Maximize Relevance: for measuring discriminant power of features mutual

information is used between target classes $h=\{h_1; h_2, \dots, h_k\}$

I(H, j) shows the relevance of feature g_i for classification. So

maximum relevance

condition is to maximize the relevance of all features in S.

$$\text{Max } V_I, V_I = \frac{1}{|S|} \sum_{i \in S} I(H, j)$$

H=target

3-Combine Redundancy and Relevance:

Additive combination: Max(V-W)

Multiplicative combination: Max (V /W)

3.1.4 Principle Component Analysis(PCA)

PCA, find a pattern in a data set and present the data in a way to highlight the similarities and differences. PCA, reduce its number of dimension, so data is compressed. This compress is done with minimum loss of information and maximum variance of the data [Smith 2002, Rady 2011].

The required steps are as below:

1-consider d - dimension data set

2-compute the mean value for every dimension in dataset

3-compute the covariance matrix of the whole data set

4-compute Eigen vector and corresponding Eigen values

5-sort the Eigen vectors, according Eigen values from high to low.

Data with higher Eigen value are more important than the others.

6-choose k most valuable eigenvectors that have largest eigenvalues

7-use eigenvector matrix to present new data

$$\text{New data} = \text{raw feature vector} * \text{raw data adjust}$$

$$\text{Raw feature vector} = \text{matrix with the eigenvectors in column transpose}$$

(considering the most significant ones in first column)

$$\text{Raw data adjust} = \text{mean} - \text{adjust data (transposed)}$$

Different feature selection methods and their algorithms are explained. Next step was classifying of a mass based on selected features. In next section some of most frequent classification techniques are explained.

3.2 Classification

A classification method consists of 2 steps:

1-Training: a model is constructed from training instances. Classification algorithm find relationship between predictors and targets. Relationship are summarized in a “model”.

2-Testing: test the model on samples whose class labels are known but not used in training phase.

The following classification algorithms were introduced and used in our proposed CAD. [Costa 2011, Daugman 1980, Dem˘sar 2014] [Cheng 2006, Chan 1995, Sahiner 2018, Vibha 2006]

3.2.1 Classification Tree and Regression Tree

A classification tree, is a tree that each non-leaf node (internal node) is as an input feature and each leaf is a class label. A tree is made by splitting the feature set in to subsets, based on attributes, it is a recursive top down process. This recursive partitioning is stopped when the subset at that node has all the same value as the target value, or when splitting doesn't add any value to the prediction. Classification tree is greedy method. Two most popular classifications trees are C4.5 and CART [Sun 2002, Leonardo 2006, Beatriz 2004, Flores 2004].

Classification and Regression tree(CART) consist of two procedures: [Liu 2011, Polat 2009]

1- **Classification tree**: when the predicted output is the class which the data belongs to.

2- **Regression tree**: when the predicted output can be a number.

As mentioned before, CART is a top down tree that chooses a variable that best split the set at each step. Different algorithms choose different methods as splitting criteria. CART chooses GINI impurity as the function for measuring quality of split. CART employ 10-fold cross validation for estimating error rate, where C4.5 uses entropy as impurity function and heuristic formulate for estimating error.

Gini Impurity

Gini is used by CART algorithm. It measures how often a randomly chosen feature from the set would be labelled in-correctly [Vibha 2006] [Loh 2011]. (if it was randomly labelled according to the distribution of labelled in the subset.)

The Gini could be computed as below:

Sum the probability of p_i of item" i "being chosen times the probability of a mistake in categorizing that item.

$$\sum_{i \neq k} p_k = 1 - p_i$$

When all cases in the node are in a same category it reaches to its minimum (0).

For computing Gini function for set of items with "J" classes and consider p_i as the fraction of items labeled with class" j " in the set.

$$\begin{aligned} I_G(\mathbf{p}) &= \sum_{i=1}^j p_i \sum_{k \neq i} p_k = \sum_{i=1}^j p_i (1 - p_i) = \sum_{i=1}^j (p_i - p_i^2) \\ &= \sum p_i - \sum p_i^2 = 1 - \sum_{i=1}^j p_i^2 \end{aligned}$$

Decision tree advantage and dis advantages:

It is based on some if-then-else rules, so interpretation of results is very simple and fast and new instances could be classified very fast. Also it is possible to use non-parametric and nonlinear functions. Decision trees are easy to interpret, fast for fitting and prediction, and low on memory usage, but they can have low predictive accuracy.

There are three types of decision trees: simple, medium and complex.

Simple tree has few leafs and makes coarse distinction between classes.

Medium tree has more leaves comparing to simple tree.

Complex tree has many leaves and makes fine distinction between classes.

In order to control growth of tree, number of branch point should be defined. For defining number of branches max number of splits should be set. The growth should be controlled to stop overfitting. Complex tree has many leaf and for training data its highly accurate but not for test set. As its leafy so it may over train and its validation accuracy is lower than its training accuracy.

Tree algorithm will reclusively and ultimately extract all information from data and this could lead to “over learning” and “over fitting”. the simplest way to solve this problem is to stop generating new split in cases that new results have little improvement. This could be done through cross validation and k- fold cross validation [Sun 2002, Leonardo 2006,Vibha 2006,Liu 2011, Polat 2009].

3.2.2 Classification KNN

K nearest neighbor(KNN) is a type of instance based or lazy learning which both metric(nearest) and number of neighbors are altered. In KNN an object is classified by majority votes of its neighbors. The object is assigned to the class most common among its K nearest neighbors. This method has lack of robustness. To solve it K could be considered bigger than 1 then the majority vote of neighbors will decide the outcome of the class labeling. The drawback is that in

case $k=n$ then new cases will belong to class most frequently represented. The problem could be avoided by assigning weight to each vote. Weight is a function of distance between the known and unknown instances. So distanced instances will have little influence on decision compared to instances in near neighbor. The distance metrics in KNN could be: Euclidean, city block, correlation, Jaccard and etc. [Hastie 2008, Manning 2008, Jump up ^ Altman 1992].

Euclidean: Euclidean distance refers to distance between two points. Consider two points with coordinate (x, y) and (a, b) . The Euclidean distance(distance) will be as below:

$$distance(x, y), (a, b) = \sqrt{(x - a)^2 + (y - b)^2}$$

City block: It's also known as Manhattan distance. It measures the path between two pixels based on 4-connected neighbors and diagonal moves are not allowed. The formula will be as below:

$$D = |x_2 - x_1| + |y_2 - y_1|$$

Correlation: It shows dependence of two random vectors (it measure statistical dependence). Distance correlation is zero for vectors that are independent. It is useful when vectors x and y have weak linear dependence but they have strong nonlinear dependence. The formula is as below:

$$dcov^2(x, y) = \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n a_{j,k} b_{j,k}$$

Jaccard: Jaccard similarity measures similarity of two data sets. It compares members in two data sets to see which members are shared and which are not.

The Jaccard coefficient for similarity measurement is as:

$$j(a, b) = \frac{|a \cap b|}{|a \cup b|} = \frac{|a \cap b|}{|a| + |b| - |a \cap b|}$$

Jaccard distance measures dissimilarity between two data sets. It is obtained by subtracting Jaccard coefficient from 1.

$$d_j = 1 - j(a, b) = \frac{|a \cup b| - |a \cap b|}{|a \cup b|}$$

When multiple classes have the same number of nearest points among the K neighbors, a tie occurs. If multiple classes have the same smallest cost, KNN uses tie-breaking methods such as nearest, random, or smallest.

KNN is a two-phase classification:

Training: It uses training examples as vectors, in a multidimensional feature space

each with a class label. Training phase store feature vector and class label of training samples.

Classification: a test sample (unlabeled vector) is classified. It assigns the label which is the most frequent among “k” training samples nearest to the test sample.

“k” is selected by user. The best choice for “k” depends on data. Larger value for “k” reduce noise while making boundaries between classes less distinct. In cases where the class is predicted to be closest training sample(k=1) is called nearest neighbour algorithm. In binary classification its best to choose “k” an odd number and this will avoid tied vote.

Noisy and irrelevant features can degrade KNN accuracy.

Nearest Neighbor Classifiers has good predictive accuracy in low dimensions, but not in high dimensions. They have fast fitting speed, medium prediction speed, high memory usage, and are not easy to interpret. Different KNN could be used such as: fine KNN, Medium KNN, coarse KNN, cosine KNN, Cubic KNN and weighted KNN.

KNN algorithm is as below:

$(x, y)(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$ pairs taking value in $R^d * \{1, 2\}$

Y is class label of x so

$x|y = r \sim p_r$ for $r=1, 2$

Given some norm $\|.\|$ on R^d appoint $x \in R^d$

Let $(x_{(1)}, y_{(1)}), \dots \dots (x_{(n)}, y_{(n)})$ a recording of training data set such that $\|x_{(1)} - x\| \leq$
 $\dots \dots \leq \|x_{(n)} - x\|$

3.2.3 Classification SVM

Support vector machine (SVM) is a supervised learning classification. SVM by finding the best hyperplane classifies the data. Hyperplane separates data points of one class from those of the other class. The best hyperplane for an SVM means the one with the largest margin between the two classes. Margin means the maximal width of the slab parallel to the hyperplane that has no interior data points [Corinna Cortes 1995, Liu 2014]. Data point that are close to separating hyperplane and are on the boundary of the slab are called support vectors. SVM can also use a soft margin, meaning a hyperplane that separates many, but not all data points. Support vector machines have high predictive accuracy, medium fitting speed, and can have good prediction speed and memory usage with few support vectors. Linear SVM is easy to interpret, but other kernel functions are less easy to interpret [Zhang 2009, Hsu 2010].

SVM uses a set of functions known as Kernel. Kernel takes inputs and transform it into another form. Different kernel functions could be used:

Linear SVM: When inputs could be linearly separable, linear SVM could be used.

SVM using linear kernel is often equivalent to non-kernel SVM.

The equation is as below:

$$\begin{cases} 1 & \text{if } W^T + x = 0 \\ 0 & \text{if } W^T + x \neq 0 \end{cases}$$

Gaussian SVM: Gaussian kernel is a weighted linear function computed between data and the support vector. When there is no prior knowledge about the data, Gaussian function is used to Makes finely detailed distinctions between classes.

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Quadratic and Cubic SVM: Other Kernels look at the features in order to determine their similarity. Polynomial (quadric and cubic) kernels beside that, also looks at the combination of features to find similarities and such combinations are called interaction features.

Consider x, y as vectors of features computed from testing and training (vectors in the input space) and $c \geq 0$ is a free parameter.

The d-degree polynomial is as below:

$$g(x, y) = (x^T + c)^d$$

In training phase SVM train the system by training samples and in classification phase it determined which class the test sample belongs to.

SVM works as below:

$(\vec{x}_1, y_1), \dots, (\vec{x}_n, y_n)$: training data set of n points

y_i is the class that point x_i belongs to, its either 1 or -1

\vec{x}_i is p dimensional real vector

The goal is to find "maximum margin hyperplane" which divide group of points \vec{x}_i for

$y_i = 1$ from the group of points for which $y_i = -1$.

This is done in a way that distance between the nearest point \vec{x}_i and the hyperplane is maximized.

Hyperplanes could be written as below:

$$\vec{w} \cdot \vec{x} - b = 0$$

\vec{w} is normal vector to the Hyperplane

$\frac{b}{\|\vec{w}\|}$ determine the offset of the hyperplane from the origin along the normal vector.

Hard margin

Two parallel hyperplane could be selected in case the training data in linearly separable. These two parallel hyperplanes can separate two classes of data so that the distance between them is quite large." margin" is the region between these two hyperplane and maximum margin hyperplane is the one that lies half way between two hyperplanes [Burges 1999]. If data set is normalized so these hyperplanes can be described as below:

$\vec{w} \cdot \vec{x} - b = 1$ point on or above this hyperplane is class label 1

$\vec{w} \cdot \vec{x} - b = -1$ point on or under this hyperplane is class label -1

$\frac{2}{\|\vec{w}\|}$ is the distance between two hyperplane , $\|\vec{w}\|$ should be minimized to maximize the distance.

Also points shouldn't fall into the margin.

These constraints are added:

$$\vec{w} \cdot \vec{x}_i - b \geq 1 \quad \text{if } y_i = 1$$

Or

$$\vec{w} \cdot \vec{x}_i - b \leq -1 \quad \text{if } y_i = -1$$

It means that each point should fall in correct side of margin.

This could be written as:

$$y_i(\vec{w} \cdot \vec{x}_i - b) \geq 1 \quad 1 \leq i \leq n$$

So:

$$\text{Minimize } \|\vec{w}\| \text{ subject to } y_i(\vec{w} \cdot \vec{x}_i - b) \geq 1 \quad \text{for } i = 1, \dots, n^n$$

\vec{w} and b solve the classification $\vec{x} \rightarrow \text{sgn}(\vec{w} \cdot \vec{x} - b)$

Maximum hyperplane is determined by x_i that fall nearest to it and these x_i are called "support vectors".

Soft margin

In cases that the data are not linearly separable this function should be used:

$$\max(0, 1 - y_i(\vec{w} \cdot \vec{x}_i - b))$$

y_i is the target (1,-1) and $(\vec{w} \cdot \vec{x}_i - b)$ is output.

If x_i lies on correct side of margin then constraint in function $y_i(\vec{w} \cdot \vec{x}_i - b) \geq 1 \quad 1 \leq i \leq n$ is satisfied then this function will be 0.

If x_i doesn't lie on the correct side of margin then the function value is far from 0 and the aim is to minimize this distance [90].

The distance is as below:

$$\frac{1}{n} \sum_{i=1}^n \max(0, 1 - y_i(\vec{w} \cdot \vec{x}_i - b)) + \rho \|\vec{w}\|^2$$

ρ determines the tradeoff between increasing the margin size and ensuring that x_i lies on the correct side of margin.

3.2.4 Classification ensemble

Ensemble method use multiple learning algorithms in order to obtain better predictions, comparing using constituent learning algorithms. Ensemble Classifiers meld results from many weak learners into one predictor. The quality of ensemble classifier depends on the used methods. Ensemble method gives better results when there is diversity of models in that method. It's a supervised learning algorithm because it can be trained and then used. Tree according their specification mentioned in 2.5.1 is commonly used in ensemble methods [Eltoukhy 2010, Sun2002, Samson2000, Breiman,1984, Kupinski1997]. Some specification of ensemble methods is presented in table 5.

Some of most common ensemble methods are:

- **Boosted tree:**

it trains each new instance in a way that it emphasizes the training instances that were previously miss-classified." Ada-boost" is a sample of it and it can be used for regression-type and classification-type problems. Boosted tree construct shallow trees. Its little time and money consuming but it needs more ensemble member than bag tree [Eltoukhy 2010].

Ada boost: This is an iterative procedure. It starts with an unweighted training set and build a classifier. If a training data is miss classified, then

its weight will be increased(boosted). New classifier will be made with this new weighted training set. A score is given to each classifier.

Final classifier is the linear combination of classifiers from each stage. Ada boost is sensitive to noisy data and outliers. Ada boost mostly is used with decision tree and have good performance. Information gathered from Ada boost classifier (information about misclassified data and their weight) in each training sample is used in the tree growing algorithm. So later trees will focus on those misclassified weighted training data.

- **Bagged decision tree**: It makes classifier on random training subsets from original training set and generate prediction (by voting or by prediction). This is an iterative procedure. Many of the training subsets may be used several times and some may not be used at all. Bag tree constructs deep tree, its time and money consuming.
- **Rusboost**: In cases that sample data are unbalanced (number of data in one class is much higher than number of data in other class), testing data could be miss classified in favor of class with more training data. To solve this random under sampling (RUS) boost method could be used. This method randomly deletes data from training datasets until the number of training data is balanced in both classes. RUS boost is a hybrid classifier. It

re-weights and resample data. This method resample data according their given weights.

- **Subspace Discriminant:** In random subspace, also called feature bagging, features are randomly selected for each learner. So classifier is trained with random samples of features instead of entire features. Each learner doesn't focus on features that are highly predictive as the correlation between estimators are reduced. On the other hand, discriminant analysis tries to maximize separability of two class by finding an optimal subspace. It minimizes the inner distance and maximize the between distance simultaneously.
- **Subspace KNN:** As mentioned before random subspace selects features randomly and trains classifier by those feature subsets. Each time that a subspace is selected, KNN is computed. Each KNN is assembled for majority vote on the class membership of the test samples. If one training sample is among the KNN in more than one subspace so that training sample may appear more than once.

Classifier Name	Predictive Accuracy	Ensemble Method	Fitting Speed	Prediction Speed	Memory Usage
<i>Boosted Trees</i>	<i>require parameter tuning but have high accuracy</i>	<i>Ada Boost, with Decision Tree learners</i>	<i>Fast with few learners, but might need more learners than bagged trees</i>	<i>Fast with few learners</i>	<i>Low</i>
<i>Bagged Trees</i>	<i>Medium to high</i>	<i>Bag, with Decision Tree learners</i>	<i>Its Slow for huge data sets</i>	<i>Slow for huge data sets</i>	<i>High for huge data sets</i>
<i>Subspace KNN</i>	<i>Good</i>	<i>Subspace, with Nearest Neighbor learners</i>	<i>Medium</i>	<i>Medium</i>	<i>High</i>
<i>Subspace Discriminant</i>	<i>Good for many predictors, accuracy dependent on data set</i>	<i>Subspace, with Discriminant learners</i>	<i>Fast</i>	<i>Fast</i>	<i>Low</i>
<i>RUS Boost Trees</i>	<i>Good for skewed data</i>	<i>RUS Boost, with Decision Tree learners</i>	<i>Fast with few learners</i>	<i>Fast with few learners</i>	<i>Low</i>

Table 5: Some specifications for ensemble methods

3.3 Meta data

Specialist beside all extracted features mentioned in 2.3, uses meta data for giving BIRAD score to a mass. Specialist consider these meta data for diagnosing, particularly for sub classifying mass with BIRAD 4 to BIRAD 4a,4b,4c:

- **Stable:** If patient has previous mammogram and the same mass exist before and there is no change in shape, density or size of the mass, so that mass will be considered stable, hence mass will have BIRAD score of 2. Stability was achieved by some Haralick features.
- **Specularity index:** If border of mass isn't smooth, its lobulated or spikey. Mass with spikey margin has higher BIRAD (Fig.28-c). Degree of Spikiness was extracted as features. High dense mass with lobulations has higher BIRAD score(Fig.28-a).
- **Existence of micro calcification:** If there is a micro calcification in the mass, specialists will give higher score of BIRAD to mass comparing the same mass with no micro classification. Existence of any micro calcification in mass was extracted as a feature (Fig.28 b).
- **Retraction:** If the mass causes any retraction on the skin, hence specialist will give higher BIRAD score to mass comparing with mass that cause no retraction in mass. Retraction in mass was extracted as a binary feature (Fig.28-d).

- **Family history:** Comparing the mass in patient with no family history of cancer with patient with family history of cancer, they will give a higher BIRAD to the mass. Family history was considered as a binary feature.
- **Change in appearance:** If mass makes any change in appearance of the breast, specialist will give higher score of BIRAD comparing the mass with same specification that didn't make any change in the appearance of breast. It was extracted as a binary feature (Fig.28-d).
- **Position:** In case the mass is close to nipple specialist will give higher BIRAD score to it comparing the mass that is not close to nipple. It was considered as a binary feature.
- **Medical history:** If patient has previous cancer history hence the given BIRAD will be higher than a patient with no medical history. It was considered as a binary feature [Kerlikowske 1996, Carney 2003, Kolb 2002] [Andersson 1997, A. C. S. 2006, Cheng 1998, Markey 2002, Mehul2005, A.C.S. 2010, Karssemeijer 2004, Freer 2001].

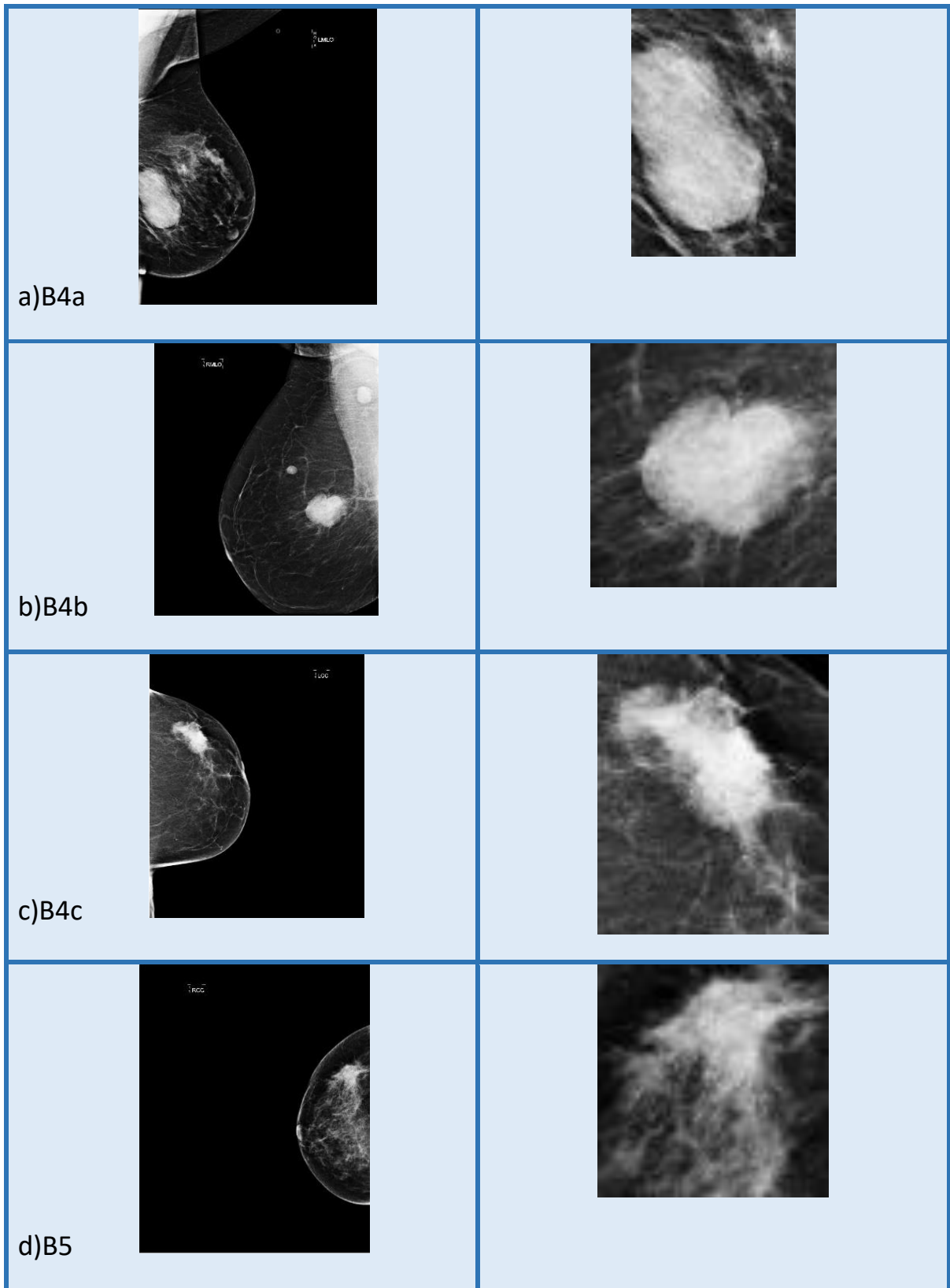


Figure28: a-Lobulated high dense mass, b- Lobulated high dense mass with micro calcifications, c- Irregular high dense ill define mass with micro calcifications, d- Irregular high dense spikey mass with micro calcifications and retraction and skin thickening

3.4 Discussion for pattern recognition

In this chapter some of most frequent feature selection methods (SFFS, KruskalWallis, MRMR and PCA) and classification techniques (Tree and regression tree, KNN, SVM, and ensemble method) and meta data were explained. From points mentioned in this chapter it could be concluded that:

Feature selection techniques: SFFS as a bidirectional method can aggregate the best feature and eliminate the worst feature, so the best possible feature subset could be selected.[Mohanty 2013, Shirbani2016] KruskalWallis, comparing to other wrapper methods, is computationally less expensive and simple to use.[Zheng 1999, Sheshadri 2007] . PCA reduce the number of selected features but it produces new features that don't have meaning of the previous features.[LindsayI Smith 2002, Rady 2011]. MRMR, comparing to other mentioned model, is better method for removing redundant and irrelevant features.[Peng 2005, Liberman 1997]

Classification techniques: Tree is based on if-the-else rule, so it is simple to interpret, fast and also new instance could be classified very fast. But they can have low predictive accuracy. [Carneiro 2015, Liu 2011] Accuracy of KNN classification could be degraded by noisy and irrelevant features. KNN has good predictive accuracy in low dimensions. They have fast fitting speed, medium prediction speed, high memory usage, and are not easy to interpret. [Hastie

2008, Liu 2014] . SVM have high predictive accuracy, medium fitting speed, and can have good prediction speed and memory usage with few support vectors. Ensemble classifiers are slower comparing to other classifiers but gives better results when there is diversity of models in that method. [Flores 2004, Beatriz 2004]

Meta data: from point mentioned in section 1.3 it could be concluded that, stability is used for diagnosing BIRAD score 2. Specularity index, retraction and existence of micro calcification are dominant meta data for giving BIRAD score over 4(4a, 4b and 4c). Therefore, it was decided to mainly focus on these meta data to subdivide BIRAD 4 to ,4a, 4b and 4c

CHAPTER 4

4. Experimental results

4.1 Introduction

Mass is categorized into 6 groups according their BIRAD score. BIRAD 0 means that the Mammography is not reliable. BIRAD score 1 means there is no abnormality found in the breast. BIRAD score of 2 means that it's a benign mass and BIRAD score 3 mean the mass has specification of a mass with BIRAD 2 but as there is no medical history of patient's mass, so it needs follow up in next six months. Also mass with BIRAD score 3 could be a lobulated, dense mass. BIRAD score 4 means that it's probably malignant mass and according to meta data specialist may diagnose BIRAD score 4a, 4b or 4c. BIRAD score 4a mean it's probably malignant, BIRAD score 4b means it has more risk of malignancy and 4c mean highly suspicious of malignancy. The mane meta data that can effect subdividing BIRAD score 4 to 4a, 4b and 4c are specularity index, micro calcification and retraction. Mass with more and higher speculation have higher BIRAD score (Fig.28-b). Existence of micro calcification also has severe effect on diagnosing. If there is micro calcification in the mass, it will have higher BIRAD score. BIRAD score of 5, is mass with BIRAD score 4 but requires more meta data which is biopsy and exact positioning of the mass. For example, highly speculated mass with retraction and micro calcifications will have BIRAD score 5 (Fig.28-d). BIRAD score 6 is given to a mass that has biopsy result indicating malignancy. Oncologist's treatment may vary for masses with BIRAD score 4a,

4b and 4c, so diagnosing BIRAD score 4a, 4b and 4c is critical issue. From point mentioned above it could be concluded that BIRAD score 2 and over 4(4a, 4b and 4c) are the most important ones, as they classify benign and malignant mass. The proposed CAD first classifies masses with BIRAD score 2 from BIRAD score over 4 and then sub classify BIRAD 4 to a, b and c.

173 masses and their diagnosing reports were gathered from Tehran, Iran. 100 of them from “Haghighat medical imaging center” and 60 from “Farokhi medical imaging center”. Among 173 masses, 80 were assigned to BIRAD score over 4 and 80 to BIRAD score 2. Of the 80 masses with BIRAD score over 4, 30 were assigned to BIRAD 4a, 30 to BIRAD 4b and 20 to BIRAD 4c. 13 masses have BIRAD score of 3.

4.2 Methodology

Diagram 2 and 3 present flow of the proposed CAD system. Diagram 4 presents evaluation of the proposed CAD system with BIRAD score 2 and over 4. Diagram 5 presents evaluation of the proposed CAD using noisy images. The following sections describe main components of the process.

4.2.1 Image enhancement

As mentioned in sections 2.1 and 2.1.1, the most frequent noise for mammograms are: Salt & pepper, Gaussian and Speckle noise which can reduce

accuracy of classification/CAD. To evaluate this, different noises were applied on small amount of masses (50 masses). Salt and pepper, Gaussian and Speckle noise, each with variance 0.0002. The CAD was evaluated by each mentioned noise.

Evaluation of CAD by masses with noise is presented in diagram 5.

As mentioned in section 2.4, it is essential to reduce any possible noise from the masses. Hence, filters mentioned in section 2.1, are applied to suppress noise and enhance the visibility of each mass. Based on results obtained from Mean square error (MSE) and PSNR functions, which are presented in table 4, step median filter for salt and pepper noise, adaptive median filter for speckle noise and Wiener filter for Gaussian noise were good candidates for noise reduction.

Mean square error (MSE)

The MSE is a form of image quality. Consider f as image with noise and f' as reconstructed image with de-noising filters, so MSE will be calculated as below:

$$MSE = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n ((f(i,j) - f'(i,j))^2)$$

Lower value of MSE shows better performance of de-noising filter.

Peak signal to ratio(PSNR) is the ratio between maximum possible power of signal and the power of noise that affects the fidelity of its representation. PSNR is calculated as below:

$$PSNR = 10 \cdot \log_{10} \frac{Max_I^2}{MSE}$$

Max_I is maximum possible pixel value of image I. Higher value of PSNR shows better performance for de-noising filter.

MSE and PSNR value for noise (Gaussian, salt & pepper and speckle) were calculate and presented in table 6.

Image with noise	Adaptive median		median		wiener	
	MSE	PSNR	MSE	PSNR	MSE	PSNR
Salt & pepper	6.693774487201984e+03	26.8321	1.150875251045171e+03	30.2301	6.780468646856455e+03	40.2101
speckle	1.028924206586984e+03	27.0252	3.021912057656508e+03	24.5310	5.526617995042068e+02	41.1143
Gaussian	9.383382607386662e+03	24.8708	3.018952789834829e+04	22.96.18	2.242419419640761e+04	42.9627

Table 6: Presenting MSE and PSNR value for three types of noises.

The results for effect of noise on classifier's accuracy before and after employing filters are presented in table 7.

	Type of noise	Classifier accuracy with original Image	Classifier accuracy by masses with noise	Classifier accuracy by masses after using noise removal filters
B4+	Salt and pepper noise	98.8%	90.0%	98.8%
B2/B4	Gaussian noise	99.4%	92.5%	97.6%
B2/B4	Speckle noise	99.4%	93.75%	98.8%

Table 7: Presents effect of noise on classifier accuracy in both levels of classification.

As it could be seen in table 7, salt and pepper noise reduced accuracy of classification B4+. Median filter removes all white spots caused by salt and pepper noise, so by using it accuracy of classifier was the same as before.

The classification accuracy for mass with BIRAD score over 4, for salt & pepper noise, was reduced to 90%. The classification accuracy for benign and malignant mass with Gaussian and speckle noise were reduced to 92.5% and 93.75% respectively. Salt and pepper noise affects accuracy of classification for mass with BIRAD score over 4. So by applying step median filter over those masses the classification accuracy was 98.8%. Gaussian and speckle noise affect accuracy of classification for masses BIRAD score 2 and over 4. So by applying Wiener and adaptive median filters the classification accuracy was 97.6% and 98.8% respectively.

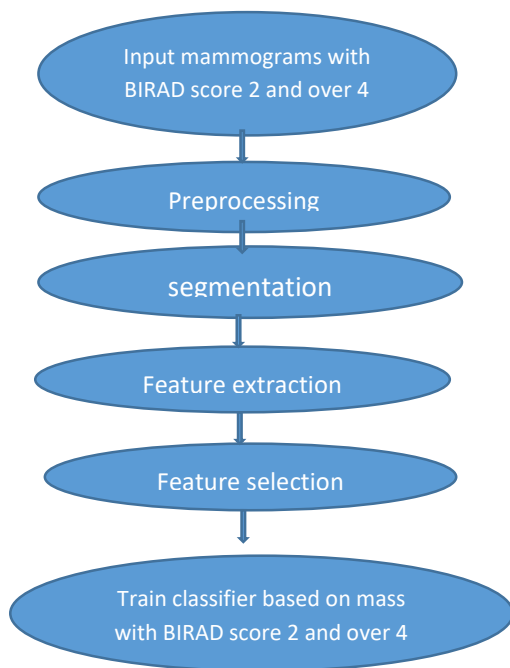


Diagram 2: Training classifier based on mass with BIRAD score 2 and over 4

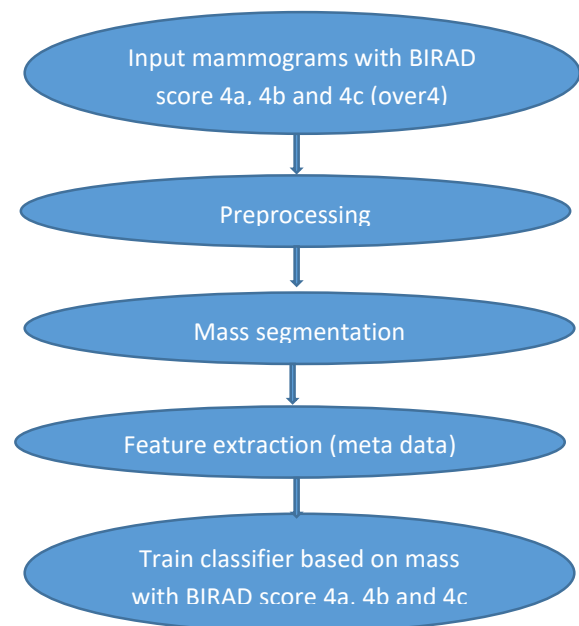


Diagram 3: Train classifier based on mass with BIRAD score over 4(4a, 4b and 4c)

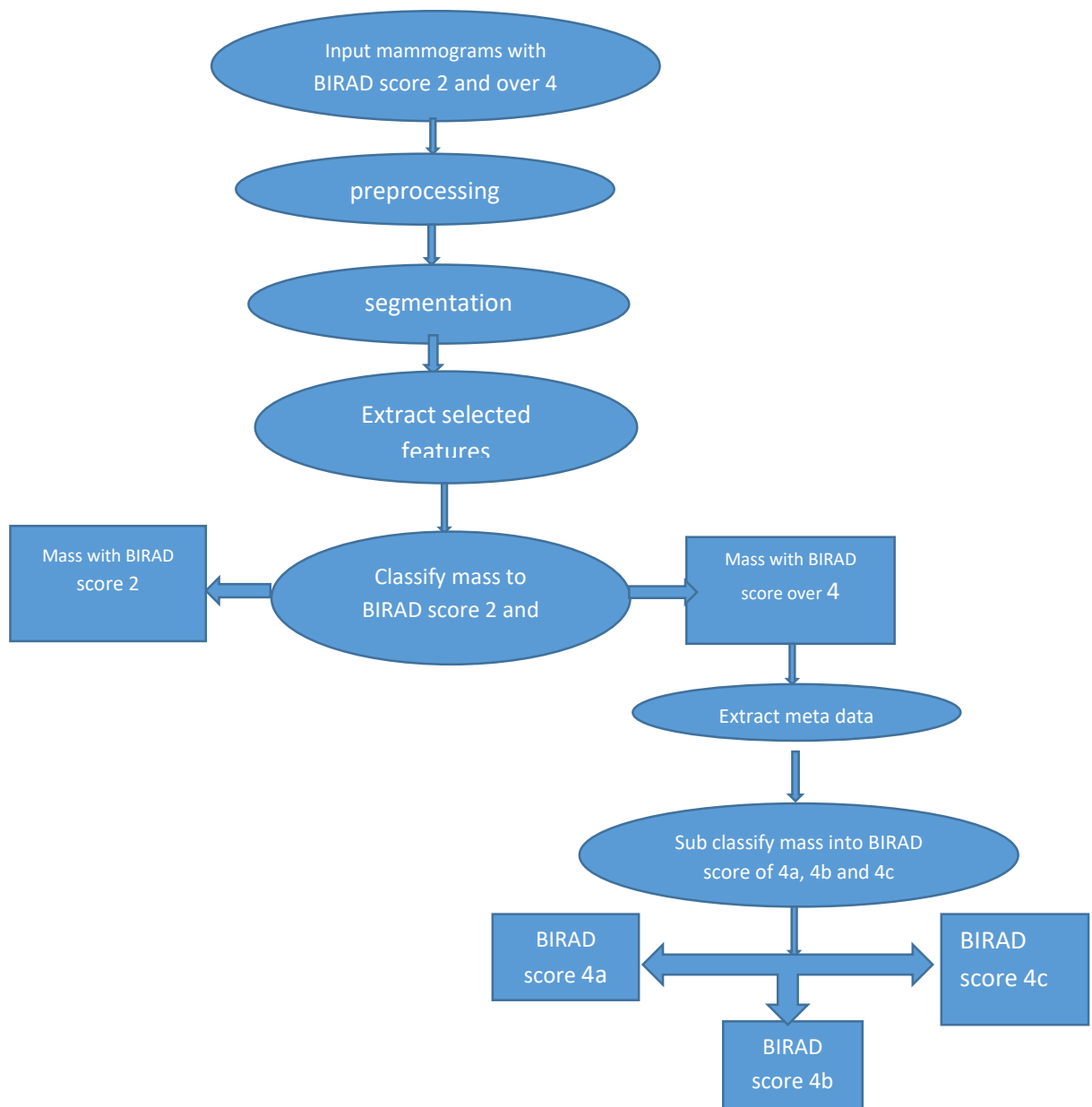


Diagram 4: Evaluation flow of the proposed CAD including a) classification of mass with BIRAD score 2 and over 4, b) Classification of mass with BIRAD score 4a, 4b and 4c. (proposed CAD is trained by masses with BIRAD score 2 and over 4 but it is also tested by mass with BIRAD score 3.)

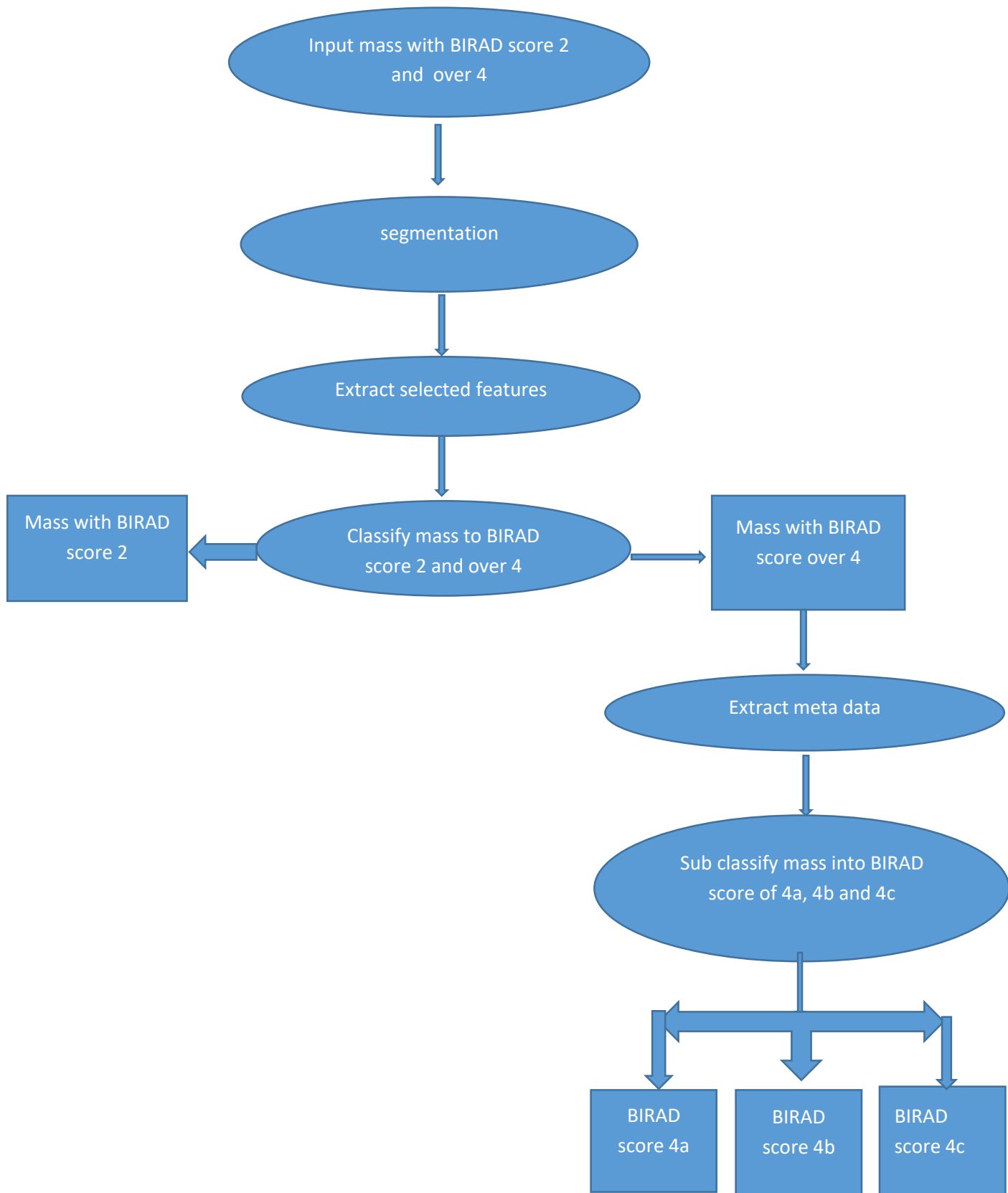


Diagram 5: Evaluation of proposed CAD using noisy images

4.2.2 Segmentation results

Primary mass annotations are done by specialist, drawing a circle around the mass which was printed on a paper report.

According to the point mentioned in 2.2.4, Chan-Vese active contour was employed to detect outline of 173 ROIs. The results are approved by 3 senior specialists.

Chan-Vese is a non-biased active contour so it will shrink or expand a free hand initialization, drawn by specialist, according to image features. Up to 300 iterations was applied on the initialization to finalize the segmentation result. Algorithm will be stopped If the contour position of the iteration is the same as the last five iterations. Afterward any possible hole in the ROI was filled.

Figure29-b present first initialization (rectangle in blue) of lobulated mass and its Chan-Vese segmentation result (curve in red). Although the initialization wasn't accurate still it was able to segment the ROI accurately. Initialization and segmentation of mass with BIRAD score 2 is presented in (Fig.29-a), as it could be seen Chan-Vese successfully detected outline of ROI margin precisely.

In (Fig.29-c, d and e) initialization and segmentation of speculated mass are also presented, as it could be seen Chan-Vese segmentation segmented speculations mass precisely.

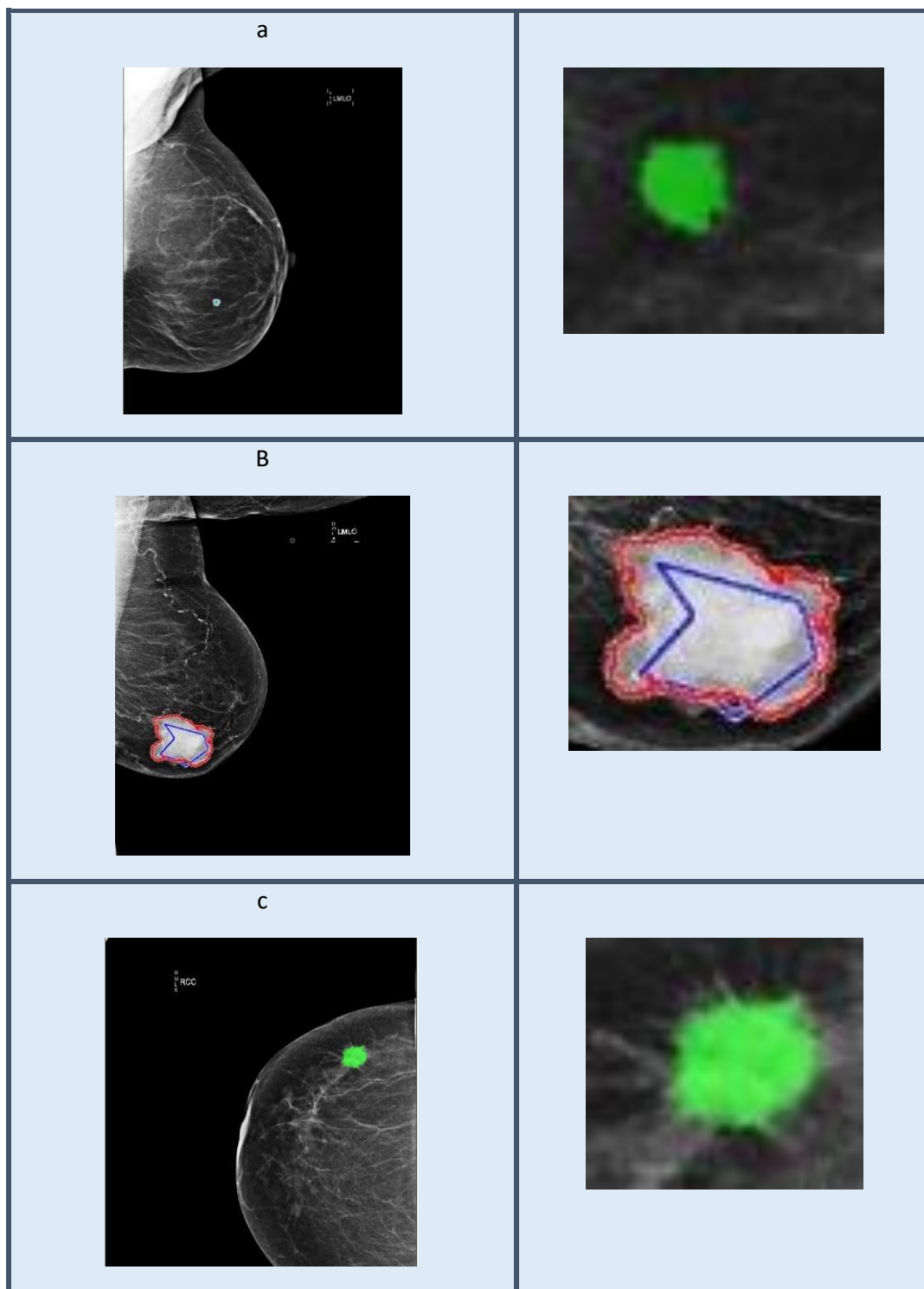


Figure 29: a- Initialization and segmentation of mass with BIRAD score 2, b- Poor initialization for lobulated mass with BIRAD score 4a, but precise Chan-Vese active contour segmentation, c- Initialization for spiky mass with BIRAD score 4a and its segmented ROI, d-Initialization for spiky mass with BIRAD score 4b and its segmented ROI, e- Initialization for spiky mass with BIRAD score 4c and its segmented ROI.

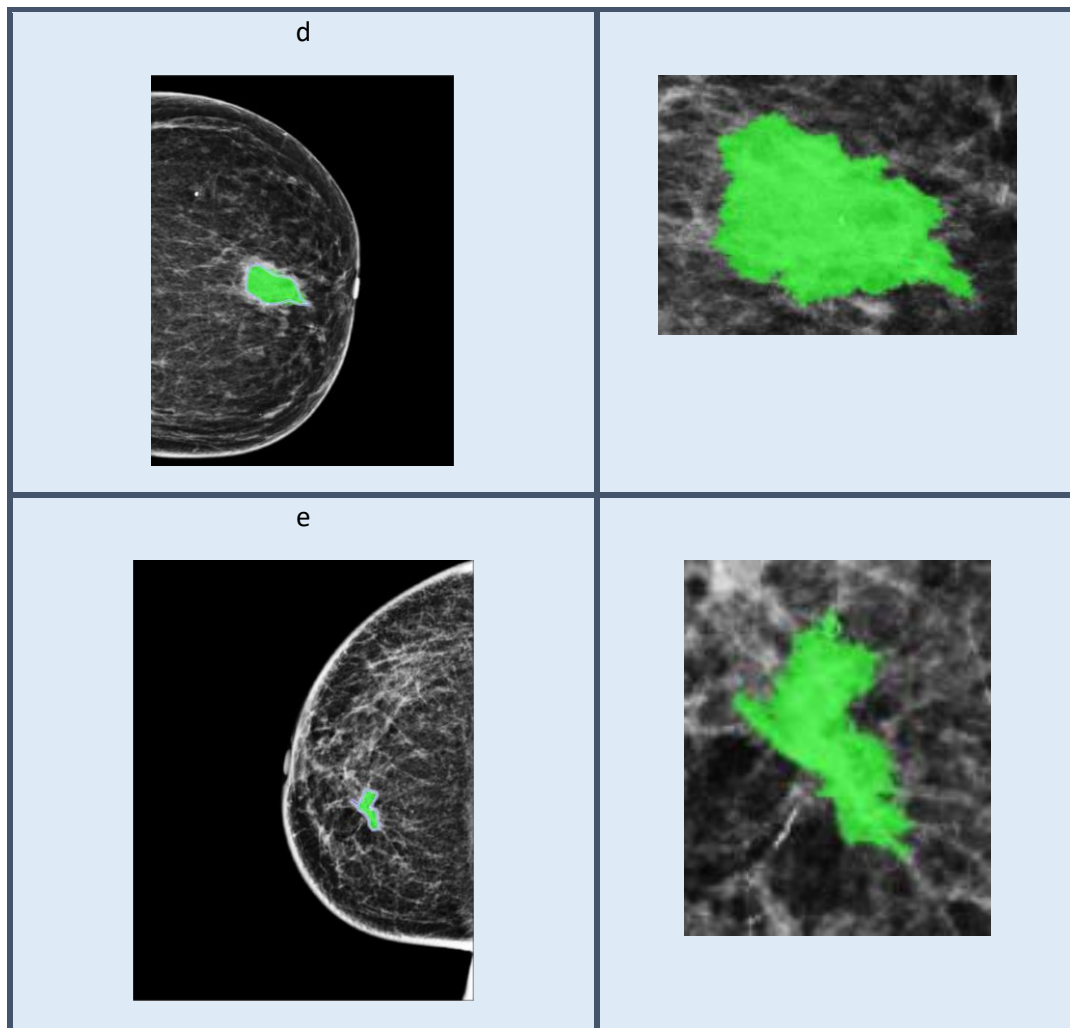


Figure 29: a- Initialization and segmentation of mass with BIRAD score 2, b- Poor initialization for lobulated mass with BIRAD score 4a, but precise Chan-Vese active contour segmentation, c- Initialization for spiky mass with BIRAD score 4a and its segmented ROI, d-Initialization for spiky mass with BIRAD score 4b and its segmented ROI, e- Initialization for spiky mass with BIRAD score 4c and its segmented ROI.

4.2.2.1 Segmentation results and its Reproducibility

All 173 masses were automatically annotated using Chan-Vese active contour segmentation method. These annotations also were approved by 3 senior specialists based on their visual judgement. From those 173 mass, a sample of 50 of the masses were 3 time automatically annotated using different

initialization free hand input (contour). The results were compared (by Jaccard index) and there was a trivial difference in the results. Figure 30-b presents result for reproducibility of automatic segmentation

Jaccard Index:

Overlap area of two segmentations were computed using Jaccard index which is defined as below:

$|A \cap B| = \text{size of the intersection } A \text{ and } B$

$|A| = \text{size of } A$

$\text{Overlap } (A, B) = |A \cap B| / \min(|A|, |B|)$

Moreover, the difference in three initialization has no effect on the segmentation and final CAD results. The average accuracy was 99.5%. Also 50 of masses were manually annotated using freehand in Matlab. Those masses also were automatically annotated using Chan-Vese active contour in Matlab. The results were compared by Jaccard index. The accuracy was 85.9%. Figure 30-a presents manual annotation in blue and automatic annotation in red.

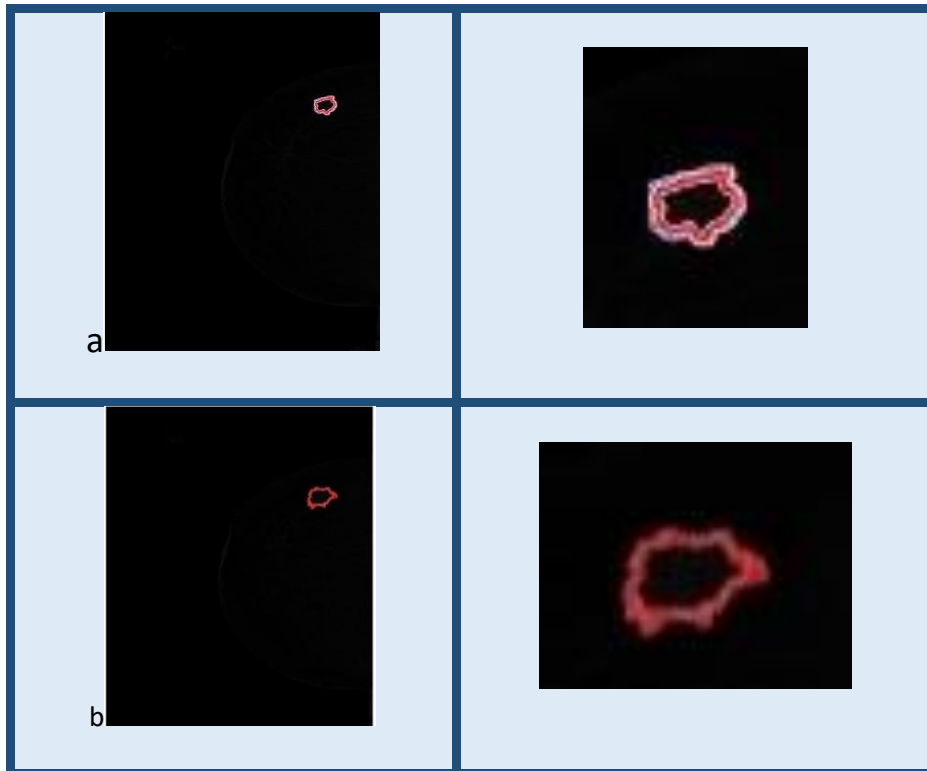


Figure 30: a-Comparing automatically segmentation ROI (red curve) with manually segmentation ROI (blue curve), b-Comparing two automatic segmented ROI

4.2.3 Feature extraction

According to points mentioned in section 2.3, there are 5 types of medical features. ROI intensity, Margin intensity, ROI shape, Margin shape and ROI texture.

Beside those features, some mathematical features also were considered in order to fully characterize a mass. Those features are presented in table 4. Totally, 68 intensity based features, shape based features and 22 texture features are extracted. Their results were compared and the most redundant and irrelevant ones were deleted manually.

Figure 31-a presents mean as a discriminate feature, as it could approximately classify BIRAD score 2 and over 4. Figure 31-b, c present orientation and CORRP Haralick feature as irrelevant feature, as they cannot distinguish between BIRAD score 2 and over 4.

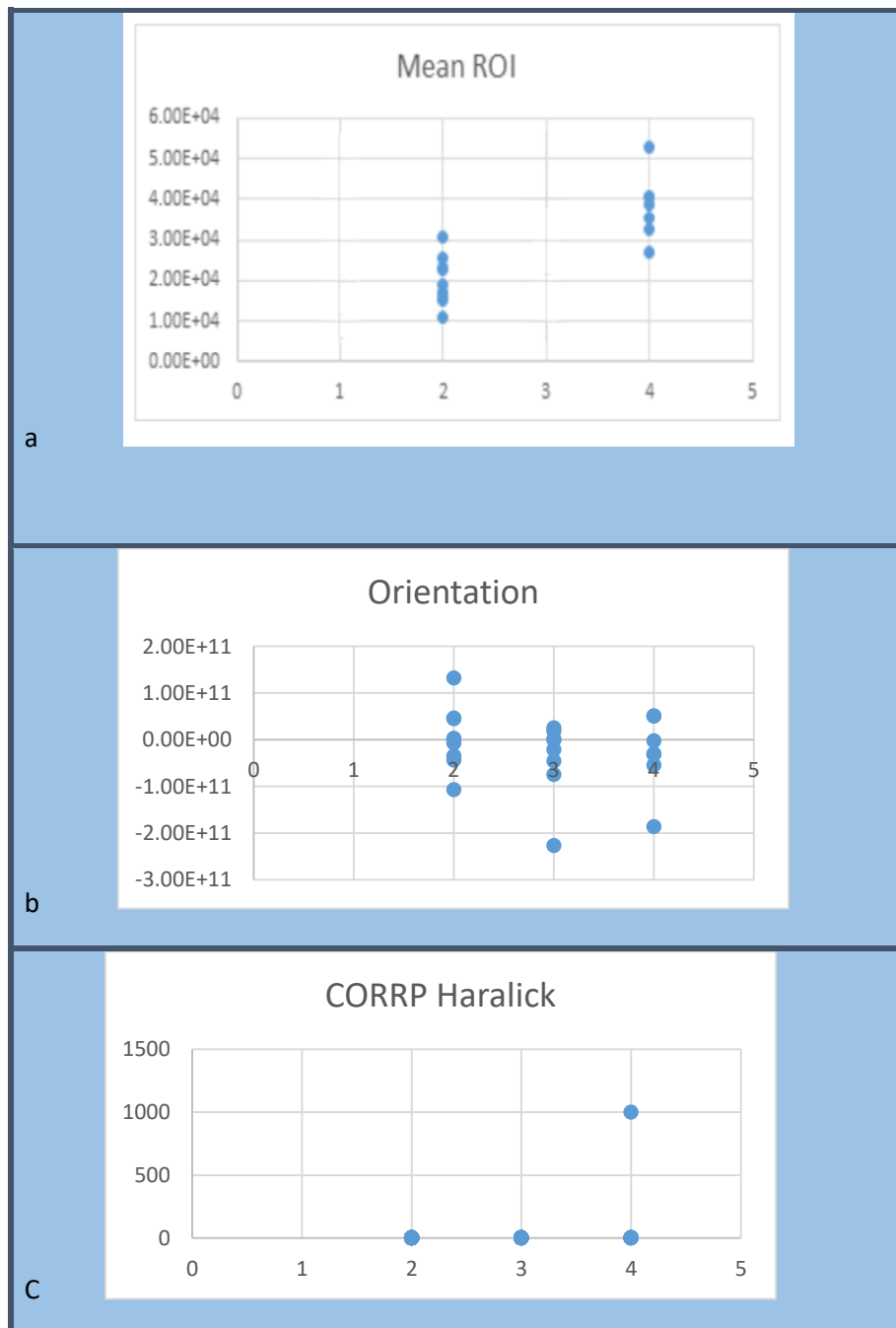


Figure 31: a- Mean ROI as a relevant feature, b- Orientation as an irrelevant feature, c- Haralick as an irrelevant feature

As a result of such manual analysis, a set of 35 medical and mathematical features were extracted from each mass, which are presented in table 4.

Features were normalized as below:

$$f'_i = \frac{f_i - \min(f_i)}{\max(f_i) - \min(f_i)}$$

Min, Max and Average of normalized value for each features in each class are presented in table 8.

Feature value								
BIRAD score		compactness	acutance	contrast	Energy	homogeneity	kurtosis	Max STD-max STD surrounding area
B2	min	-0.07375	-0.02537	-0.007974967	-0.38764	-0.26766	-0.16178	-0.42193
	max	0.065314	0.20092	0.992020222	0.082489	0.124526	0.757038	0.496893
	average	-0.02302	-0.01439	0.008694828	0.028144	0.03306	-0.00431	-0.00603
B4	min	-0.0214	-0.0254	-0.007979901	-14.4529	-2738.37	-0.4227	-0.25343
	max	0.962699	0.974601	-0.007629083	0.077046	0.153018	0.421472	0.578069
	average	0.15605	0.085873	-0.007864923	-0.91701	-144.307	0.00073	0.003194

Table 8: Min, max and average value for 35 extracted features

Feature value									
BIRAD score		mean global1	mean ROI	diff moment2	STD sur-STD ROI	diff smoothness	solidity	STD - mean	diff variance
B2	min	-0.2939	-0.4267	-0.21187	-0.49672	-0.08432	-0.09847	-0.59101	-0.45245
	max	0.619449	0.534627	0.527862	0.422105	0.204151	0.135749	0.376279	0.274511
	average	0.021549	0.053634	0.001475	0.006202	0.008721	0.078616	-0.04456	-0.00751
B4	min	-0.47242	-0.43814	-0.1858	-0.57789	-0.08154	-2.02653	-0.62238	-0.17128
	max	0.019675	0.573299	0.284068	0.253609	0.05843	0.116572	0.386629	0.547554
	average	-0.16944	0.038418	-0.00244	-0.00302	0.012253	-0.44216	-0.04302	0.004648

Table 8: Min, max and average value for 35 extracted features

Feature value		sum Hist ROI	uniformity	contrast sur	diff Hist	AUTOC	CONTR	CORRM	CPRM
B2	min	-0.0688	-0.00782	-0.49093	-0.96946	-0.07575	-0.02478	-0.08673	-0.11214
	max	0.345847	-0.00781	0.509075	0.011524	0.384321	0.115597	0.027466	0.654109
	average	-0.01683	-0.00782	0.08711	-0.0087	-0.02114	-0.02182	0.003165	-0.02265
B4	min	-0.0688	-0.00782	-0.57574	-0.01245	-0.52524	-0.02488	-0.97253	-0.11342
	max	0.931598	0.992185	0.366564	0.030237	0.924251	0.97512	0.027466	0.88786
	average	0.046867	0.044816	-0.07396	0.0064	0.10654	0.133654	-0.03595	0.127771

Table 8: Min, max and average value for 35 extracted features

Feature value		CSHAD	DISSI	ENTRO	MAXPR	SOSVH	SAVGH	SVARH	SENTH
B2	min	-0.08349	-0.01979	-0.10691	-0.39785	-0.0787	-0.01987	-0.00781	-0.09855
	max	0.404978	0.365387	0.444909	0.076756	0.382958	0.052826	-0.00781	0.439518
	average	-0.02382	-0.00886	-0.02343	0.021218	-0.02402	-0.01056	-0.00768	-0.01985
B4	min	-0.08501	-0.01979	-0.11138	-28.8596	-0.51602	-0.65523	-0.00782	-0.10303
	max	0.916509	0.980214	0.893088	0.073862	0.921302	0.980134	0.992185	0.877777
	average	0.14557	0.064729	0.149715	-1.65649	0.122798	0.030586	0.044819	0.127392

Table 8: Min, max and average value for 35 extracted features

Feature value		DVARH	DENTH	INF1H	INF2H	
B2	min		-0.01438	-0.00036	-0.13694	-0.02384
	max		-1E-06	0.018032	0.815563	-1.3E-05
	average		-0.0141	0.007783	-0.01846	-0.02312
B4	min		-0.01438	0.007035	-0.16795	-0.02384
	max		0.246352	0.018028	0.633627	0.976159
	average		0.022927	0.007879	0.105746	0.13513

Table 8: Min, max and average value for 35 extracted features

As mentioned in section 1.2.6, benign mass is round or oval, dense, well defined, homogenous with smooth border. From feature values presented in table 8, these points could be concluded:

Shape of mass is defined by compactness. Benign mass is more compact comparing with malignant one, as it is round or oval. As the value of compactness was computed as $\frac{1}{\text{perimeter}}$ so malignant mass has higher compactness value comparing to benign mass.

Considering texture, benign mass is homogenous while malignant mass is heterogeneous.

For evaluating mass as a well-defined mass or ill-defined mass, acutance and contrast were extracted.

Acutance is the difference of mean value in ROI and mean value its surrounding area. Benign mass mean value was closer to 1 and malignant mass mean value was closer to 0, so benign mass acutance was a negative value while malignant mass will had more positive value.

Contrast is the ratio of contrast of two areas, ROI and its surrounding area. Benign mass had higher contrast value comparing to malignant mass.

As for density, benign mass also is denser comparing malignant mass. Benign mass value for mean, STD goes toward 1 while malignant mass values go toward 0. Also benign mass comparing to malignant mass had higher value for diff moment 2, STD ROI-STD surrounding area, kurtosis, max STD ROI-max STD surrounding area, solidity, maximum probability, variance and mean global.

Feature selection methods were used to identify the most distinguishing features. These features and their weights were used to classify each mass.

4.2.4 Feature selection

From each feature selection method (filter, wrapper and embedded) one technique was chosen and implemented such as: SFFS, MRMR, KruskalWallis, PCA. As MRMR remove redundant and irrelevant ones so some mixed methods also were tested such as: SFFS+ MRMR and KruskalWallis+ MRMR. 35 extracted features and their BIRAD were used as inputs for feature selection methods.

Feature selection methods return a weight for each feature which was used as a base to select a feature. Weighted feature was calculated as below:

$$f_i * \frac{w_i}{\sum_{i=1}^n w_i}$$

Where f_i is selected feature

w_i is weight of f_i

N is number of selected features

According to weight of each features their impact is calculated as below:

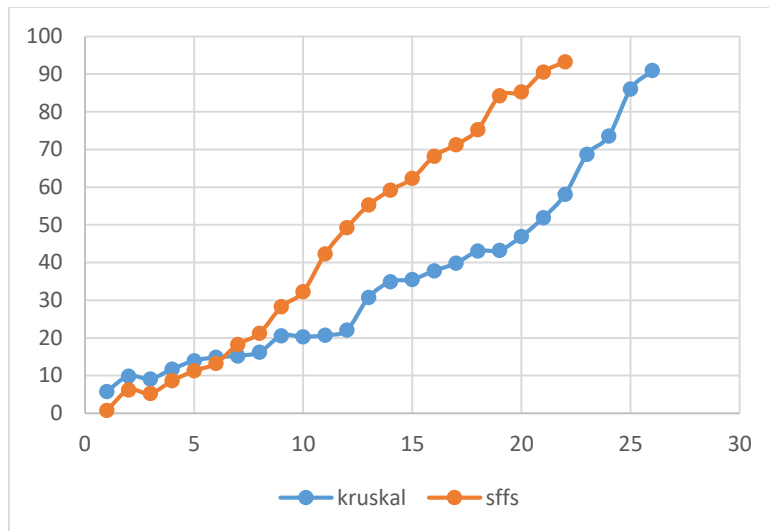
$$\frac{w_1}{\sum_{i=1}^n w_i}, \frac{w_1+w_2}{\sum_{i=1}^n w_i}, \frac{w_1+w_2+w_3}{\sum_{i=1}^n w_i}, \dots$$

In this thesis it was decided to choose number of features that have more than 90% of the values. Table 9 presents effect of number of features on the values.

- **MRMR feature selection** method give rank to each 35 of features. The features according to their priority are: solidity, Entropy, compactness, SOSVH, acutance, SVARH, mean global, CPROM, contrast, SENTH, energy, CORRM, uniformity, Diff variance, Diff HIST, mean ROI, CONTR, DISSI, Sum Hist, CSHAD, SVAGH, homogeneity, DVARH, STD-mean, max STD ROI-max STD surrounding area, contrast, AUTO, DENTH, Diff smoothness, STD ROI-STD surrounding area, MAX PR, kurtosis, Diff moment2.
- **KruskalWallis feature selection** method identified that 26 features accounted for 90.9% of the values. The features according their priority are: Compactness, CORMM, STD-Mean, CSHAD, SENTH, INF2H, SOSVH, entropy, SAVGH, AUTO, CPROM, Contrast, Sum Histogram ROI, Diff moment2, DVARH, Kurtosis, Diff Smoothness, Acutance, Energy, Homogeneity, Mean Global, Mean ROI, (STD surrounding area – STD ROI), solidity, SVARH and Max STD ROI- Max STD Surrounding area.
- **Kruskal Wallis+ MRMR:** by applying MRMR on 26 features selected by Kruskal Wallis, features were prioritized as follow: Compactness, energy, SVARH, STD –Mean, SOSVH, diff smoothness, mean global, contrast, CPRROM, solidity, homogeneity, STD ROI-STD Surrounding area, Max STD ROI- Max STD surrounding are, diff Hist, acutance, Diff variance, Diff moment2, kurtosis, mean ROI, contrast, Sum histogram ROI.

- **Sequential forward floating feature (SFFS)** selection identified 22 features that accounted for 93.23% of the values. Those features according their priority are: Solidity, Acutance, CPROM, Compactness, STD-Mean, Mean ROI, SVARH, CSHAD, Entropy, Uniformity, SENTH, SOSVH, Sum Histogram ROI, AUTOC, (Max STD ROI - Max STD Surrounding area), (STD surrounding area -STD ROI), DVARH, Difference entropy, INF2H, Contrast, CORMM, Diff variance and Diff Moment2.
- **SFFS+MRMR** applying MRMR on SFFS selected features, 22 features according to their priority are ranked as follow: Solidity, entropy, SOSVH, compactness, acutance, SVARH, mean global, CPROM, contrast, SENTH, energy, Diff variance, CORRM, uniformity, Diff Hist, mean ROI, CONTR, DISSI, sum Hist, CSHAD, DVARH, homogeneity, STD –mean, Max STD ROI- Max STD Surrounding area, DENTH.
- Finally, **PCA** needed only the 4 most significant eigenvalues to account for 93.32% of the information.

Influence on value



Number of features

Table 9: SFFS and KruskalWallis number of weighted features and their impact on value

4.2.5 Classification of benign and malignant mass (BIRAD 2 and over 4)

Each of the subsets (selected features) were evaluated by different binary classifiers with various settings, which are mentioned in section 3.1. For experimental results, all different Feature subsets were incorporated to different tree classifiers (complex tree, simple tree and medium tree) with different number of splits and split criterions (Ginis diversity index, towing rule, max deviance reduction). Feature subsets were incorporated to different SVMs (linear, quadratic, cubic and Gaussian) with different Kernel functions such as: cubic, linear, quadratic and Gaussian. KNN classifiers (coarse, cosine, cubic and weighted) with different number of neighbors and distance metrics (Euclidean, city block, cubic, cosine, correlation and spearman) also were tested. Finally, some ensemble classifiers were evaluated this includes: boosted tree, bagged

tree, subspace discriminant and subspace KNN. These were tested with several ensemble methods such as: bag, Ada boost, Rus-boost and subspace using different learner types, number of learners and learning rates.

The CAD had been evaluated using two different schemes: Cross validation and hold out.

Cross validation method: In this system, data were partitioned into 5 fold, randomly. Each fold was hold out in turn for testing. The system uses other folds to train the system. Accuracy was assessed by average performance of test sets. (it require multiple fits but makes efficient use of all).

Hold out method: 20% of data is selected randomly (data from BIRAD score of 2 and over 4). They are used as test set and then the system is trained with the remaining data. In hold out method accuracy was assessed with performance of test data. (recommended for large data sets). Best result from each classifier using different feature subsets are presented in table 10. (Scheme #1 is hold out method and scheme#2 is cross validation method results.)

The best accuracy in this scheme using appropriate classifier and feature selection schemes, was 99.4%. while the average accuracy was 98.62% when cross validation was run for 10 times. Average accuracy when cross validation was run 10 times was 99.76% and best accuracy was 100%.

Feature selection	Classifier	Scheme #	Accuracy	
KruskalWallis	bag tree ensemble method: bag # of learner: 100	1	100	
		2	99.4	
	Complex tree Max split=20 Split criterion= Gini	1	100	
		2	96.9	
	Quadric SVM Kernel function=Gaussian	1	97	
		2	95	
	KNN #of neighbors=4 Distance Metric=cubic	1	52	
		2	50.6	
	MRMR	Subspace KNN Ensemble method=bag #of learners=200	1	92
			2	91.3
Complex tree Max #split=10 Split criterion=Gini		1	96.5	
		2	98.1	
KNN #of neighbor=20 Distance Metric=quadric		1	78.6	
		2	50.9	
SVM Kernel function=Gaussian		1	76	
		2	75	
SFFS		Complex tree Max #split=10 Split criterion=Gini	1	98.1
			2	96.9
	Boosted tree Ensemble method=Ada-boost #of learners=100	1	100	
		2	99.4	
	SVM Kernel function=Gaussian	1	98.4	
		2	98.1	

Table 10: Selected weighted features are sent to different classification methods. In each method best result is presented in this table. Accuracy is presented in two methods. 1-Hold out method, 2- Cross validation method.

	KNN #of neighbor=20 Distance Metrix=quadric	1	98
		2	91.3
PCA	Boosted tree Ensemble method=Ada boost #of learners=100	1	70.1
		2	50.5
	KNN #of neighbor=10 Distance Metrix=quadric	1	70.5
		2	50
	SVM Kernel function=Gaussian	1	99.1
		2	97.5
	Complex tree Max # split=100 Split criterion=Ginis	1	97.4
		2	90
KruskalWallis+ MRMR	boosted tree ensemble method: Ada boost # of learner: 100 Learning rate=0.1	1	100
		2	99.4
	Complex tree Max split=20 Split criterion= Ginis	1	100
		2	96.9
	Quadric SVM Kernel function=Gaussian	1	97
		2	95
	KNN #of neighbors=4 Distance Metrix=cubic	1	52
		2	50.1
SFFS + MRMR	Complex tree Max #split=10 Split criterion=Ginis	1	98
		2	96
	Boosted tree Ensemble method=Ada-boost #of learners=100	1	100
		2	99.4
	SVM Kernel function=Gaussian	1	98.5
		2	98
	KNN #of neighbor=20 Distance Metrix=quadric	1	91.9
		2	98.1

Table 10: Selected weighted features are send to different classification methods. In each method best result is presented in this table. Accuracy is presented in two methods.1-Hold out method, 2- Cross validation method.

The classifications that have higher accuracy were checked with different number of learners and results are presented in table 11.

Feature section method	Classification	Ensemble method	Decision tree. #of learners	results
Kruskal Wallis	Bag	bag	15	96.3
			25	96.3
			50	98.8
			100	99.4
			200	99.4
SFFS	Boosted tree	Ada boost	15	98.1
			25	98.1
			50	93.1
			100	99.4
			200	99.4
Kruskal Wallis+ MRMR	Boosted tree	Ada boost	15	98.1
			25	98.1
			50	98.1
			100	99.4
			200	99.4
SFFS+MRMR	Boosted tree	Ada boost	15	96.3
			25	96.3
			50	96.3
			100	99.4
			200	99.4

Table 11: Classifications that have higher accuracy are checked with different number learners in order to see how it may affect the classification accuracy.

Classifications with higher accuracy each were run 10 time and average accuracy for each of them is presented in table 12.

Feature selection	Classification	# of learners	minimum	maximum	average
Kruskal Wallis	Bag/bag	100	97.5	99.4	98.26
SFFS	Boosted tree/ada boost	100	98.8	99.4	98.62
Kruskal Wallis+ MRMR	Boosted tree/ada boost	100	95.6	99.4	96.74
SFFS+MRMR	Boosted tree/ada boost	100	98.1	99.4	98.62

Table 12: Each classification is run for 10 times. Minimum, maximum and average accuracy of each of them is calculated.

Confusion matrix and ROC curve of SFFS selected features followed by boosted tree classifier with ensemble method Ada boost, Decision tree, 100 learners as the best classifier is presented in figure 32.

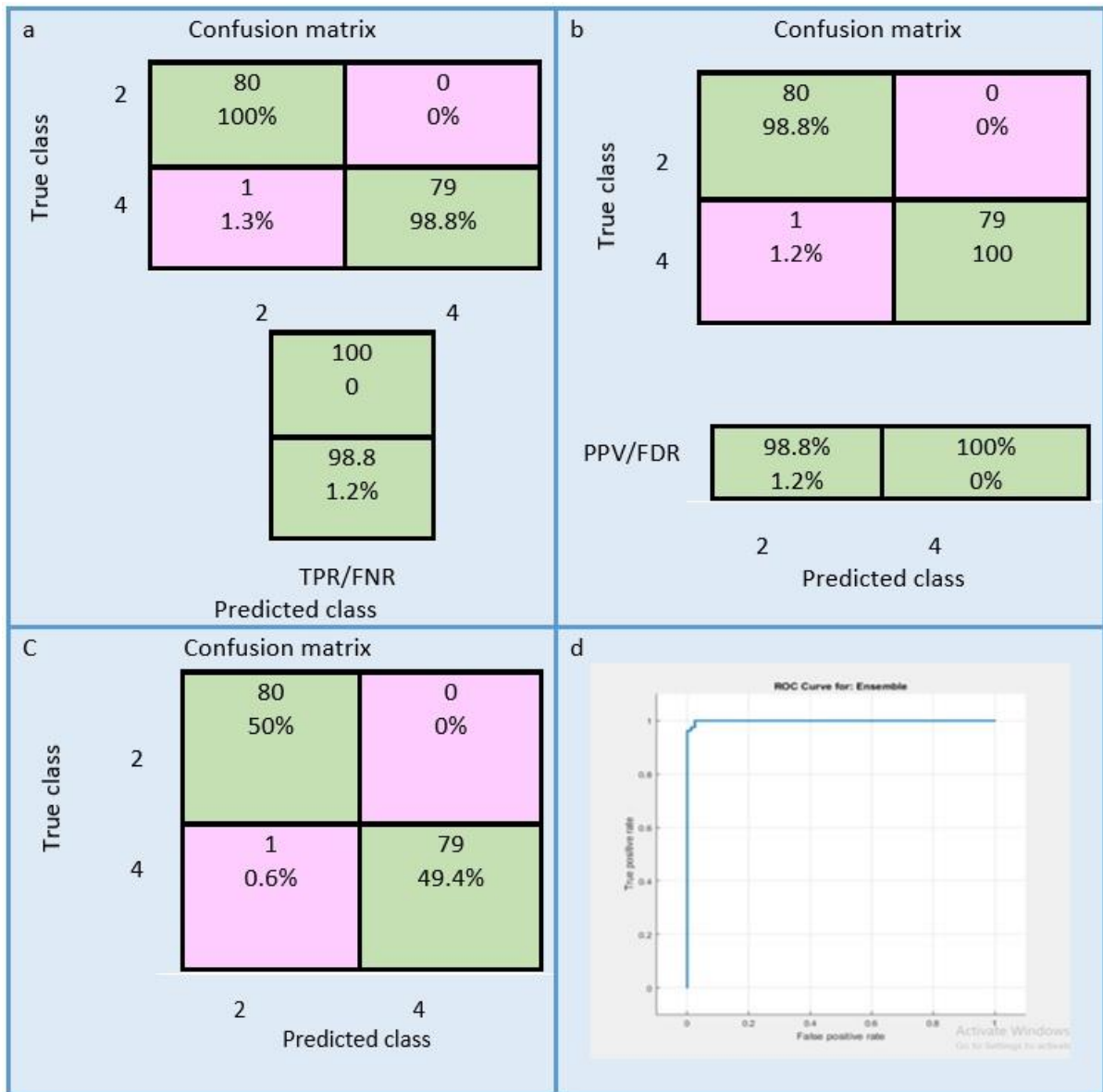


Figure 32: Confusion matrix and ROC curve for SFFS selected features with boosted tree classifier with ensemble method Ada boost, Decision tree, 200 learners and learning rate 0.1

Figure 32-a, presents per true class of confusion matrix. It shows that one malignant mass was miss classified. So it was 1.3% false positive and 98.8% true positive and 100% true negative. Figure 32-b presents per predicted class confusion matrix and it showed 1.2% of benign mass were false negative. Figure 32-c, presents over all Confucian matrix for the mentioned classifier. It reveals that one of the malignant masses was miss classified.

It was also tried to evaluate mentioned classification algorithms without using feature selection algorithms. So 35 extracted features were sent to classification algorithms. In order to show how changes in setting of each method affect the classification accuracy, in each algorithm different settings for each method of algorithm was tested and results are presented in Table 13.

Classification Tree	method	# of split	Cross validation results
Tree	simple tree	2	96.9
		4	96.9
		6	97.5
	Medium tree	2	96.9
		4	97.5
		6	97.5
	complex tree	2	96.9
		4	97.5
		6	97.5

Table 13: Accuracy results for different classification algorithms and different ensemble methods. (All extracted features are fed to classifiers)

Classification SVM	method	Kernel function	results
SVM	Linear	Linear	95.0
		Gaussian	95.6
		Quadric	96.9
		Cubic	96.9
	Fine Gaussian	Linear	96.3
		Gaussian	83.8
		Quadric	95.6
		Cubic	89.4
	Medium Gaussian	Linear	95.6
		Gaussian	95.0
		Quadric	96.9
		Cubic	96.3
	Coarse Gaussian	Linear	90.6
		Gaussian	93.8
		Quadric	94.4
		Cubic	94.4
	Quadric	Linear	95.0
		Gaussian	95.6
		Quadric	96.9
		Cubic	96.9
Cubic	Linear	95.0	
	Gaussian	98.6	
	Quadric	96.9	
	Cubic	96.9	

Table 13: Accuracy results for different classification algorithms and different ensemble methods.
(All extracted features are fed to classifiers)

KNN classification	#Of neighbors	Distance metric	Results
Fine	2	Euclidean	92.5
		City box	95.5
		Correlation	93.8
		Jaccard	98.8
Medium	2	Euclidean	89.8
		City box	94.4
		Correlation	92.5
		Jaccard	98.8
Coarse	2	Euclidean	92.5
		City box	95.6
		Correlation	93.8
		Jaccard	98.8
Weighted	2	Euclidean	93.1
		City box	95.6
		Correlation	95.6
		Jaccard	98.8

Table 13: Accuracy results for different classification algorithms and different ensemble methods. (All extracted features are fed to classifiers)

Ensemble classification	Ensemble method	Decision tree # of learners	Results
Boosted tree	Ada boost	15	98.1
		25	98.1
		50	98.8
		100	98.8
		200	99.4
	Bag	15	98.8
		25	98.8
		50	98.8
		100	99.4
		200	100
	Rusboost	15	93.1
		25	93.1
		50	93.1
		100	93.1
		200	93.1
Bag Tree	Bag	15	100
		25	98.8
		50	98.8
		100	98.8
		200	99.4
	Ada boost	15	98.1
		25	98.1
		50	98.8
		100	99.4
		200	99.4

Table 13: Accuracy results for different classification algorithms and different ensemble methods. (All extracted features are fed to classifiers)

Subspace Discriminant	Subspace	15	96.3
		25	96.3
		50	96.3
		100	96.3
		200	96.3
	Bag	15	98.8
		25	98.8
		50	98.1
		100	98.8
		200	98.8
Subspace KNN	Subspace	15	98.1
		25	98.8
		50	98.8
		100	96.3
		200	96.3
	Bag	15	98.8
		25	98.1
		50	98.1
		100	98.8
		200	98.8
	Ada boost	15	98.8
		25	98.8
		50	98.8
		100	98.8
		200	99.4

Table 13: Accuracy results for different classification algorithms and different ensemble methods. (All extracted features are fed to classifiers)

Rusboost	Rusboost	15	93.1
		25	93.1
		50	93.1
		100	93.1
		200	93.1
	Bag	15	98.8
		25	99.4
		50	99.4
		100	98.8
		200	98.8
	Ada boost	15	98.1
		25	98.1
		50	98.1
		100	98.1
		200	99.4

Table 13: Accuracy results for different classification algorithms and different ensemble methods. (All extracted features are fed to classifiers)

As its presented in table 10, bag tree with ensemble method bag and 15 learners had the highest ccuracy,100%. Some other classifications had 99.4% accuracy. Classifications with higher accuracy were run 10 times and their average accuracy is presented in table 14.

	Classification/ensemble method	#of learners	Min accuracy	Max accuracy	Average accuracy
1	Boosted tree/Ada boost	200	98.1	99.4	98.75
2	Boosted tree/bag	100	88.8	99.4	95.16
3	Boosted tree/bag	200	87.5	100	89.14
4	Subspace KNN/Ada boost	200	93.1	99.4	96.25
5	Rusboost/bag	25	93.1	99.4	96.25
6	Bag tree/bag	15	87.5	100	91.68
7	Bag tree/ bag	200	87.5	99.4	92.26
8	Bag tree/ Ada boost	100	88.8	99.4	94.1
9	Rusboost/ Ada boost	200	93.1	99.4	97.51

Table 14: Classification algorithms with higher accuracy each are run 10 times and their average accuracy are presented.

Confusion matrix for best classification algorithm is presented in figure 33.



Figure 33: Presents confusion matrix for boosted tree classification with Ada boost ensemble method and 200 learners.

As presented in figure 33, the overall result of the proposed CAD was 1.3% false positive and 98.8% true positive and 100% true negative. In other words, the

proposed CAD using the above specification was 98.8% leading to missing only one mass.

4.2.6 CAD results on BIRAD score 3

In cases that specialist gives BIRAD score of 3 to a mass, patient will be advised to do another modality and also to follow up in next 6 months. Stability is a key issue in diagnosing BIRAD 3 and to make decision to change the score either to BIRAD 2 or BIRAD 4. If patient, with mass diagnosed as BIRAD score of 3, has previous mammogram and the same mass exists before and there is no change in shape, density or size of the mass, so that mass will be considered stable and will have BIRAD score of 2.

Although the CAD system was trained with masses with BIRAD score 2 and over 4, it was tested with 13 masses with BIRAD score of 3 and the results were the same as the specialist's follow up obtained after six months follow up report. CAD evaluation by mass with BIRAD score of 3 is presented in diagram 4.

Up to now, the proposed CAD was evaluated for classifying benign and malignant mass (mass with BIRAD score 2 and over 4). Oncologist's treatment may vary for mass with BIRAD score 4a, 4b and 4c. So it's critical to differentiate between these masses too. To do this, meta data defined by specialist should be extracted, which is explained in next section.

4.2.7 Meta data for sub classification of BIRAD score 4

In section 1.8 it is mentioned that the proposed CAD classifies mass in two levels: First it classifies mass with BIRAD score of 2 and over 4 as benign or malignant mass, then mass with BIRAD score over 4 was sub classified as mass with BIRAD score 4a, 4b and 4c. For this, meta data should be extracted from 80 masses with BIRAD score over 4 in order to characterize them. As mentioned in section 3.3 and 4.1, meta data plays important role in diagnosing procedure, especially for diagnosing masses with BIRAD score over 4.

A BIRAD score of four (4a,4b and 4c) means that it is suspiciously a malignant mass. BIRAD score 4a is a mass with irregular shape, lobulated or a little spikey with ill-defined margin. Mass with BIRAD score 4b means a mass with more speculated margin comparing with mass with BIRAD score 4a. Mass with BIRAD score 4c has higher level of speculation comparing mass with BIRAD score 4b and also it may have micro calcification or it may be close to nipple, there may also be dermal or facial retraction. Mass with a little speculation could be considered as mass with BIRAD score 4a but when there are micro calcifications in that mass it will be diagnosed as mass with BIRAD score 4c.

Well defined Lobulated mass could be diagnosed as mass with BIRAD score 3, but when there are micro calcifications in it or it causes retraction on the tissue, so it will be diagnosed as mass with BIRAD score 4c.

So micro classification, specularity index (margin shape) and retraction are important meta data that can help the CAD to classifies mass. Extraction of micro calcification and specularity index are described in section 4.2.7.1 and 4.2.7.2.

4.2.7.1 Micro calcification

The following algorithm were used to extract any possible micro calcification in ROI.

1- Intensity value of micro calcification is defined.

2- Thresholding is applied on the ROI to find any possible micro calcification.

Note: Density of micro calcification and macro calcification are the same. There is no problem with Macro calcification in the mass, but if there is even one micro calcification in the mass specialist will give higher BIRAD to the mass.

3- Area of the detected density is computed and in case it is lower than macro calcification it defines as micro calcification and the algorithm return value of "1" for that mass, otherwise it will return "0".

Figure 34 presents automatic detection of micro calcification in ROI.

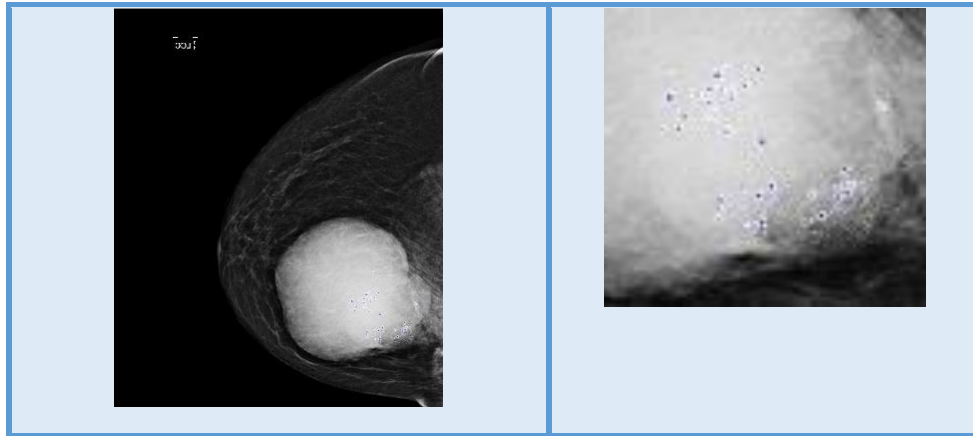


Figure 34: Mass with micro calcification

Shape of margin (smooth, lobulated and speculated (over4)) as mentioned in 3.2.2, is very important for diagnosing a mass. As mentioned in section 3.3, it was decided to extract specularity index, shape of margin for mass with BIRAD score over 4, as meta data and use it in order to classify mass with BIRAD score over 4. Specularity index is computed in next section.

4.2.7.2 Specularity index

For computing specularity index, first margin of ROI was extracted and then smoothed by Gaussian filter in order to remove small redundant edges. Figure 35 present margin extraction of ROI.

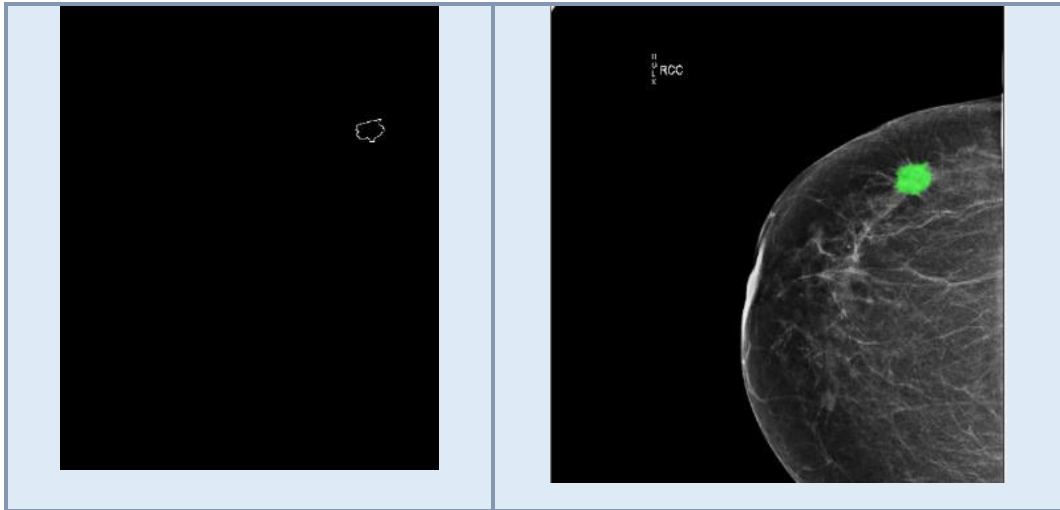


Figure 35: Extracted margin from ROI

- 1- The margin is generated by applying a morphological dilation (8 dilations with ones (3,3) structure element) followed by subtraction from the mass region.
- 2- Gaussian filter and thinning is applied to smooth the edge; otherwise lots of redundant small spikes will be detected on the margin.
- 3- For each pixel on the margin (A), previous pixel B (3 pixels backward) and next pixels C (3 pixels forward) positions (x, y) are obtained.
- 4- Normalized vector AB and AC are considered. "A" as middle point, and the inside angle of "A" is computed for all pixels on the edge as below:

$$A = [a_1; a_2]$$

$$B = [b_1; b_2]$$

$$C = [c_1; c_2]$$

$$AB = \frac{B - A}{\text{norm}(B - A)}$$

$$AC = \frac{C - A}{\text{norm}(C - A)}$$

$$A \cdot B = a_1 * b_1 + a_2 * b_2$$

$$A \cdot C = a_1 * c_1 + a_2 * c_2$$

$$|AB| = \sqrt{((b_2 - a_2)^2 - (b_1 - a_1)^2)}$$

$$|AC| = \sqrt{((c_2 - a_2)^2 - (c_1 - a_1)^2)}$$

$$\theta(i) = \text{acos} \frac{AB \cdot AC}{|AB||AC|} * 180/\pi \quad (\text{inside angle for A})$$

```

5- for i = 0 to size of  $\theta$       (size of  $\theta$  is number of detected angles on the edge)
    if  $\theta(i) < 90$  or  $\theta(i) > 185$  then it is start of lobulation or spike and
         $\theta(i)$  is considered as critical angle
        save the location of i
    go forward till another critical angle is founded
    if  $\theta(i) > 185$  ( $\theta(i)$  is next detected critical angle)
        then a lobulation is detected
        lobulation = lobulation + 1
    else if  $\theta(i) < 90$  then a spike is detected
        length of spike is calculated
        Spike = spike + 1
Go forward

```

Number of spikes, length of the tallest spike, the average length of spikes and number of lobulation were calculated as specular index. According to those features and existence of micro calcification malignant mass were diagnosed. From points mentioned in 4.2.7, it could be concluded that, specialist will give higher BIRAD score to mass that has retraction in the breast tissue. So, in order to fully characterize mass with BIRAD score over 4(4a, 4b and 4c), beside those features (meta data) that were automatically extracted, retraction also was extracted as a binary feature.

4.2.8 Classification mass with BIRAD score over 4

Masses with BIRAD score over 4, (30 with BRAD score 4a, 30 with BIRAD score 4b and 20 with BIRAD score 4c), with meta data automatically extracted as features (mentioned in section 4.2.7) were sent to different classification

methods. Similar to previous section, the proposed CAD has been evaluated using the hold out and cross validation methods.

Feature value for mass with retraction and micro calcifications will be “1”. Mass with BIRAD score 4a are less lobulated and have shorter spikes comparing to masses with BIRAD score 4b and 4c. Mass with BIRAD score 4b have more number of spikes and taller spikes, comparing to mass with BIRAD score 4a. Number of lobulation or length of spikes for a mass with BIRAD score 4c may be the same as those for mass with BIRAD score 4a, but it may have micro calcifications or retraction.

Extracted features from masses with BIRAD score over 4 were sent to different classification algorithms introduced earlier. The results are presented in table 15.

	Classification Tree	Method	# of split	Cross validation results
1	Tree	simple tree	2	95.0
2			4	96.3
3			6	96.3
4		Medium tree	2	95.0
5			4	96.3
6			6	96.3
7		complex tree	2	95.0
8			4	96.3
9			6	96.3

Table 15: Extracted features from mass with BIRAD score over 4 are send to different classification algorithms with different methods.

	Classification SVM	Method	Kernel function	Cross validation results
10	SVM	Linear	Linear	98.8
11			Gaussian	97.5
12			Quadric	98.8
13			Cubic	98.8
14		Fine Gaussian	Linear	97.5
15			Gaussian	95.0
16			Quadric	95.0
17			Cubic	96.3
18		Medium Gaussian	Linear	97.5
19			Gaussian	95.0
20			Quadric	95.0
21			Cubic	95.0
22		Coarse Gaussian	Linear	97.5
23			Gaussian	97.5
24			Quadric	98.8
25			Cubic	98.8
26		Quadric	Linear	98.8
27			Gaussian	97.5
28			Quadric	98.8
29			Cubic	98.8
30		Cubic	Linear	98.8
31			Gaussian	97.5
32			Quadric	98.8
33			Cubic	98.8

Table 15: Extracted features from mass with BIRAD score over 4 are send to different classification algorithms with different methods.

	KNN classification	#Of neighbors	Distance metric	Cross validation results
34	Fine	2	Euclidean	97.5
35			City box	97.5
36			Correlation	87.5
37			Jaccard	81.3
38	Medium	2	Euclidean	97.5
39			City box	97.5
40			Correlation	88.8
41			Jaccard	73.8
42	Coarse	2	Euclidean	97.5
43			City box	97.5
44			Correlation	87.5
45			Jaccard	77.5
46	Weighted	2	Euclidean	97.5
47			City box	97.5
48			Correlation	88.8
49			Jaccard	81.3

Table 15: Extracted features from mass with BIRAD score over 4 are send to different classification algorithms with different methods.

	Ensemble classification	Ensemble method	Decision tree learner. # of learners	Cross validation results
50	Boosted tree	Ada boost	15	95.0
51			25	95.0
52			50	95.0
53			100	95.0
54			200	95.0
55		Bag	15	97.5
56			25	97.5
57			50	97.5
58			100	97.5
59			200	98.8
60		Rusboost	15	60.0
61			25	77.5
62			50	88.8
63			100	88.8
64			200	91.3
65	Bag Tree	Bag	15	97.5
66			25	97.5
67			50	97.5
68			100	91.3
69			200	98.8
70		Ada boost	15	70.0
71			25	70.0
72			50	95.0
73			100	95.0
74			200	95.0

Table 15: Extracted features from mass with BIRAD score over 4 are send to different classification algorithms with different methods.

75	Subspace Discriminant	Subspace	15	91.3
76			25	96.3
77			50	96.3
78			100	96.3
79			200	98.8
80		Bag	15	97.5
81			25	97.5
82			50	97.5
83			100	97.5
84			200	97.5
85	Subspace KNN	Subspace	15	92.5
86			25	92.5
87			50	92.5
88			100	92.5
89			200	92.5
90		Bag	15	96.3
91			25	96.3
92			50	98.1
93			100	98.1
94			200	98.1
95		Ada boost	15	70.0
96			25	95.0
97			50	95.0
98			100	95.0
99			200	95.0

Table 15: Extracted features from mass with BIRAD score over 4 are send to different classification algorithms with different methods.

100	Rusboost	Rusboost	15	70.0
101			25	70.0
102			50	70.0
103			100	85.0
104			200	85.0
105		Bag	15	96.3
106			25	97.5
107			50	97.5
108			100	97.5
109			200	98.8
110		Ada boost	15	95.0
111			25	95.0
112			50	95.0
113			100	95.0
114	200		95.0	

Table 15: Extracted features from mass with BIRAD score over 4 are send to different classification algorithms with different methods.

As it could be seen in table 15, classifications in rows: 59,69, 79 and 109 had highest accuracy in cross validation method. So average accuracy for 10 times run for each of them was calculated and results are presented in table 16.

	Classification/ensemble method	#of learners	Min accuracy	Max accuracy	Average accuracy
1	Boosted tree/bag	200	97.5	98.8	98.41
2	Bag tree/bag	200	91.3	98.8	94.3
3	Sub discriminant/subspace	200	91.3	98.8	94.85
4	Rusboost/bag	200	96.3	98.8	97.8

Table 16: Average accuracy for ten times run of classifiers with higher accuracy

Confusion matrix of the classification result is presented in figure 36.



Figure 36: a-Per true class result, b-Per predicted class result, c- Over all result of Confusion matrix for classification result



Figure 36: a-Per true class result, b-Per predicted class result, c- Over all result of Confusion matrix for classification result

In figure 36, per true class, per predicted class and over all confusion matrix for classification results are presented. ("1" is BIRAD score of 4a, "2" is for BIRAD score 4b and "3" BIRAD score 4c.). Figure 36-a, shows that, one of the masses

with BIRAD score 4c, which was 5% of masses with BIRAD score 4c (5% false negative), was miss classified. Figure 36-c, shows that BIRAD score 4b has 1.3% false positive result, mass with BIRAD score 4c which was miss classified as mass with BIRAD score 4b. All mass with BIRAD score 4a and 4b were classified accurately, and one mass with BIRAD score 4c was miss classified.

In hold out method, the average accuracy was 99% and the best accuracy was 100% by boosted tree, bag ensemble method and with 200 learners of decision tree classification.

In cross validation method, the best accuracy with same ensemble method was 98.8% and the average accuracy for ten times evaluation was 98.41%.

4.3 Discussion and conclusion

From point mentioned in 4.1 it could be concluded that BIRAD score 2 and over 4(4a, 4b and 4c) are the most important ones. In order to accurately classify different BIRAD scores, number of masses in each BIRAD score should be approximately the same. So it was decided to classify in two stage.

From point mentioned in 4.2.1 and table 5 it could be concluded that it is essential to reduce effect of noise as it effect classification accuracy result.

Automatic Chan-Vese active contour was used to segment masses. Accuracy of reproducibility was 99.5%. Also manually segmented masses were compared with automatic segmented ones and the accuracy was 85.9%.

35 medical and mathematical features were extracted from each mass; their normalized values are presented in table 6. Shape of mass was defined by compactness. Acutance and contrast were extracted to differentiate ill-defined and well defined masses. Some intensity based features were extracted to define mass as dense, Iso-dense or low dense mass. Haralick texture based features were used to define mass as homogenous or heterogeneous.

From points mentioned in 4.2.4 it could be concluded that:

MRMR chose mostly texture features and then intensity based features. It also chose compactness as second most important one. **Kruskal Wallis** chose compactness, shape based feature, as the most important one and intensity based features the least important ones. **Kruskal Wallis+ MRMR** also chose compactness as the most important one but texture features and intensity based features have the same priority approximately. **SFFS** chose acutance and solidity as the most important ones. Compactness has the 4th priority. **SFFS+ MRMR** still chose solidity the most important one and then texture features and intensity features.

Selected features were sent to classifiers. The results in table 10 reveals that average accuracy in cross validation method for SFFS and SFFS+MRMR as feature selection method followed by boosted tree hybrid classifier, and Ada boost ensemble method, with decision tree learner type and 100 learners was 98.62%.

To evaluate effect of feature selection methods all 35 extracted features were sent to classifiers. From data presented in table 11 it could be concluded that although classifications in rows 3 and 6 have highest accuracy (100%), still Boosted tree classification with Ada boost ensemble method and 200 learners had highest average accuracy,98.75%.

From points mentioned in 4.2.7 it was decided to extract micro calcification, retraction and specularlarity index as meta data (extra features) in order to classify B4a, B4b and B4c. Those features were send to classifiers and average accuracy for ten times evaluation was 98.41% for boosted tree, bag ensemble method, with 200 learners of decision tree classification (cross validation method).

CHAPTER 5

5 General discussion and Conclusion

A computer aided diagnosis system (CAD) was developed to fully characterize and classify mass and give BIRAD scores using mammographic image.

Comparing all breast abnormalities mass is the most common one and most complicated one to diagnose (mentioned in section 1.2.6). Considering points mentioned in chapter 1 such as: critical patients are recommended to start the check up with mammogram, comparing to MRI, mammogram is less expensive, Ultra sound is not a diagnosing system, hence it was decided to focus on giving BIRAD score to mass in mammograms.

One hundred seventy three masses (80 masses with BRAD score 2, 80 mass with BIRAD score over4 and 13 mass with BIRAD score 3) were automatically annotated using active contour. Based on points mentioned in section 4.2 and images presented in Figure 29, Chan-Vese active contour segmentation method with free hand initialization were applied.

According to BIRAD score system mentioned in 4.1, classifying mass with BIRAD score 2 and over 4 is the most important ones. In order to have more accurate classifier number of data in all classes (2, 4a, 4b and 4c) should be approximately the same. So the proposed CAD first classified benign and malignant mass (mass with BIRAD score 2 and BIRAD score over4). For this, feature extraction scheme is

applied to fully characterize a benign and malignant mass by extraction of the most dominant features that had a large impact on the outcome of a patient biopsy (mentioned in section 4.2.3). Thirty-five medical and mathematical shape, intensity and texture features associated to the mass were extracted and their values are presented in table 8. As benign mass is round or oval, so it's more compact comparing malignant mass. Value for density based features in benign mass are higher, comparing to malignant mass. Benign mass is well defined, so the value for acutance and contrast features were higher in benign mass comparing to malignant mass.

This was followed by Several feature selection schemes, mentioned in section 4.2.4, that were applied to select the most relevant features. MRMR chose texture features and then intensity based features. It also chose compactness as second most important one. Kruskal Wallis chose compactness, shape based feature, as the most important ones and intensity based features the least important ones. Here, texture features also had higher priority comparing to intensity based features. Kruskal Wallis+ MRMR also chose compactness as the most important one but texture features and intensity based features had the same priority approximately. SFFS chose acutance and solidity as the most important ones. Compactness had the 4th priority. Beside those feature, it selected texture features.

SFFS+ MRMR still chose solidity the most important one and then texture features and intensity features.

Selected weighted features were classified by a variety of classifiers (mentioned in section 4.2.5), including ensemble classifiers. In summary, the results (presented in table 12) revealed that boosted tree hybrid classifier with Ada boost ensemble method, decision tree learner type, 100 learners had 99.4% accuracy, when using SFFS, SFFS+MRMR.

KruskalWallis and KruskalWallis + MRMR feature subsets achieved 100 % accuracy in hold out method and 99.4 % accuracy in cross validation method when using bag tree classification with Bag ensemble method, decision tree, 100 learners. Comparing 3 mentioned feature subsets (SFFS and SFFS+MRMR and Kruskal Wallis), Kruskal Wallis had lower number of features comparing to SFFS, 22 and 26 respectively.

For evaluating the effect of feature selection methods all extracted features were fed into different mentioned classifiers and the results are presented in table 14. Results revealed that boosted tree classifier with Ada boost ensemble method and 200 learner had 99.4% and 98.75% best and average accuracy in cross validation method respectively. The same classifier had 100% and 99% best and average accuracy in hold out method. It could be concluded that although results were approximately close to each other, still using feature

selection methods leads to faster CAD, as number of learners were less in that system.

For sub classification of BIRAD 4 to a, b and c, automatically extracted meta data (micro calcification and specular index, which are described in section 4.2.7) and retraction, as features were fed into different classifiers. In hold out method, the average accuracy was 99% and the best accuracy was 100% by boosted tree, bag ensemble method and with 200 learners of decision tree classification. In cross validation method, the best accuracy with same ensemble method was 98.8% and the average accuracy for ten times evaluation was 98.41%. (The results are presented in table 16).

Reproducibility of automatic segmentation (mentioned in section 4.2.2) of masses (each mass was segmented 3 times with three different initialization) was assessed by Jaccard index and 99.5% accuracy was achieved. The difference in segmentation didn't have any influence on CAD's final results.

The proposed CAD was trained with masses with BIRAD score 2 and over 4 but it was tested with masses with BIRAD score of 3(mentioned in section 4.2.6). The CAD's results were the same as specialist follow up result in six months.

Finally, to evaluate the effect of noise (mentioned in section 4.1), based on results presented in table 7, it was demonstrated that Salt & pepper noise reduced classification accuracy for mass with BIRAD score over 4 to 90%,

Gaussian and speckle noise reduced classification accuracy for benign and malignant mass to 92.5% and 93.75% respectively. So the CAD includes a preprocessing step to de-noise mammograms. For this based on results presented in table 6, median, adaptive median and wiener filters were employed to reduce Salt & pepper, Gaussian and speckle noise. Hence classification accuracy for mass with BIRAD score over 4 became 98.8% and classification accuracy for benign and malignant mass increased to 97.6% and 98.8%.

The best specification that give the best performance was used to build the optimum proposed CAD system to predict BIRAD scores are summarized as follow: sequential forward floating feature (SFFS) selection and a boosted tree hybrid classifier with Ada boost ensemble method, decision tree learner type and 100 learners, achieved sensitivity and specificity of 99.4% and 100% respectively for classifying benign and malignant mass. Boosted tree hybrid classifier Bag ensemble method, decision tree learner type and 200 learners, achieved sensitivity and specificity of 98.8% and 100% respectively for classifying mass with BIRAD score over 4(a, b and c).

The above results demonstrate the high sensitivity and high specificity of the proposed CAD compared to other research. This is due to the fact that most of research only used mathematical features including texture features without

considering medical features and meta data which play important role in diagnosis.

Reviewed CAD, some authors extracted just texture features for diagnosing a mass and they didn't mention any thing about noise removal and pre-processing, despite the fact that noise can affect classification results (table 7).

Automatic segmentation of ROI and its accuracy is important in diagnosing system. Most of reviewed CADs didn't mention any thing about detecting margin/border of mass and thus no report was given about the accuracy of segmentation. Shape of ROI's margin is important in diagnosing system. Despite such important fact, some of researchers used a box for approximate the segmentation of ROI without applying a segmentation method to segment the exact ROI. In addition, the size of the box can affect the result and complexity of computation.

Because of complexity of exact segmentation of ROI, some of researchers mostly focused only on implementing segmentation methods while the left the other important components of a full CAD system as a CAD diagnosing system beside accurate segmentation needs feature extraction, feature selection and classification.

Most of authors in reviewed CADs just used either shape based features, texture features or intensity based features and none of the reviewed papers, used all types of features together to fully characterize ROI.

Most of the current CADs had either lower accuracy comparing to our proposed CAD, or the feature dimension was huge which add extra overhead. Furthermore, accuracy of such works dependents on the size of windows chosen for feature extraction. Last and most important one, none of the reviewed CADs assigned a BIRAD to a mass where it is ultimate goal of our proposed CAD

Future works:

- Although the current result is very promising, the robustness of scheme needs further investigation on a larger dataset.
- Fully automatic detection and segmentation of mass without interaction with users.
- Mass with BIRAD score 4+ are checked in different modality (medical imaging devices). Exact positioning of a mass and correlate it in different modality is important. There are lots of cases where there are some masses behind each other in a breast and just one is malignant and other ones are benign. So it's important to address that malignant mass in

different modality, as oncologist should know the exact position of malignant mass.

- Asymmetry in tissue is one the abnormalities that is mostly missed in very first stages as its difficult for specialist to find that abnormality in the breast.

To find asymmetry in tissue specialist should read all four images of mammogram together and compare them all to each other in order to find that asymmetry. It is time consuming, difficult and of course diagnosing it needs lots of experiences. So an extended CAD that can diagnose asymmetry could help specialists in not missing it.

References:

1. Altman N., (1992). "An introduction to kernel and nearest-neighbor nonparametric regression", *The American Statistician*, PP. 175–185.
2. Andersson I, et al, (1997),” Number of projections in Mammography: influence on detection of breast disease”, *AJR Am J Roentgenol* PP.349–351
3. A. C. S. (AMS)., (2006), “Learn about breast cancer”, <http://www.cancer.org>
4. ArzavJain D., (2007),” Breast Mass Classification from Mammograms using Deep Convolutional Neural Networks”, *Stanford University*, P.458
5. Altekruse SF., et al, (2007),” SEER Cancer Statistics Review”, *National Cancer Institute, Bethesda*, PP.1-15.
6. A.C.S., (2010),” Learn about breast cancer“, Available at: [〈http://www.cancer.org〉](http://www.cancer.org)
7. Alam RN, (2014),” Computer-aided mass detection on digitized mammograms using a novel hybrid segmentation system”, *Int J Biol Biomed Eng* PP.51–58
8. Aziz Makandar, et al, (2015), “Breast Cancer Image Enhancement using Median Filter and CLAHE”, *International Journal of Scientific & Engineering Research*, Volume 6, Issue 4, 462 ISSN 2229-5518
9. Athira, et al, (2016), “An Overview of Mammogram Noise and De-Noising Techniques “, *International Journal of Engineering Research and General Science* Volume 4, Issue 2, ISSN 2091-2730
10. Arevalo J., (2016), “M.A.G. Representation learning for mammography mass lesion classification with convolutional neural networks”, *Comput. Methods Programs Biomed.* PP. 248–257.
11. Arevalo J., et al., (2016), “Representation learning for mammography mass lesion classification with convolutional neural networks”, *Computer methods and programs in biomedicine*, PP.248–257
12. Aghaei F., et al, (2016),” Applying a new quantitative global breast MRI feature analysis scheme to assess tumor response to chemotherapy”, *J. Magn. Reson. Imaging*, PP.1099–1106.
13. Arivazhagan S, yet al, (2017), “Texture classification using wavelet transform”, *Pattern Recognit Lett* 24, PP.1513–1521
14. Breiman L., et al, (1984), “Classification and Regression Trees”, *Boca Raton, FL: CRC Press*, P.234-251.
15. Boser BE, et al, (1992),” A training algorithm for optimal margin classifiers”, In *Proc. of the fifth annual workshop on Computational learning theory*, PP. 144–152

16. Baker J., et al, (1996), "Breast imaging reporting and data system standardized Mammography lexicon: observer variability in lesion description", *AJR Am J Roentgenol*, pp. 773-778
17. Burges C, (1998)," Tutorial on support vector machines for pattern recognition". *Data Min Knowl Disc* PP.955–974
18. Bochud FO, et al, (1999)," Estimation of the noisy component of anatomical backgrounds", *Med Phys* PP.1365–1370
19. Burgess AE, et al, (1999)," Human observer detection experiments with mammograms and power-law noise ", *Med Phys* PP.419–437
20. Berg W., et al, (2000), "Breast Imaging Reporting and Data System: inter- and intraobserver variability in feature analysis and final assessment", *AJR Am J Roentgenol*, pp. 1769-1777
21. Bovis K., et al, (2000), "Detection of masses in mammograms using texture features", in: *Proceedings of the 15th International Conference on Pattern Recognition, ICPR00*, pp. 267–269.
22. Blot L. et al, (2000),"Extracting background texture in mammographic images: a co-occurrence matrices based approach", in *Proceedings of the 5th International Workshop on Digital Mammography, Toronto (Canada)*, pp. 142-148.
23. Boris K., et al, (2002), "Classification of mammographic breast density using a combined classifier paradigm", in *Proc. Med. Image Understanding Anal. Conf.*, pp. 177–180.
24. Beatriz A., et al, (2004)," Data mining with decision trees and neural networks for calcification detection in mammograms", in: *Proceedings of the Third Mexican International Conference on Artificial Intelligence, LNCS*, pp. 232–241.
25. Birdwell RL, et al, (2004)," Mammographic characteristics of 115 missed cancers later detected with screening Mammography and the potential utility of computer-aided detection", *Radiology* PP.192–202
26. Berg W.A., et al, (2004), "Supplemental screening sonography in dense breasts", *Radiol Clin North Am*, pp. 845-851
27. Bissonnette M, et al, (2005)," Digital breast tomo synthesis using an amorphous selenium flat panel detector", *Proc SPIE* PP.529–540
28. Balleyguier C., et al, (2007), " BIRADS™ classification in Mammography ",*European Journal of Radiology*, Volume 61, Issue 2, PP. 192-194
29. Ball J., et al, (2007)," Digital mammographic computer aided diagnosis (cad) using adaptive level set segmentation", In *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, PP. 4973–4978.
30. B ath M, et al, (2007)," Visual grading characteristics (VGC) analysis: a non-parametric rank-invariant statistical method for image quality evaluation", *Br J Radiol* PP.169–176
31. Bowyer K., et al, (2011), "Tutorial: Masses", Available from: <http://imagnis.com/>
32. Bhangale T., et al, (2015)," An unsupervised scheme for detection of micro calcifications on mammograms", *Proc. IEEE Int Conf Image Proc. Vancouver, BC, Canada*, pp. 184–187
33. Connors RW, et al, (1980), "A theoretical comparison texture algorithm", *IEEE Trans Pattern Anal Mach Intell* 2: PP.204–222

34. Caselles V., et al, (1993) "A geometric model for active contours in image processing", Numer. Math. PP. 1–31.
35. Chan H., et al. (1995), "Computer-aided classification of mammographic masses and normal tissue: linear discriminant analysis in texture feature space", Journal of Physics in Medicine and Biology, Vol. 40, pp. 857-876.
36. Caselles V., et al, (1997), "Geodesic active contours", Int. J. Comput. Vis. PP. 61–79.
37. Cheng H., et al, (1998), "The ACR Breast Imaging Reporting and Data System (BIRADS)", American College Radiology, Third Edition, http://www.imaginis.com/pro/breast_imag_resrc/acrbirads.asp
38. Chan T, et al, (2001), "Active contours without edges", IEEE Trans. Image Process., pp. 266-277
39. Costa DD, et al, (2001)," Classification of breast tissue in mammograms using efficient coding", Bio-Medical Engineering, On-Line, PP. 10:55.
40. Cheng H., et al, (2003), "Computer-aided detection and classification of micro calcifications in mammograms: a survey, Pattern Recognition", pp. 2967-2991
41. Carney PA, et al, (2003), "Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography", Ann Intern Med PP.168–175
42. Cheng Hh., et al, (2006), "Approaches for automated detection and classification of masses in mammograms", Pattern Recognition, Vol. 39, pp. 646-668.
43. Christiyanni J., (2006), "Fast detection of masses in computer aided Mammography", IEEE Signal Process Mag , PP. 54–64 .
44. Cascio D, et al, (2006), "Mammogram segmentation by contour searching and mass lesions classification with neural network", IEEE Trans. Nucl. Sci., PP. 2827-2833
45. Cheng H, et al, (2008), "Approaches for automated detection and classification of masses in mammograms", Pattern Recognition, Vol. 39, pp. 646-668.
46. Carney PA, et al, (2004), "Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening Mammography", Ann Intern Med PP.168–175
47. Cao G., et al, (2008), "Optical aerial image partitioning using level sets based on modified Chan–Vese model", Pattern Recognit. Lett. PP. 457–464.
48. Costa DD., et al, (2011)," Classification of breast tissue in Mammograms using efficient coding", Bio-Medical Engineering, PP. 10:55
49. Chun-yan Y., et al, (2013)," oval active contour model for image segmentation using distance regularization term", Computers and Mathematics with Applications journal, PP.1746–1759
50. Carneiro G., et al., (2015), "Unregistered multi view Mammogram analysis with pre-trained deep learning models", In International Conference on Medical Image Computing and Computer-Assisted Intervention, PP.652–660.
51. Cherif chaabani A, et al, (2018), "An automatic-pre-processing method for mammographic images", JDCTA Int J Digit Content Technol Appl, PP.190–201
52. Dirac delta function, From Wikipedia, the free encyclopedia

53. Daugman JG, (1980), "Two-dimensional spectral analysis of cortical receptive field profiles", *Vision Res* PP.847–856
54. Duda RO, et al, (2000), "Pattern classification, 2nd edn". Wiley, New York, P.688
55. Duda RO, et al, (2001), "Pattern classification", 2nd edn. Wiley, New Yorkzb MATH, P.637.
56. Doyle J, et al, (2005), "Can the radiologist accurately predict the adequacy of sampling when performing ultrasound-guided core biopsy of BI-RADS category 4 and 5 lesions detected on screening Mammography?", *Clin Radial*, pp. 999-1005
57. Ding C, et al., (2005), "Minimum redundancy feature selection from microarray gene expression data", *J Bioinforma Comput Biol*, PP.185–205.
58. Dean C., (2006), "Ilven to Improved cancer detection using computer-aided detection with diagnostic and screening Mammography: prospective study of 104 cancers", *Am. J. Roentgenol*, pp. 20-28
59. Domínguez AR, et al, (2009), "Towards breast cancer diagnosis based on automated segmentation of masses in Mammograms", *Pattern Recognition* PP.1138–1148
60. Duda R., et al, (2012), "Pattern Classification, John Wiley and Sons, second edition", *IEEE Trans. Pattern Anal. Res* 20:847–856
61. Demšar J, (2014), "Statistical comparisons of classifiers over multiple data sets", *Mach Learn Res* PP.1–30.
62. DeSantis C., et al., (2014), "Breast cancer statistics,2013". *CA: a cancer journal for clinicians*, PP.52–62.
63. Dheeba J., et al, (2014), "Computer-aided Detection Improves Early Breast Cancer Identification", *Journal of Biomedical Informatics*, Volume 49, PP.45-52
64. Demšar J, (2014), "Statistical comparisons of classifiers over multiple data sets", *Mach Learn Res* PP.1–30.
65. Dhungel N., et al., (2015), "Automated mass detection in Mammograms using cascaded deep learning and random forests", In *Digital Image Computing: Techniques and Applications (DICTA)*, International Conference on, PP. 1–8.
66. Dehmeshki J., et al , (2015), "CT Attenuation Values of Blood and Myocardium: Rationale for Accurate Coronary Artery Calcifications Detection with Multi-Detector CT", *PLoS One*; San Francisco Vol. 10, Iss. 4, e0124175. DOI: 10.1371/journal.pone.0124175
67. Ertas H., et al, (2001), "Feature extraction from mammographic mass shapes and development of a Mammogram database", in: *Proceedings of the 23rd Annual EMBS International Conference, Istanbul, Turkey*, pp. 2752–2755.
68. El-Naqa I, et al, (2002), "A support vector machine approach for detection of micro calcifications", *IEEE Trans Med Imaging* PP.1552–1563
69. Elter M, et al, (2009), "CADx of mammographic masses and clustered micro calcifications: a review", *Med Phys* PP.2052–2068
70. Eltoukhy, M.,et al, (2010), "A comparison of wavelet and curve let for breast cancer diagnosis in digital mammogram", *Computers in Biology and Medicine* Volume 40, Issue 4, PP. 384-391
71. Esteve J, et al, (2015), "Facts and figures of cancer in the European Community", In: *Tech. Rep.*, International Agency for Research on Cancer, PP. 832-978

72. Elmann M., et al, (2015), "Implementation of Practical Computer Aided Diagnosis System for Classification of Masses in Digital Mammograms", Publisher: Scholars's Press, ISBN: 978-3639763881.
73. Freer T., et al, (2001), "Ulisse Screening Mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center", *Radiology*, pp. 781-786
74. Flores B., et al, (2004), "Data Mining with Decision Trees and Neural Networks for Calcification Detection in Mammograms", in: *Proceedings of the Third Mexican International Conference on Artificial Intelligence, LNCS, Springer*, pp. 232–241.
75. Fisher RA, (2012), "The use of multiple measures in taxonomic problems." *Ann Eugen* PP.179–188
76. Grigorescu S, et al, (1999), "Comparison of texture features based on Gabor filters", *IEEE Trans Image Proc* PP.1160–1167
77. Gulsun M., et al, (2003), "Evaluation of breast micro calcifications according to Breast Imaging Reporting and Data System criteria and Le Gal's classification", *Eur J Radiol*, pp. 227-231.
78. Gao S., et al, (2005), "Image segmentation and selective smoothing by using Mumford–Shah model", *IEEE Trans. Image Process.* PP.1537–1549.
79. García S, et al, (2008), "An extension on Statistical Comparisons of Classifiers over Multiple Data Sets" for all pairwise comparisons. *Mach Learn Res* PP.2677–2694
80. Gibbs P, et al, (2008), "Textural analysis of contrast enhanced MR images of the breast.", *Magn Reson Med*, PP.92–98
81. Hilleren DJ, et al, (1991), "Invasive lobular carcinoma: mammographic findings in a 10-year experience.", *Radiology* PP.25–26
82. Haralick, et al, (1992), "Computer and Robot Vision: Vol. 1", Addison-Wesley, p. 459.
83. Heath M., et al, (2000), "The digital database for screening Mammography, in: *Proceedings of the International Workshop on Digital Mammography*", pp. 212–218.
84. Hanmandlu M., et al, (2004), "A fuzzy approach to texture segmentation", *IEEE Explore*, P.12
85. Hanchuan Peng, (2005), "Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy", *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, VOL. 27, NO. 8.
86. Hiremath PS, et al, (2006), "Texture classification using wavelet packet decomposition", *ICGSTs GVIP J* 6(2), P.77–80
87. Hastie T., et al., (2008), "The Elements of Statistical Learning", Second Edition. NY: PP.2450-2470.
88. Hsu CW, et al, (2010), "A practical guide to support vector classification", Technical report, Department of Computer Science and Information Engineering, National Taiwan University, P.16.
89. Haralick, et al, (2010), "Textural Features for Image Classification", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-3, pp. 610-621.
90. Hanmandlu M, et al, (2015), "A fuzzy Approach to texture segmentation. *Proc Int Conf Inf Technol*", *Coding and Computing* 1, PP.636–642
91. Huynh Q., (2016), "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks.", *Med. Imaging*.P.3

92. Hall-Beyer M, (2017), "GLCM texture: a tutorial", v.2.7.1, on-line document, www.fp.ucalgary.ca/mhallbey/tutorial.htm, P.10.
93. Ikeda DM, et al, (2007), "Interval carcinomas in the Malmö Mammographic Screening Trial: radiographic appearance and prognostic considerations", *AJR Am J Roentgenol* PP.287–294
94. Jain A.,et al, (1997), "Feature-selection: Evaluation, application, and small sample performance", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 2, pp. 153-158.
95. Jain A., et al, (2008), "Statistical pattern recognition: a review", *IEEE Trans. Pattern Anal. Mach. Intel.* PP. 4–37.
96. James JJ, et al, (2008)," *The breast*", 5th ed. New York, NY: Churchill Livingstone; chap 52.
97. Jiao Z., (2016)," A deep feature based framework for breast masses classification", *Neuro computing*, PP. 221–231.
98. Kerlikowske K, et al, (1996), "Effect of age, breast density, and family history on the sensitivity of first screening Mammography ", *JAMA*, PP.33–38
99. Kupinski M. A., et al (1997), "Feature selection and classifiers for the computerized detection of mass lesions in digital Mammography", *Proc. Intl. Conf. Neural Networks*, PP.2460–2463.
100. Kass M, et al, (1998), "Snakes: Active contour models", *International Journal of Computer Visio*, vol. 1, pp. 321-331
101. Kudo M.,et al, (2000), "Comparison of algorithms that select features for pattern classifiers", *Pattern Recognition*, vol. 33, no. 1, pp. 25-41.
102. Kolb TM, et al, (2002), "Comparison of the performance of screening Mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations", *Radiology*, PP.165–175
103. Karssemeijer N., e al, (2004)," Effect of independent multiple reading of Mammograms on detection performance, *SPIE Med. ", Imaging*, 5372, pp. 82-89
104. Kopans D., et al (2007), "Breast Imaging (3rd ed.)", Philadelphia, USA
105. Kulak E, (2008)," Image analysis of textural Features for Content based retrieval", Sabanci University, ID code:8153, P. 125.
106. Karella A., et al., (2015), "computer aided detection and diagnosis in medical imaging", CRC press, a Taylor & Francis
107. Kannan S., et al, (2016)," PERFORMANCE COMPARISON OF NOISE REDUCTION IN MAMMOGRAM IMAGES", *IJRET: International Journal of Research in Engineering and Technology* PP.2319-1163
108. Kim E., (2018)," Applying data-driven imaging biomarker in mammography for breast cancer screening: preliminary study", P.8
109. Liberman, L., et al, (1997), "Stereotactic core biopsy of calcifications highly suggestive of malignancy", *Radiology*, pp. 673-677
110. Lewin JM, et al, (2001)," Comparison of full-field digital Mammography with screen-film Mammography for cancer detection.", Results of 4,945 paired examinations. *Radiology* PP.873–880

111. Li H., et al, (2001), "Computerized radiographic mass detection-part II: lesion site selection by morphological enhancement and contextual segmentation", IEEE Trans. Med. PP. 302–313.
112. Lehman C., et al, (2001), "Effect of training with the American college of radiology breast imaging reporting and data system lexicon on mammographic interpretation skills in developing countries", Acad Radiol, pp. 647-650
113. Liberman L., et al, (2002), "Breast imaging reporting and data system (BI-RADS)", Radiol Clin North Am, pp. 409-430.
114. Lkeda DM, et al, (2002), "Interval carcinomas in the Malmö Mammographic Screening Trial: radiographic appearance and prognostic considerations.", AJR Am J Roentgenol PP.287–294
115. Lehman C., et al, (2002), "Use of the American College of Radiology BI-RADS guidelines by community radiologists: concordance of assessments and recommendations assigned to screening mammograms", AJR Am J Roentgenol, pp. 15-20
116. Leconte L., et al., (2003)," Mammography and subsequent whole-breast sonography of nonpalpable breast cancers: the importance of radiologic breast density", AJR Am J, Roentgenol, pp. 1675-1679
117. LO JY, et al, (2006), "Breast tomosynthesis: Initial clinical experience with 100 human subjects (abs), Radiological Society of North America scientific assembly and annual meeting program.", Oak Brook, Ill: Radiological Society of North America SSG01-03
118. Leonardo de O. Martins, et al, (2006), "Classification of Normal, Benign and Malignant Tissues Using Co-occurrence Matrix and Bayesian Neural Network in Mammographic Images", SBRN'06, IEEE computer society, pp. 24-29
119. Lladó X, et al, (2009), "A textural approach for mass false positive reduction in Mammography", Compute Med Imaging Graph, PP.415–422.
120. Liincott Williams & Wilkins Publishers, (2010), "ACR Breast imaging reporting and data system (BI-RADS), Breast Imaging Atlas (4th ed.)", American College of Radiology, Reston
121. Loh W., (2011)," Classification and regression trees ", wires.wiley.com/widm, Volume 1. P.10.
122. Ioan B, et al, (2011)," Directional features for automatic tumor classification of Mammogram images", Biomed Signal Process Control, P.370–378
123. Lahmiri S, et al, (2011)," Hybrid discrete wavelet transform and Gabor filter banks processing for mammogram features extraction", Proc. NEWCAS, France. IEEE Comput Soc, PP.53–56
124. Liu Y, et al, (2011), "Region based image retrieval with high-level semantics using decision tree learning", Pattern Recognit 41, P.2554–2570 8.
125. Liu S., et al, (2012), "A local region-based Chan–Vese model for image segmentation Pattern Recognition", pp. 2769-2779
126. Liu X., et al, (2014)," Mass Classification in Mammograms Using Selected Geometry and Texture Features, and a New SVM-Based Feature Selection Method", IEEE SYSTEMS JOURNAL, PP. 175–185.
127. Liu X., et al, (2014), "Mass Classification in Mammograms Using Selected Geometry and Texture Features, and a New SVM-Based Feature Selection Method", IEEE SYSTEMS JOURNAL, VOL. 8, NO. 3.

128. Li H., et al,(2017),“Breast masses in Mammography classification with local contour features”, *BioMed. Eng. OnLine* PP. 44–54.
129. DanielLévy, et al, (2017),” Breast Mass Classification from Mammograms using Deep Convolutional Neural Networks”, Stanford University
130. Mumford D., (1989), “Optimal approximation by piecewise smooth functions and associated variational problems”, *Commun. Pure Appl. Math*, vol. 42, pp. 577–685.
131. Malladi R,et al, (1995), “Shape modeling with front propagation: a level set approach”, *IEEE Trans. Pattern Anal. Mach. Intell.* PP. 158–175.
132. Metz CE., et al, (1998), “The computer program ROCKIT 0.9B”, Available from C. Metz, Dept of Radiology, University of Chicago, Chicago, IL, USA, <http://www-radiology.uchicago.edu/krl/>
133. Markey M. K., et al., (2002),” Cluster analysis of BI-RADS descriptions of biopsy-proven breast lesions”, In: *Medical Imaging: Image Processing, Proceedings of SPIE* Vol. 4684, pp. 363-370
134. Markey M., et al, (2002), “Cluster analysis of BI-RADS descriptions of biopsy-proven breast lesions”, *Medical Imaging: Image Processing, Proceedings of the SPIE*, vol. 4684, pp. 363–370
135. Markey M, et al, (2003), “Self-organizing map for cluster analysis of a breast cancer database”, *Artificial Intelligence in Medicine*, Vol. 27, No. 2, pp.113–127
136. Mendez A., et al, (2004), “Evaluation of Breast Imaging Reporting and Data System Category 3 mammograms and the use of stereotactic vacuum-assisted breast biopsy in a nonacademic community practice”, *Cancer*, pp. 710-714
137. Mehul P., et al, (2005),” Classification of mammographic lesions into BI-RADS shape categories using the Beamlet Transform”, *Medical Imaging: Image Processing, Proc. of the SPIE*, vol. 5747, pp.16-25
138. Mertelmeier T, et al, (2006),” Optimizing filtered back projection reconstruction for a breast tomo synthesis device”, *Proc SPIE*, PP.131–142
139. Manning C., et al., (2008), “Introduction to Information Retrieval”, NY: Cambridge University Press, P.81
140. Mohammed J., et al, (2010),” AN EFFICIENT AUTOMATIC MASS CLASSIFICATION METHOD IN DIGITIZED MAMMOGRAMS USING ARTIFICIAL NEURAL NETWORK”, *International Journal of Artificial Intelligence & Applications (IJAIA)*, Vol.1, No.3, P .13.
141. Mohanty AK, et al, (2013),” A novel image mining technique for classification of mammograms using hybrid feature selection”, *Neural Comput Appl.* PP. 1151–1161
142. Mohanty A., et al, (2013), “Texture-based features for classification of mammograms using decision tree”, *Neural Comput & Applic*, P.23.
143. Manas S.,et al , (2015), “Soft, Hard and Block Thresholding Techniques for Denoising of Mammogram Images”, PP.186-191
144. Muramatsu C., et al, (2016),” Breast mass classification on mammograms using radial local ternary patterns”, *Comput. Biol. Med.* PP.43–53.
145. Mohammed A., et al, (2018),” Computer Methods and Programs in Biomedicine”, *Computer Methods and Programs in Biomedicine*, PP. 85–94.

146. Niklason LT, et al, (1997),” Digital tomosynthesis in breast imaging “, Radiology PP.399–406
147. Nagel Rufus N., et al, (1998), “Analysis of methods for reducing false positives in the automated detection of clustered micro calcifications in mammograms”, Med. Phys., pp. 1502-1506
148. NEMA Standards Publications, (2004),” Digital Imaging and Communications in Medicine (DICOM)”, part 14, National Electrical Manufacturers Association 1300 N. 17th Street Rosslyn, Virginia 22209 USA, P.55
149. Nugroho H., et al, (2014), “Identification of Malignant Masses on Digital Mammogram Images based on Texture Feature and Correlation based Feature Selection”, 6th International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, Indonesia, PP. 254-261.
150. Osher S., (1998), “Fronts propagation with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations”, Journal of Computational Physics, vol. 79, pp. 12–49.
151. Orel S., et al, (1999), “BI-RADS categorization as a predictor of malignancy”, Radiology, pp. 845-850
152. Osher S., (2004), “Level Set Method. In: Numerical Geometry of Images”, Springer, New York, NY, PP.978-998
153. Obenauer S., (2005), “Applications and literature review of the BI-RADS classification”, Eur Radiol, pp. 1027-1036
154. Orman J, et al, (2006),” Adaptation of image quality using various filter setups in the filtered back projection approach for digital breast tomo synthesis”, International Workshop on Digital Mammography, pp. 175-182
155. Pudil P., et al, (1994), “Floating search methods in feature selection”, Pattern Recognition Letters, vol. 15, no. 11, pp. 1119-1125.
156. Pijnappel R., et al, (2004), “Reproducibility of mammographic classifications for non-palpable suspect lesions with micro calcifications”, Br J Radiol, pp. 312-314
157. Pisano, et al, (2005),” Diagnostic performance of digital versus film Mammography for breast-cancer screening.”, N Engl J Med PP.1773–1783
158. Papadopoulos A., et al, (2008), “Improvement of micro calcification cluster detection in Mammography utilizing image enhancement techniques”, Computers in Biology and Medicine, pp:1045 – 1055
159. Polat K, et al, (2009),” A novel hybrid intelligent method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problems”, Expert Syst Appl 36, P.1587–1592 9.
160. Park SH, et al, (2018),” Receiver operating characteristic (roc) curve”, practical review for radiologists. Korean J Radiol 5(1), P.11–18
161. Rafferty EA, (2003), “Breast tomosynthesis. Adv Digital Radiography”, RSNA Categorical Course in Diagnostic Radiology Physics, pp. 219–226
162. Roentgenol A., et al, (2003),” Rationale for a trial of screening breast ultrasound”, American College of Radiology Imaging Network (ACRIN), pp. 1225-1228
163. Rashed E., et al, (2006), “Neural networks approach for Mammography diagnosis using wavelets features”, in: Proceedings of the First Canadian Student Conference on Biomedical Computing.P.12

164. Rangayyan R, et al, (2007), "A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs", J. Franklin Inst., PP.312-348
165. Ruschin M, et al, (2007)," Dose dependence of mass and micro calcification detection in digital Mammography: free response human observer studie", Med Phys PP.400–407
166. Rojas A., et al, (2008)," Detection of masses in mammograms via statistically based enhancement, multilevel-thresholding segmentation, and region selection", Comput. Med. Imaging Graph., pp. 304-315
167. Rosin L., (2009), "Classification of pathological shapes using convexity measures", Pattern Recognition, Lett., pp. 570-578
168. Rady H., (2011), "Face recognition using principle component analysis with different distance classifiers, genetic algorithms and Ada Boost classifier", International Journal of Computer Science and Network Security, vol. 11, pp. 134–144.
169. Riyahi alam N., et al.,(2015)," Computer-aided mass detection on digitized mammograms using a novel hybrid segmentation system", Int Biol Biomed Eng PP.51–58
170. Samuel J., et al., (1998)," Illustrated Breast Imaging Reporting and Data System (BIRADS TM)", American College of Radiology (ACR), Reston, P.325
171. Somol P., et al, (1999), "Adaptive floating search methods in feature selection", Pattern Recognition Letters, vol. 20, pp. 1157-1163.
172. Samson S., et al, (2000), "A variational model for image classification and restoration", IEEE Trans. Pattern Anal. Mach. Intell. PP.460–472.
173. Smith L., (2002)," A tutorial on Principal Components Analysis", PP.56-71.
174. Sun Y, et al, (2002), "Normal mammogram classification based on regional analysis", in: Proceedings of the IEEE Midwest Symposium on Circuits and Systems, vol. 45, pp. 375–378.
175. Soltanian-Zadeh H., et al, (2004)," Comparison of multiwavelet, wavelet, Haralick, and shape features for microcalcification classification in mammograms", Pattern Recognition, pp. 1973-1986
176. Skaane P, et al, (2004)," Screen-film Mammography versus full-field digital Mammography with soft-copy reading: randomized trial in a population-based screening program-the Oslo II study.", Radiology PP.197–204
177. Sampat M., et al, (2005), "Classification of mammographic lesions into BI-RADS shape categories using the Beamlet Transform", Medical Imaging: Image Processing, Proceedings of the SPIE, vol. 5747, pp. 16–25.
178. Sun Y., et al, (2005), "A Comparison of Feature Selection Methods for the Detection of Breast Cancers in Mammograms: Adaptive Sequential Floating Search vs. Genetic Algorithm.", Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference Shanghai, China, P.12
179. Sampat, M., et al, (2005) "Classification of mammographic lesions into BI-RADS shape categories using the Beamlet Transform, Medical Imaging: Image Processing", Proc. of the SPIE, Vol. 5747, No. 16, pp.16–25.
180. Sheshadri S., et al, (2006)," Breast tissue classification using statistical feature extraction of mammograms", Med. Imaging Inform. Sci., pp. 105-107

181. Sheshadri H., et al,(2007),” Experimental investigation on breast tissue classification based on statistical feature extraction of mammograms”, *Computerized Medical Imaging and Graphics*, no.31, pp.46-48.
182. Skaane P, et al, (2007),” Randomized trial of screen-film versus full-field digital Mammography with soft-copy reading in population-based screening program: Follow-up and final results of Oslo II study”, *Radiology* PP.708–717
183. Samulski M., et al, (2010),” computer-aided detection in Mammography as a decision support, *Eur. Radiol*”, pp. 2323-2330
184. Sohail A, et al, (2011),” Classification of ultrasound medical images using distance based feature selection and fuzzy-SVM”, P.22
185. Surendiran B., et al, (2011),” Feature Selection using Stepwise ANOVA Discriminant Analysis for Mammogram Mass Classification”, *ACEEE Int. J. on Signal & Image Processing*, Vol. 02, No. 01
186. Shirbani F. et al, (2016),” Fast SFFS-Based Algorithm for Feature Selection in Biomedical Datasets”, *Amirkabir International Journal of Science& Research (Electrical & Electronics Engineering) AIJ-EEE*, PP.43-56.
187. Salabat Khan, et al, (2017), “A comparison of different Gabor feature extraction approaches for mass classification in Mammography”, *Multimed Tools Appl*, PP.76:33–57.
188. Song Yuheng, et al, (2017), “Image Segmentation Algorithms Overview”, *SiChuan University*, P.124
189. Salama M., et al , (2018),“An Improved Approach for Computer-Aided Diagnosis of Breast Cancer in Digital Mammography”, *IEEE International Symposium on Medical Measurements and Applications (MeMeA)*,INSPEC Accession Number: 18027668 , Rome, Italy,p.12
190. Sahiner B., et al, (2018), “Computer-aided characterization of mammographic masses: accuracy of mass segmentation and its effects on characterization”, *IEEE Trans. Med. Imaging*, Vol. 20, No. 12, pp. 1275–1284.
191. Torheim G, et al, (2001), “Feature extraction and classification of dynamic contrast-enhanced T2*-weighted breast image data”, *IEEE Trans Med Imaging* 20, PP.1293–1301.
192. Tzikopoulos St., et al., (2010), “Shape-based tumor retrieval in mammograms using relevance-feedback techniques”, *Proceedings of the 20th International Conference on Artificial Neural Networks*, Thessaloniki, Greece. pp. 201-215
193. Taghanaki S., et al, (2017), “Pareto-optimal multi-objective dimensionality reduction deep auto-encoder for Mammography classification”, *Comput. Methods Progr. Biomed.* P. 85–93.
194. Ting F., (2019),” Convolutional neural network improvement for breast cancer classification”, *Expert Syst.* PP. 103–115
195. Vese L., et al, (2002),” A multiphase level set framework for image segmentation using the Mumford and Shah model”, *Int. J. Comput. Vis.* PP. 271–293.
196. Vibha L., et al, (2006),” Classification of Mammograms Using Decision Trees”, In: *10th International Database Engineering and Applications Symposium (IDEAS'06)*. PP.263-266

197. Vadivel A., et al,(2013),” A fuzzy rule-based approach for characterization of mammogram masses into BI-RADS shape categories”, *Computers in Biology and Medicine*, Volume 43, Issue 4, PP. 259-267
198. Virmani J., et al, (2016), “PCA-PNN and PCA-SVM Based CAD Systems for Breast Density Classification”, Springer International Publishing, Warsaw, Poland, pp. 159–180.
199. Vijikala V., et al, (2016), “Identification of most preferential de-noising method for mammogram images“, *Conference on Emerging Devices and Smart Systems (ICEDSS)*
200. Wei D, et al, (1995), “Classification of mass and normal breast tissue on digital mammograms: multiresolution texture analysis”, *Med Phys*, PP.1501–1513.
201. Warren LJ., et al, (2000), “Potential contribution of computer-aided detection to the sensitivity of screening Mammography”, *Radiology* PP.554–562
202. Wu T, Stewart, et al, (2003),” Tomographic Mammography using a limited number of low-dose cone-beam projection images”, *Med Phys* PP.365–380
203. Wei C., et al, (2005), “A general framework for content-based medical image retrieval with its application to mammograms”, *Proceedings of the SPIE International Symposium on Medical Imaging*, PP. 134–143.
204. Wei C., et al, (2005), “A general framework for content-based medical image retrieval with its application to mammograms”, in: *Proceedings of the SPIE International Symposium on Medical Imaging*, pp. 134–143.
205. Wang X., et al, (2010), “An efficient local Chan–Vese model for image segmentation”, *Pattern Recognit.* PP.603–618.
206. Wang X., et al, et al, (2016), “Discrimination of breast cancer with micro calcifications on Mammography by deep learning”, *Scientific reports*, P.6
207. Xinsheng Zhang, (2009), “Boosting twin support vector machine approach for MCs detection”, *Proceedings of the Asia-Pacific Conference on Information Processing*, pp. 149–152.
208. Xu C, (2010), “Snakes, Shapes, and Gradient Vector Flow”, *IEEE Transactions on Image Processing*, vol. 7(3), pp. 359-369
209. Yang J.,et al, (1998), “Feature subset selection using a genetic algorithm”, *IEEE Intelligent Systems*, vol. 13, no. 1, pp. 44-49.
210. Yang S., et al, (2005), “A Computer-aided system for mass detection and classification in digitized mammograms”, *Journal of Biomedical Engineering- Applications, Basis and Communications*, Vol. 17, pp. 215-228.
211. Yufeng Z, (2010), “Breast cancer detection with Gabor features from digital mammograms”. *Algorithms*, P .44–62.
212. Yeh W.,et al, (2015), “Implementation of Practical Computer Aided Diagnosis System for Classification of Masses in Digital Mammograms”, *International Conference on Computing, Control, Networking, Electronics and Embedded Systems Engineering*, PP. 2769-2779
213. Zheng B., et al, (1999), “Application of a Bayesian belief network in a computer-assisted diagnosis scheme for mass detection”, *Proc. SPIE*, PP.1553–1561
214. Zheng B., et al, (1999), “Feature selection for computerized mass detection in digitized mammograms by using a genetic algorithm,” *Acad. Radiol.* PP.1016-1076

215. Zaiane, O., et al, (2002) "Mammography classification by an association rule-based classifier", in Proc. of the Third International Workshop on Multimedia Data Mining, pp.62–69.
216. Zoetelief J, et al, (2003)," European protocol on dosimetry in Mammography", Publication EUR 16263 EN, Brussels, Belgium: European Commission
217. Zhang X., et al ,(2009)," detection with combined image features and twin support vector machines", J. Comput, pp. 215-221
218. Zhao W., et al, (2015)," Subspace Linear Discriminant Analysis for Face Recognition", CAR-TR-914 N00014-95-1

Abbreviations

Maximum Intensity Images (MIP)

Short TI Inversion Recovery (STIR)

Breast Magnetic Resonance Imaging (MRI)

Cranial-Caudal (CC)

Medio Lateral-Oblique(MLO)

Upper Outer Quadrant(UOQ)

Upper Inner Quadrant(UIQ)

Lower Outer Quadrant(LOQ)

Lower Inner Quadrant (LIQ)

Digital Imaging and Communications in Medicine(DICOM)

Digital Data base for Screening Mammography (DDSM)

Sequential Floating Forward Feature Selection algorithm(SFFS)

Adaptive Sequential Floating Forward Feature Selection algorithm (ASFFS)

Simple Genetic Algorithm (SGA)

Linear Discrimination Analysis Algorithm (LDA)

Fuzzy C- Mean(FCM)

Support Vector Machine(SVM)

Support Vector Machine Recursive Feature Elimination (SVM-RFE)

Quadric Discriminant Analysis(QDA)

Minimum Redundancy and Maximum Relevance(MRMR)

Normalized Mutual Information Feature Selection(NMIFS)

Grey Level Co-occurrence Matrix (GLCM)

Grey Level Run Length Matrix(GLRLM)

Magnetic Gabor Response-LDA(MGR-LDA)

Statistic MGR(SMGR)
Window Based First Order SMGR(WSMGR)
Medical Applications on a Grid Infrastructure Connection(MAGIC-5)
Mammographic Image Analysis Society Images (MIAS)
Local Temporary Pattern (LTP)
Successive Enhancement Learning Based Weighted SVM (SELwSVM)
Artificial Neural Network(ANN)
Discrete Wavelet Transform (DWT)
ANOVA Discriminant Analysis (ANOVA DA)
Association Rule based Classification by Category(ARC-BC)
Spatial Gray Level Dependency(SGLD)
Mean Square Error(MSE)
Peak Signal to Noise Ratio(PSNR)
Region Of Interest(ROI)
Cluster Prominence(CPROM)
Sum of Variance(SVARH)
Cluster Shade(CSHAD)
Sum of squares(SOSVH)
Difference Variance(DVARH)
Sum Average(SAVGH)
Maximum Probability(MAXPR)
Sum Entropy(SENTH)
Autocorrelation(AUTOOC)

Information Measure of Correlation²(INF²H)

Information Measure of Correlation¹(INF¹H)

Correlation(CORRM)

Difference Entropy(DENTH)

Standard Deviation(STD)

Normalized Histogram (NH)

Under the ROC Curve(AUC)

Principal Component Analysis (PCA)

Classification and Regression Tree(CART)

K Nearest Neighbor(KNN)

Convolutional Neural Network(CNN)