PREPROCESSING OF MAMMOGRAPHY IMAGE FOR IMPROVING ACCURACY& EFFICIENCY OF CAD IN EARLY DETECTION OF BREAST CANCER

A THESIS SUBMITTED TO THE GRADUATE SCHOOL OF APPLIED SCIENCES OF NEAR EAST UNIVERSITY

By
RAHEL DESTA FISSEHA

In Partial Fulfillment of the Requirements for the Degree of Master of Science in Biomedical Engineering

NICOSIA, 2020

RAHEL DESTA FISSEHA

PREPROCESSING OF MAMMOGRAPHY IMAGE FOR IMPROVING THE ACCURACY AND EFFICIENCY OF CAD IN EARLY DETECTION OF BREAST CANCER

NEU 2020

PREPROCESSING OF MAMMOGRAPHY IMAGE FOR IMPROVING ACCURACY& EFFICIENCY OF CAD IN EARLY DETECTION OF BREAST CANCER

A THESIS SUBMITTED TO THE GRADUATE SCHOOL OF APPLIED SCIENCES OF NEAR EAST UNIVERSITY

By
RAHEL DESTA FISSEHA

In Partial Fulfillment of the Requirements for the Degree of Master of Science in Biomedical Engineering

NICOSIA, 2020

RAHELDESTAFISSEHA: PREPROCESSING OF MAMMOGRAPHY IMAGE FOR IMPROVING THE ACCURACY AND EFFICIENCY OF CAD IN EARLY DETECTION OF BREAST CANCER

Approval of Director of Graduate School of Applied Sciences

Prof. Dr. Nadire Çavuş

We certify that this thesis is satisfactory for the award of the degree of Master of Science in Biomedical Engineering

Examining Committee in charge:

Prof. Dr. Ayşe Günay Kibarer — Supervisor, Head of the Department of Biomedical Engineering, NEU

Asst. Prof. Dr. Elbrus Imanov — Computer Engineering, NEU

Asst. Prof. Dr. Boran Şekeroğlu — Information Systems Engineering, NEU

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by the rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Rahel Desta, Fisseha

Signature:

Date: 24/08/2020

ACKNOWLEDGEMENT

My sincere appreciation goes to Prof. Dr. Ayşe Günay Kibarer, Head of the Department of Biomedical Engineering, Near East University, for her encouragement and comprehensive advices throughout my stay and my thesis work, and thank you for being part of my story, you are truly an inspiration. Again, from the bottom of my heart, I thank you and appreciate all you have done.

My husband and my lovely little baby boy, you have my life in such a positive energy that words cannot express my appreciation and love. You are truly an inspiration.

Last but not least, I would like to thank Near East University for the scholarship, Ethiopian government for the sponsorship and my friends living here in Near East University and Ethiopia for their support and care.

To my parents…

ABSTRACT

Breast cancer is a global health malice which causes the second largest cancer related mortality among women worldwide. Its earliest possible detection, therefore, is of vital importance in the reduction of the associated fatality and increment of survival rate with high probability of full recovery. Mammography is among the infamous contemporary imaging modalities for early visualization of breast abnormalities. However, it's usually suspected to different background artifacts, unknown noises and poor image contrast which negatively influence in interpretation and detection of small abnormalities in automated CAD system for mammograms. Extracting ROI for the accuracy of CAD is challenging task because of existing of pectoral muscle in digital mammograms. In this paper, an effective preprocessing technique which removes irrelevant and unwanted background artifacts and pectoral muscles is proposed. The proposed algorithm suppresses background artifacts using thresholding method and morphological operation. Then, pectoral muscle accurately segmented using automated seeded region growing method. Finally, mammography image quality enhanced using median filter and CLAHE. The proposed method tested on mini-MIAS database and works effectively. The accuracy of pectoral muscle segmentation is evaluated by correctness and completeness and was reported as 97% and 96% respectively. Image quality assessment metrics of PSNR, MSE and SNR against mean, wiener and median filters had been calculated and found Median filter and CLAHE precious result in mammography image quality enrichment. The result of this study shows that, the proposed method definitely suits for enhancement of mammogram image quality and extraction ROI in improving accuracy and efficiency of auto-CAD in early detection of breast cancer.

Keywords: ROI; breast cancer; mammography; preprocessing; CAD; CLAHE; region growing

ÖZET

Göğüs kanseri küresel bir sağlık problemi olup, dünya çapında kadınlar arasında ölüme yol açan ikinci en büyük kanser türlerinden biridir. Dolayısı ile mümkün olan en erken teşhis, ilgili ölümlerin azaltılması ve yaşam ömrünün yüksek olasılıkla kesin iyileşme ile uzatılması için hayati önem taşımaktadır. Göğüsteki anormalliklerin erken görüntülenmesinde mamografi çağdaş görüntüleme yöntemlerinden biridir. Ancak, arkaplandaki yapılar, bilinmeyen sesler ve zayıf görüntü kontrastı, mamogramlar için geliştirilen otomatikleştirilmiş CAD sistemlerinde ufak anormalliklerin yorumlanması ve teşhisinde negative etkiler yaratır. CAD sisteminin kesinlik kazanabilmesi için sayısal (dijital) mamogramlarda ROI metodu ile ayıklama işlemi göğüs kasının varlığına bağlı olarak oldukça zorlu bir külfettir. Bu tez çalışmasında, ilgisiz ve istenmeyen arkaplan yapıların ve göğüsle ilgili kasların görüntüden temizlenebilmesi için kullanılacak etkin bir önişlem tekniği önerilmektedir. İleri sürülen algoritma, ikili görüntünün siyah ve beyaz olarak görüntülenmesi metodunu ve morfolojik görüntü işleme yöntemini kullanarak arkaplan yapıları ortadan kaldırmayı amaçlamaktadır. Daha sonrasında, göğüs kası kesin bir şekilde otomatik tohumlandırmalı bölge büyütme metodu kullanılarak bölünür ve ayrılır. En son olarak, mamografi imge kalitesi medyan filtresi ve histogram eşitlemesinin ayrık dalgacık dönüşümü (CLAHE) kullanılarak arttırılır. Önerilen metot mini-MIAS databaz üzerinde denenmiş ve etkin olarak çalıştığı gözlenmiştir. Göğüs kası segmantasyonunun kesinliği, doğruluk ve bütünlük üzerinden değerlendirilmiş ve sırası ile %97 ve %96 olarak bildirilmiştir. Görüntü kalitesi değerlendirme metriklerinden PSNR, MSE ve SNR a karşın ortalama, wiener ve medyan filtreleri hesaplanmış ve Medyan Filtre ve CLAHE' nin mamografi imge kalitesini zenginleştirme açısından fevkalade sonuçlar verdiği gözlenmiştir. Bu çalışmanın sonuçları göstermektedir ki, önerilen metot ve ROI (ilgi duyulan alan) ayıklaması, mammogram görüntü kalitesini ve göğüs kanserinin erken teşhisinde auto-CAD'in kesinliği ve etkinliğini geliştirmede kullanılmak için gayet uygundur.

Anahtar kelimeler: ROI; göğüs kanseri; mamografi; önişlem; CAD; CLAHE; otomatik tohumlandırmalı bölge büyütme

TABLE OF CONTENTS

ACKNOWLEDGEMENT i

ABSTRACT iii

ÖZET iv

LIST OF TABLES viii

LIST OF FIGURES ix

LIST OF ABBREVIATIONS xi

CHAPTER 1 : INTRODUCTION 1

1.1 Statement of the Problem 2

1.2 Aim of the Study 3

1.3 The Importance of the Study 3

1.4 Goals and Objectives 4

1.4.1 General objectives 4

1.4.2 Specific objectives 4

1.5 Thesis outline 4

CHAPTER 2: LITERATURE REVIEW 5

2.1 Mammographic Anatomy of Human Breast 5

2.2 Common Problems in Breast 7

2.3 Developmental Anomalies in Breast 8

2.4 Inflammatory Diseases 9

2.5 Breast Cancer 10

2.6 Diagnosis and Screening of Breast Cancer 11

2.7 Basics of Mammography 12
2.8 Advances of Imaging Test in Diagnostic Breast Cancer 13
2.9 Limitations of a Mammography 13
2.10 Related Works 14
CHAPTER 3: MATERIALS AND METHODS 17
3.1 Dataset Used 17
3.2 Methodology 19
3.2.1 Proposed Method 19
3.2.2 Reduction of artifacts and separation of background 20
3.2.3 Pectoral muscle segmentation 22
3.2.3.1 Image segmentation 22
3.2.3.2 *Region growing* 23
3.2.4 Image enhancement 26
3.2.4.1 Contrast improvement 26
3.2.4.2 Noise reduction 27
CHAPTER 4: RESULTS AND DISCUSSIONS 28
4.1 Accuracy Performance Evaluation 30
4.2Quality assessment metrics 36
4.2.1Performance of image quality based on MSE 36
4.2.2Performance of Image Quality Based on PSNR 37
4.2.3Performance of Image Quality based on SNR 38
CHAPTER 5: CONCLUSION AND RECOMMENDATION 41
5.1 Conclusion 41

5.2 Recommendations..42

REFERENCES 43

APPENDIX …………………………………………………….………………………... 48

APPENDIX 1: Ethical Approval letter……………………………….………………………………... 48

APPENDIX 2: Similarity Report -TURNITIN (Consultant signed) ……………………...……... 49

LIST OF TABLES

Table 4.1: shows number of properly segmented, over segmented and under segmented among the randomly selected mini-MIAS images for the proposed algorithm 31

LIST OF FIGURES

Figure 2. 1: Schematic representation of functional anatomy of mammary Gland.................... 6
Figure 2. 2: Appearance of normal Mammograms MLO and CC 7
Figure2. 3: Breast cancer incidence worldwide 10
Figure2. 4: stage of breast cancer, tumor size and survival rate... 12

Figure 3. 1: sample mammography image from mini-MIAS database 18
Figure 3. 2: block diagram of proposed method.. 20
Figure 3. 3: Sample mini-MIAS database image that contains many background objects and artifacts 21
Figure 3. 4: classification of image segmentation techniques... 23
Figure 3. 5: Flow chart of proposed automated seeded region growing method 25

Figure 4. 1: shows the result of removal artifacts and separation of background,.................. 29
Figure4.2: sample result of pectoral muscle segmentation using proposed algorithm………..30
Figure 4. 3: shows sample mini-MIAS images with pectoral muscle intensity value similarity or lower than the remain breast profile... 32
Figure 4. 4: shows result of image quality enhancement and different filters 33
Figure 4. 5: shows experimental result of proposed method on mdb028 reference number of MIAS with Fatty (F)breast tissue type and Malignant (M) 34
Figure 4. 6: shows experimental result of proposed method on mdb048 reference number of MIAS with Fatty-glandular(G) breast tissue type and Normal(N)... 34
Figure 4. 7: shows experimental result of proposed method on mdb058 reference number of MIAS with Dense-glandular (D) breast tissue type and Malignant (M) 35
Figure 4. 8: shows experimental result of proposed method on mdb199 reference number of MIAS with Dense-glandular(D) breast tissue type and Benign(B).. 35

Figure 4. 9: MSE result on comparison of different filters 37
Figure 4. 10: PSNR result on comparison of different filters.. 38
Figure 4. 11: SNR result on comparison of different filters.. 39

LIST OF ABBREVIATIONS

2D:	Two dimensions
AHE:	Adaptive histogram equalization
B:	Benign
CAD:	Computer aided detection/diagnosis
CC:	Craniocaudal
CLAHE:	Contrast limited adaptive histogram equalization
D:	Dense-glandular
F:	Fatty
FN:	False negative
FP:	False positive
G:	Fatty-glandular
MIAS:	Mammographic image analysis society
MLO:	Mediolateral oblique
MRI:	Magnetic resonance imaging
MSE:	Mean square error
PET:	Positron emission tomography
PSNR:	Peak signal to noise ratio
ROI:	Region of interest
SNNs:	Supernumerary nipples
SNR:	Signal to noise ratio
T:	Threshold
TP:	True positive
WHO:	World health organization

CHAPTER 1

INTRODUCTION

Of all cancer in women, breast cancer ranks with high prevalence rate which causes the second greatest global mortality among females. According to American Cancer Society, about 1 in 8 woman develops breast cancer in her lifetime and only 5 – 10 percent of these are attributed to genetic causes (NCI2019; WHO2018). Its earliest detection and timely screening, therefore, are very dependable techniques to reduce the associated fatalities and increase the number of available treatment options which leads to higher probability of full recoveries and increased survival. With low x-ray radiation dose and excessive dynamic range, mammography is among the best and widely adopted imaging modality employed for visualization of internal structure of breast abnormalities (Bilimoria et al. 2008; Halalli and Makandar 2018).

On the average, Mammography detects about 80 – 90 percent of breast cancer for asymptomatic and clinically unsuspected women. Timely screening and early breast cancer detection able to diminish mortality rate up to 25%. However, mammography image is difficult to interpret and usually they are suspected to artifacts and noises. Furthermore; due to low x-ray dose radiation, their image contrast intensity is poor. This negatively influence medical images to detect and understand breast cancers at primary stage (Heath et al. 1998; Kerlikowske et al. 1995).

In screening mammography, low specificity property may cause some unnecessary biopsy. Researches indicate that about 10% - 30% breast radiologist fail to recognize during routine screening. Therefore, standardization of mammogram images is essential to limit the search for abnormalities (George and Dhas 2017). In extraction region of interest (ROI), chest wall location i.e. pectoral muscle presents a big challenge in the pre-processing and it influences the detection process. The CAD system detection procedures could be biased by the presence of artifacts and pectoral muscle (Bird, Wallace, and Yankaskas 1992; Pitas 2000) and It affects the accuracy of extracting ROI in intensity-based image processing. Thus, pre-processing presents a decisive role in an automated computer aided diagnostics system for mammography images (Halalli and Makandar 2018).

Preprocessing of mammography images is turned as vital step in image processing. As for the rest CAD steps like segmentation, feature extraction and classification, their probability of success is determined by the accuracy of the pre-processing. Unknown noise, inhomogeneity, weak boundaries, poor image contrast and unrelated parts are common inherited characteristic of medical images and they affect their clinical imaging essences. These issues can be redressed by effective preprocessing algorithms. Preprocessing involves a score of activities which include image re-sampling, grey scale contrast enhancement, noise removal, mathematical operation and manual correction (George and Dhas 2017; Halalli and Makandar 2018; Makandar and Halalli 2016).

In this research work, a preprocessing algorithm for enhancing mammography image quality, which reduce background artifacts, suppress unknown noises and segments the pectoral muscle is proposed. It is divided into three main steps: in the first step, Separation of Backgrounds and removal of Artifacts by using thresholding method and morphological operation techniques respectively; in the second step, segmenting the pectoral muscle using automatic seeded region growing method; and the third step is enhancing image quality by using median filter and Contrast limited adaptive histogram equalization (CLAHE).

1.1 Statement of the Problem

Breast cancer is global health malice and most common in women around the world. It is the most fatal in developing countries and the second leading in America for cancer related mortality (Breast Cancer Research Foundation,2018). According to WHO statistics, around 2.09 million breast cancer cases were diagnosed in 2018. X-ray mammography is commonly used image modality in early stage detection of breast cancer, and thus mammograms are conducted regularly in breast screening. Hence, timely screening, early detection and accuracy of image interpretation plays a great role in reducing mortality, enhances successful treatment options and increase probability of complete recovery.

Comparing to other image modality, mammography images are difficult to interoperate and breast radiologists fail to understand and detect breast cancer at early stage that of poor image quality and background artifacts of the mammogram. This leads to unnecessary biopsies and some breast cancers are missed. So, mammography images need effective preprocessing procedure to overcome such difficulties (Lattanzio and Simonetti 2010).

The ultimate goal of this thesis is to enhance image quality of screened mammography breast images thereby it also suppresses background artifacts and reduce unknown noises and enhance the intensity contrast for magnify small image details which is essential for early detection of women breast cancer at primary level. Furthermore, pectoral muscles would be extracted for accuracy and efficient segmentation of region of interest and for simplifies interpretation.

1.2 Aim of the Study

To work on effective preprocessing algorithm for mammography breast images that helps radiologist and doctors to diagnose and identify the breast abnormality early and easily.

1.3 The Importance of the Study

- The proposed algorithm helps radiologist to diagnosis the disease more accurately and takes decision swiftly.
- It assists breast radiologist to detect breast abnormalities at early stage
- It plays great role in minimize the possibility of false negative
- It benefits women for higher quality of life and economically treated thereby increase survival rate
- It increases numbers of available treatment options, thus increase the probability of fully curried from breast cancer and decrease mortality from breast cancer
- It helps to interpret breast images correctly and conspicuously

1.4 Goals and Objectives

1.4.1 General objectives

- To enhance mammography image quality and to extract ROI for increasing the accuracy and efficiency of CAD system in early breast cancer detection

1.4.2 Specific objectives

- To enhance contrast intensity of breast image in order to better distinguish details contained in the image
- To suppress unknown noise by preserving detail of the image
- To reduce background artifacts &suppress image distortion
- To improve the detection and accuracy in extraction region of interest by segmenting pectoral muscle to increase the performance of CAD system

1.5 Thesis outline

Chapter 1 provided a general information, aim, importance, general objective and specific objective of the thesis. In chapter 2, literature reviews on Mammographic Anatomy of Human breast, common problems in breast, diagnosis and screening of breast cancer, limitation of mammography and review of related works. Chapter 3 describes the database used and methodology parts. Chapter 4 explains the result and discussion portion of the paper, finally in Chapter 5 conclusion and recommendation are stated.

CHAPTER 2
LITERATURE REVIEW

2.1 Mammographic Anatomy of Human Breast

Breasts are known to be an organ containing exocrine glandular tissue specialized milk production and secretion important for new born. It is located immediate to the pectoral fascia anterior to the pectoralis major muscle. Its regional position lies from 2^{nd} to 6^{th} rib extending medially to outer margin of sternum and laterally anterior axillary line. These accessory glands do exist in both sexes with similar architecture and differ later due to effect of sex hormones testosterone and estrogen in male and female respectively, effect evident on size enlargement in females during puberty onwards. As an organ the breast consists of different tissues with synthesis secretory and support functions. The mammary gland (*glandulamammaria*) consists of 15-20 tubule alveolar lobes that further divide and give rise to smaller lobules. From lobules each milk secretory unit of alveolar is found. Structurally duct system from each alveolar space will join to form bigger duct system draining each lobule and converging to form duct system draining each lobe of mammary gland referred as ductus lactiferus (Ɵ=2-4mm). Ductus lactiferus from group of lobes will then form wider sinuses referred as sinus lactiferus which opens to the nipple through constricted openings called poruslactiferus (Ɵ=0.4-0.7mm).The supportive component comprises of the connective tissue and the subcutaneous adipose tissue (Pohlodek 2014).

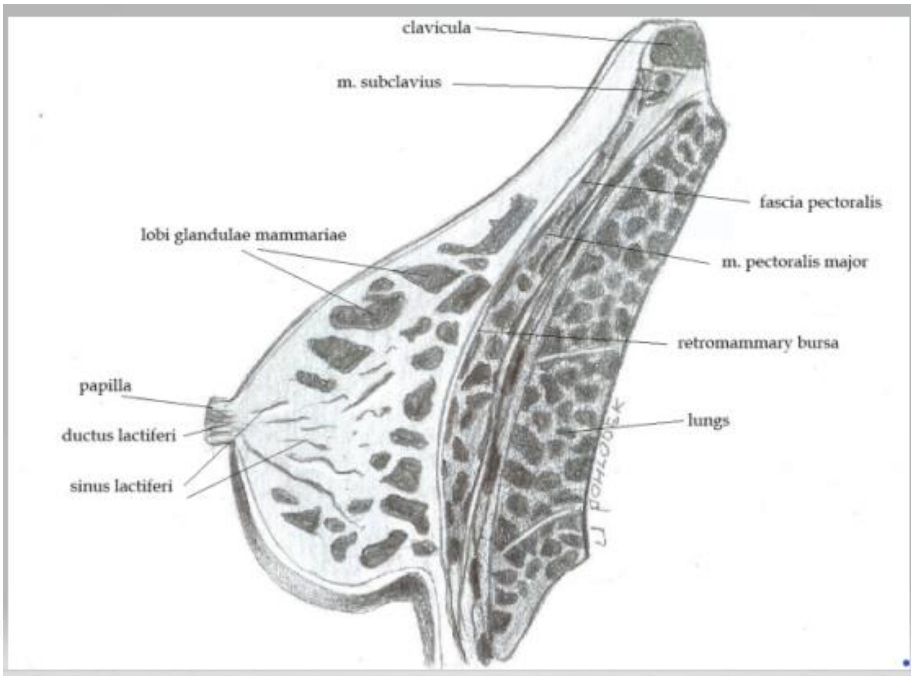


Figure2. 1: Schematic representation of functional anatomy of mammary Gland (Pohlodek 2014).

Breast is supplied by mammary arteries branching from internal thoracic artery and drained by dense network of venous and lymphatic drainage passing through axilla to internal thoracic vessels. Abnormalities in mammary gland follows drainage roots in their advance and progression which may be reflected through enlargement of nearby lymph nodes around the mammary gland such as axillary lymph nodes. Therefore, changes in size of axillary lymph nodes is used as tool for tentative diagnosis for problems in breast tissue. The outer skin, axial bony skeleton muscles and ligaments provide support to the breast mass (Drake, Vogl, and Mitchell 2009; Wineski 2018; Lattanzio and Simonetti 2010).

The mammographic anatomy is represented by the corpus mammae (the main body of breast with glandular tissue and its supporting fibrous tissue), the Chassaignac'sbursa (the discreet space between breast and pectoralis Fascia referred as retro-mammary bursa), crests of the Duret (attachments of outer lobes to fascia) and the anterior nipple areolar complex. With anterolateral pectoralis major muscle fold (Indicated in Figure) [Lattanzio, 2010].

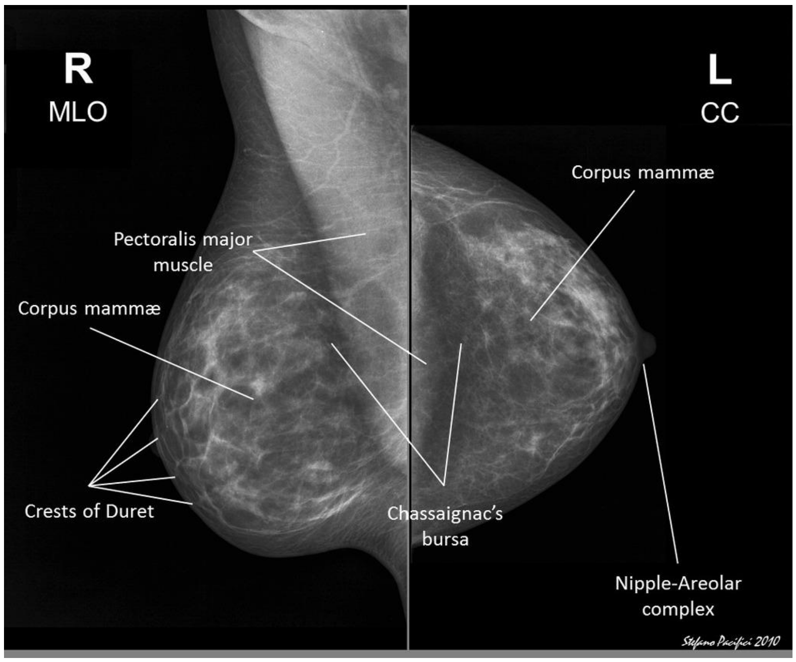


Figure2. 2: Appearance of normal Mammograms MLO and CC (Lattanzio and Simonetti 2010).

The functional milk secretory tissue of mammary gland in breast contains 12-15 lobes arranged in radial fashion and converging ducts finally open at tip of the nipple. Physiologically breast assumes different shapes and sizes varying from individuals to individuals technically termed as flat and shallow, hemispheroid, pear shaped and pendulous breast(Pohlodek 2014).

2.2 Common Problems in Breast

Many diseases and abnormalities of the breast has been described and documented. Both developmental and acquired anomalies has also been demonstrated in literatures(De Silva and Brandt 2006a; De Silva and Brandt 2006b). Among the most common observations that seek medical attention presence of gross palpable masses, painful conditions around the breast and observations for any nipple discharge has been described (Salzman, Collins, and Hersh 2019).

Although may not be always breast masses are associated with breast cancer. Palpable breast mass presentation need to be examined thoroughly with adequate history, clinical diagnosis accordingly can be supported with diagnostic imaging techniques (Salzman, Collins, and Hersh 2019).

2.3 Developmental Anomalies in Breast

Among the most common developmental anomalies, Poland syndrome a presentation representing thoracic wall malformation and absence or hypoplastic pectoralis muscle and breast related to abnormal shortness of the upper limb commonly referred as brachy syndactyly has been indicated. It takes its name after it is characterized by Alfred Poland in 1841. The syndrome is known to have clinical presentation of malformation or devoid of ribs, pectoralis muscles partial or total absence , specially the sterno-costal head end of the pectoralis major, resulting in the absence of the upper axillary fold (Shamberger, Welch, and Upton 1989). Supernumerary nipples (SNNs) referred commonly polythelia, is also an important presentation to be considered for preponderance for hormonal management. It may develop abnormalities like a normal breast such as breast cancer. Supernumerary teat appear along milk line of the embryonic development, on the back, shoulder, thigh, face, or vulva (Caouette-Laberge and Borsuk 2013).

The other common anomaly reported was polymastia, the clinical presentation of having more than the expected number of breasts, similarly occurring along the embryonic milk line. Both polythelia and polymastia are related to familial inheritance (Casey, Chasan, and Chick 1996). Contrary to polythelia, polymastiamay be related to anomalies around thoracic and renal bodies (Caouette-Laberge and Borsuk 2013). Polymastia results due to developmental anomalies when embryonic mammary ridge fail to undertake normal regression(Bland et al. 2017).

Athelia/Amastia,a term used to denote absence of nipple and/or functional component of breast tissue though rare occurrence have different presentations associated to its underlying pathologies. Ectodermal embryonic defects may cause absence of both right and left breasts in males and females as related to the ectodermal layer and its appended ages developmental defect. Treatment modalities of amastia and atheliacan be similar to post-oncologic breast reconstruction with attention given to the placement of the inframammary fold and nipple(Caouette-Laberge and Borsuk 2013).

2.4 Inflammatory Diseases

Various forms of breast inflammatory disorders have been encountered in imaging centers. Breast Inflammatory abnormalities constitute a wide range of breast disorders based on responsible causes, from benign infectious to non-infectious and inflammation resulting breast malignancy. Because of difficulties to diagnose mastitis and breast cancer on clinical presentation alone, imaging features may be benefic for better management of breast disorders (Leong, Chotai, and Kulkarni 2018).Mastitis is a common name given to diverse abnormalities characterized by inflammatory reaction in breast tissue (Kamal, Hamed, and Salem 2009; Lepori 2015).

Being as a superficial organ the clinical presentation of breast inflammation include common cardinal signs of inflammation heat, pain and redness as well as skin thickening. Historical events in clinical development of the inflammatory process should be thoroughly taken from the patient. Inflammation that has progressively development is regarded as atypical. Breast inflammation can arise from causes of infectious or non-infectious nature, with possibility also from breast cancer. The possibility of involvement of lactiferous duct abnormalities should be diagnosed. Involvement of lymphadenopathy should also be thoroughly thought in examining inflammatory breast disease (Lepori 2015).

2.5 Breast Cancer

Abnormal tissue proliferation related to breast are becoming the common cause of oncologic health problem and death in women globally(Becker and Obstetrics 2015). Earlier reports have also compared the relative importance of breast cancer in European countries accounting for 350000 annual cases in European women *(Barth 2011)* and worldwide as indicated in Figure 2.3.

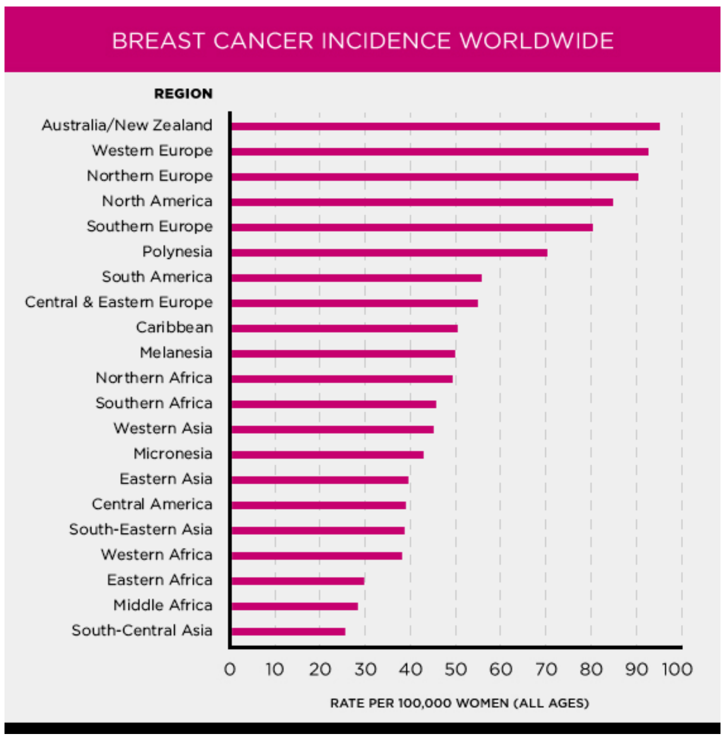


Figure2. 3: Breast cancer incidence worldwide (IARC and WHO 2018).

In depth knowledge of the breast cancer molecular level characteristics related to subtypes of breast cancer is important to shape the process of diagnosis, prognosis and treatment of breast cancer in the future. The different categories of breast cancer were described as invasive(infiltrating)carcinoma and noninvasive (in situ carcinoma) based on mobility to adjacent tissues. In situ carcinoma is also further classified as ductal and lobular based on growth patterns and cytological features. Ductal in situ carcinoma is more common than lobular in situ carcinoma. Ductal in situ carcinoma is further sub-classified to five well characterized categories referred as Cribriform, Comedo, Micropapillary, Solid and Papillary based on specific architectural features(Lester et al. 2009). Similarly, invasive carcinomas are also subclassified as infiltrating ductal, ductal/lobular, invasive lobular, mucinous (colloid), tubular, medullary and papillary carcinomas. Among all subcategories, infiltrating ductal carcinoma (IDC) is most common subtype. All these classifications depend on histological sections (Malhotra et al. 2010).

2.6 Diagnosis and Screening of Breast Cancer

It is well characterized and evidenced that early detection of breast cancer is important factor for survival of patients. Hence well-organized screening and diagnostic schemes is inevitably important for increasing life span of patients. In this case four of the following procedures lead to the efficient detection of breast cancer as early as possible: (1) Thorough clinical examination and case history analysis; (2) mammography; (3) ultrasonography; and (4) MRI of breast mass (Becker 2015).

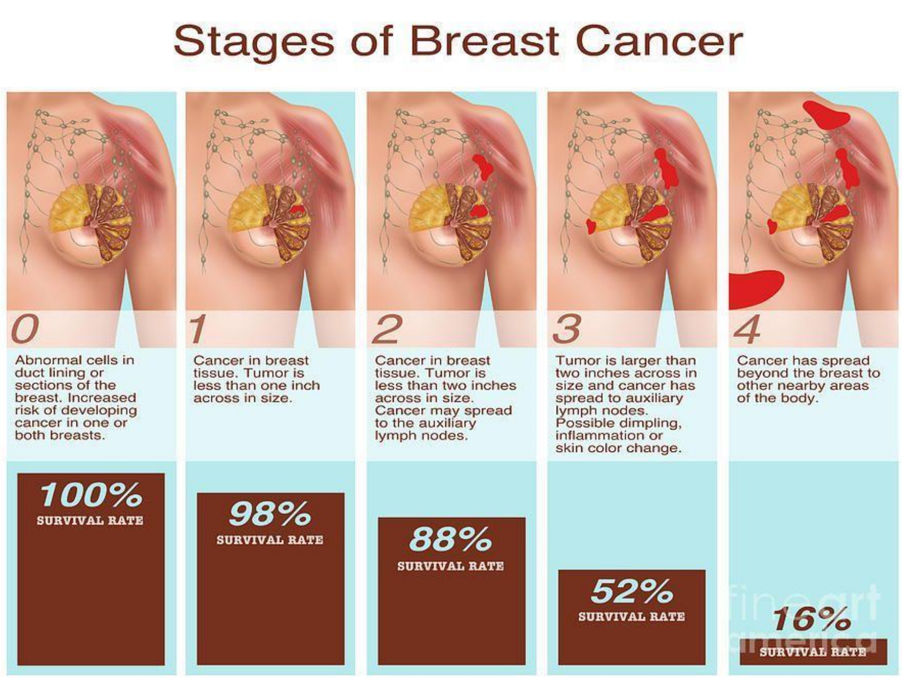


Figure2. 4: stage of breast cancer, tumor size and survival rate (Science photo library,2020)

2.7 Basics of Mammography

A piece of information on RadiologyInfo.org for patients (2018) asserts;

> "Mammography is specialized medical imaging that uses a low-dose x-ray system to see inside the breasts. A mammography exam, called a mammogram, aids in the early detection and diagnosis of breast diseases in women. An x-ray (radiograph) is a noninvasive medical test that helps physicians diagnose and treat medical conditions. Imaging with x-rays involves exposing a part of the body to a small dose of ionizing radiation to produce pictures of the inside of the body. X-rays are the oldest and most frequently used form of medical imaging.".

The process is designed to find abnormalities and irregularities in the imaged breasts. Mammography might serve screening or diagnostic evaluation purposes. Screening mammography is employed for an early detection of breast abnormalities with no prior sign of irregularities while diagnostic mammography is used to detect suspicious changes (such as pain, lumps and unusual appearance!) in breasts and to evaluate abnormal findings in mammograms. Screening test is likely to yield smaller number of breast cancer confined to breast area. The look of a woman with breast cancer is predicted through the size of the breast and reaches among others (Mayo clinic,2017; American cancer society, 2019).

Screening mammography comes with its own benefits and risks. It reduces the risk of death by improving the capability of health professionals to detect small abnormalities. X-rays are known to have no side effects for this type tests and no radiation remains with the patient. The risks of screening mammography include the existence of false positives, an exposure to excessive radiations, and its difficulty during pregnancy.

In orientational view of mammography image, Cranio-Caudal (CC) and Medio-Lateral Oblique (MLO) are the common projections mammography views. The pro of medio-lateral oblique projection is that it helps to make visible almost all part of the breast, often including lymph nodes. The main con is that portion of chest wall location, or pectoral muscle, is shown is upper right corner for right oriented and upper left corner for left oriented mammography image of right and left radiographic image of the breast respectively, as shown in figure 2.2. In contrast, the cranio-caudal view (CC) is taken from above, resulting in an image that sometimes does not show the area close to the chest wall. The pectoral muscle has a slightly higher intensity compared to the rest of the breast tissue and appear in upper left corner or upper right corner of MLO view of mammogram.

2.8 Advances of Imaging Test in Diagnostic Breast Cancer

The contemporary advances in mammography exercise can be seen in three categories. The first category has to do with the full-field digital mammography or digital mammography in general, where an electronic replacement to the x-ray film is employed. The process converts the x-rays into mammographic pictures of the breast. The second group involves the so-called computer aided detection (CAD) systems. The system searches the digital mammograms for possible indications cancer. And the final category is a three-dimensional mammography otherwise known as digital breast tomosynthesis (DBT). This is an advanced technique where many breast images from various angles are captured and synthesized into a three-dimensional image set.

Here, it is worth mentioning that there are new imaging techniques being developed and tested. Some of them include; Scinti-mammography (molecular breast imaging), Electrical impedance imaging (EIT), Positron emission mammography (PEM) and Elastography.

2.9 Limitations of a Mammography

Despite the benefits and recent advancements, mammography is not a perfect to the end; it has its own limitations. It is not a one hundred percent accurate in detecting all breast cancers. The fact that breasts looks different as per the woman, and the possibility of breast transplants may distort mammograms. There are false negative and false positive results. False negative happens when mammograms are unable to detect cancer and false positives happen when a mammogram looks abnormal when in fact there is no breast cancer. The difficulty of visualizing a breast cancer leads professionals to images with past results (RadiologyInfo.org for patients, 2018; American cancer society,2018).

2.10 Related Works

Classical mammography-based breast cancer CAD systems follow pre-processing, segmentation, feature extraction and classification steps. The need for improved mammogram picture quality is of considerable importance which makes preprocessing an essential element in mammogram analysis. Pre-processing is used for image enhancement where less important parts of the image are reduced. Image enhancement and segmentation are the integral parts of most CAD systems (George and Dhas 2017; Makandar and Halalli 2016).

All breast cancer CAD systems result in the automatic anatomical extraction of nipple, breast border, and pectoral muscle. Performing pectoral muscle segmentation is a challenging task due to its considerable variability. Invisibility of the pectoral boundaries, the presence of false positive, and the overlapping nature of the fibro-glandular tissue are among the challenges. Segmentation, being putting an image into sub divisions for further processing, is a very important step in CAD systems. Segmentations makes the search for irregularities easier(Ibrahim et al. 2016; Rampun et al. 2019).

(Rampun et al. 2019) summarized the importance of accurate segmentations of a pectoral muscle for mammographic analysis as an essential pre-processing breast cancer CAD system as follows: the inclusion of the pectoral muscle region into breast density may result in inaccurate density, and the correct estimation of the pectoral muscle boundary leads to accurate mammogram pair registration.

There are various segmentation methods. Many researchers have reviewed different segmentation techniques (for example Ibrahim et al., nd; Rampun et al., 2019; and Moghbel et al., 2020; Akram et al., 2013). Moghbel et al. (2020) holistically and thoroughly reviewed the pectoral segmentation techniques in to: Thresholding and Identity based models, Region-growing methods, Line estimation methods, Curve estimation methods, Cliff detection methods, and other methods which employ combinations of different methods. Recent developments in the field which employ deep learning techniques are also worth mentioning here (Rampun et al., 2019).

Akrametal. (2013), by employing active contour method, attempted to set up an algorithm for a pre-processing of automated computer aided detection system of mammograms. They have managed to develop a pre-processing technique for digital mammograms which removes pectoral muscles, labels, and artifacts. The desired image without the pectoral muscle and artifacts is obtained in three steps; firstly, they used a thresholding technique to remove labels and artifacts; secondly, the acquired a contour with information on pectoral boundaries through multi-phase active contour technique, and finally, they extracted the pectoral muscle region using the contour in step two and came up with a final mammogram image without unwanted features. They have tested the proposed algorithm on mammograms from the mini-MIAS database (although much dependent on the intensity of the images) and found it to give accurate pectoral segmentation. The results found are tested to give accurate outputs in comparison to older techniques. The accuracy of segmentation results of the proposed algorithm were 77.1 percent for bad pre-processing and up to 97.8 percent for accurate pre-processing results.

Another study to be mentioned on pre-processing of mammography images for breast cancer early detection is undertaken by Makandar and Halalli (2006). The study used three steps. The first step has to do with the removal of the background artifacts. The pectoral muscle is reduced in the second step and the enhancement of the digital mammography by wiener filter and CLAHE is the final step. The proposed algorithm was tested on mammograms from the mini-MIAS database and the results were compared with 98 percent completeness and 97 percent correctness. The results showed that the proposed technique works well for mammogram image improvement. The Wiener filter, compared to other filter techniques, was also found to be superior in removing undesired noises of mammogram images.

Rampun et al. (2019), on the other hand, published a paper on breast pectoral muscle segmentation in mammograms using a slightly different technique which used a modified holistically nested edge detection network. The paper draws its inspiration from by the Holistically-nested Edge Detection (HED) network and employed Convolutional Neural Network (CNN) which results in automatic pectoral muscle segmentation in mediolateral oblique mammograms. The paper boasts to have come up with a solution for the limitations of traditional hand-crafted techniques which are incapable of dealing with complex shape differences. A neural network framework which includes multi-scale learning is included to solve this limitation. The study has made use of extensive experimental evaluation employing scanned film mammograms and full field digital mammograms exploiting data sets from four different sources; i.e. Mammographic Image analysis Society (MIAS), Breast Cancer Digital Repository (BCDR), In Breast and Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM). The method that the authors proposed is totally automated with no need of user intervention and with easiness to identify the absence of pectoral muscle in the image.

The proposed method gives an average value of 94.8±8.5 percent for the Jaccard and 97.5±6.3 percent for Dice similarity metrics across the various data bases included in the study.

CHAPTER 3
MATERIALS AND METHODS

3.1 Dataset Used

The breast mammographic image used for analysis and testing of the proposed algorithm are taken from publicly accessible from mini-MIAS databases, the mini-MIAS datasets are made for scientific research purpose. the database consists of 322 MLO mammogram images (among this 202 MLO normal and 120 MLO abnormal digital mammogram images) with fatty, dense-glandular and fatty-glandular breast tissue categories with Calcification, speculated masses, circumscribed masses and Others with severity abnormalities (like benign and malignancy) as shown in Figure 3.1. The original digitalized 50 microns pixel image of MAIS diminished to 200-micron pixel padded/clipped resolution. Each image consists of 1024 pixels × 1024 pixels size, represented with 8-bit gray level image scale and they are in ".jpg" image format (J Suckling et al, 1994). For the proposed algorithm, 150 randomly selected mammographic image of mini-MIAS database with all type of breast tissue, size, shape and abnormalities has been used.

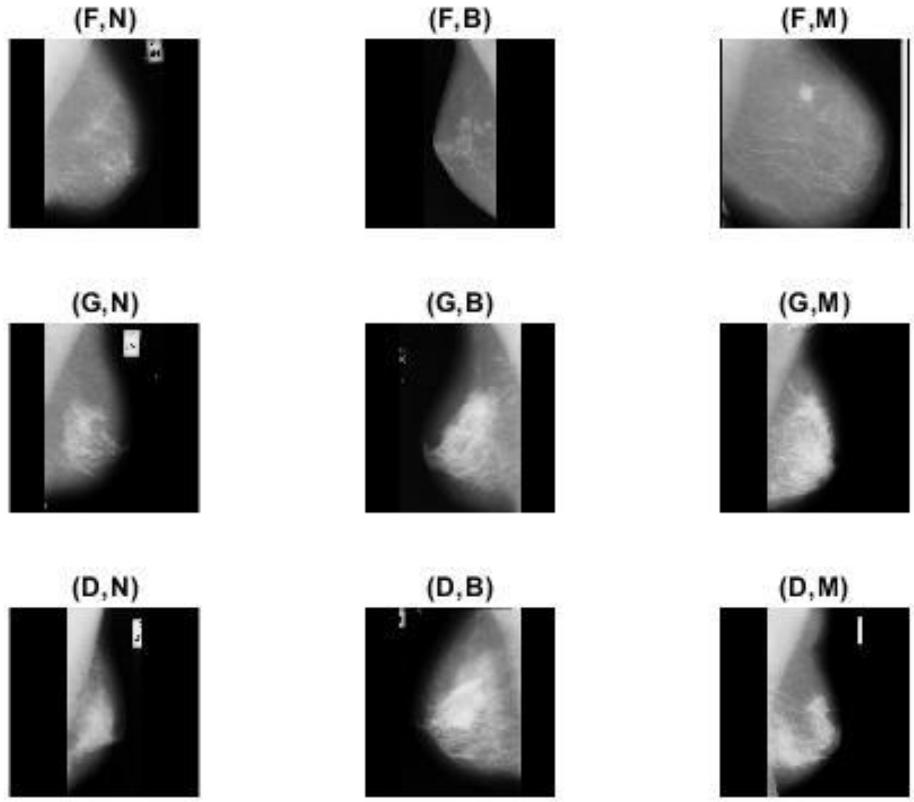


Figure 3.1: sample mammography image from mini-MIAS database: (F,N)Fatty and Normal: (F,B) Fatty with Benign: (F,M) Fatty with Malignant: (G,N) Fatty-glandular and Normal: (G,B) Fatty-glandular with Benign: (G,B) Fatty-glandular with Malignant: (D,N) Dense-glandular and Normal: (D,B) Dense-glandular with Benign: (D,M) Dense-glandular with Malignant

3.2 Methodology

3.2.1 Proposed Method

Image quality enhancement and suppressing of irreverent portion of mammography screening image plays a significant role for discovering fine small detail of digital mammograph image in early finding of breast abnormalities, for the reason of breast mammography image are difficult to interpret comparing with other medical image modalities, thus preprocessing is essential. The main goal of breast cancer CAD system is to improve productivity and performance of breast radiologist in making decision. Appropriate extraction of ROI helps in identifying abnormality swiftly and enlarge the efficiency of CAD system. The proposed algorithm possesses three main steps as briefly explained in Figure 3.2 and Figure 3.5. The first stage is all about removing of image distortions, backgrounds and artifacts, followed by pectoral muscle extraction. finally, image contrast intensity is improved and reduction of unknown noise takes place by CLAHE and median filter respectively that enriches the quality of mammography images.

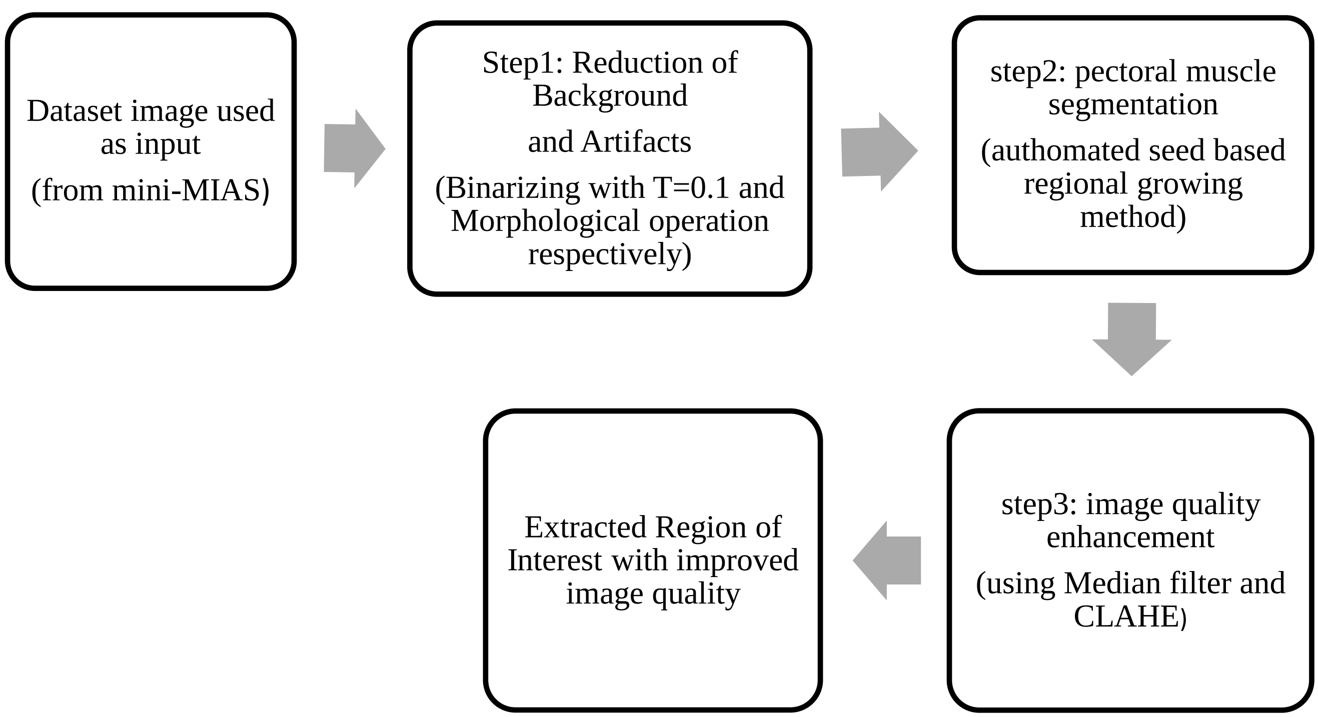


Figure 3. 2:block diagram of proposed method

3.2.2 Reduction of artifacts and separation of background

Existing of artifacts (such as label, wedges and markers) and digitization noise (as expressed in figure 3.3) causes failure of proper segmentation algorithms in digital mammogram region. Suppressing irreverent background information, removing digital noise and extracting ROI is mammography preprocessing method that significantly determine the accuracy and efficiency of CAD system in its segmentation, feature extraction and classification steps. To remove unwanted background (black pixels), firstly the gray scale images are converted to binary image (with T=0.1) by thresholding method. less important artifacts like label, wedges, and markers are suppressed using morphological operation method as follow:

Step 1: Gray scale images are binarized with threshold (T=0.1). It is separating of image into foreground and background. the 0.1 threshold value binarized the mini-MIAS database mammographic images in better way and found by try and error.

Step 2: labeling the binary image which contains radiopaque artifacts using *bwlabel* MATLAB function from the breast profile region.it is labeling of pixels as belonging to one of two (or more) classes, then counting number of pixels by *regionprops* MATLAB function which measure properties of image regions.

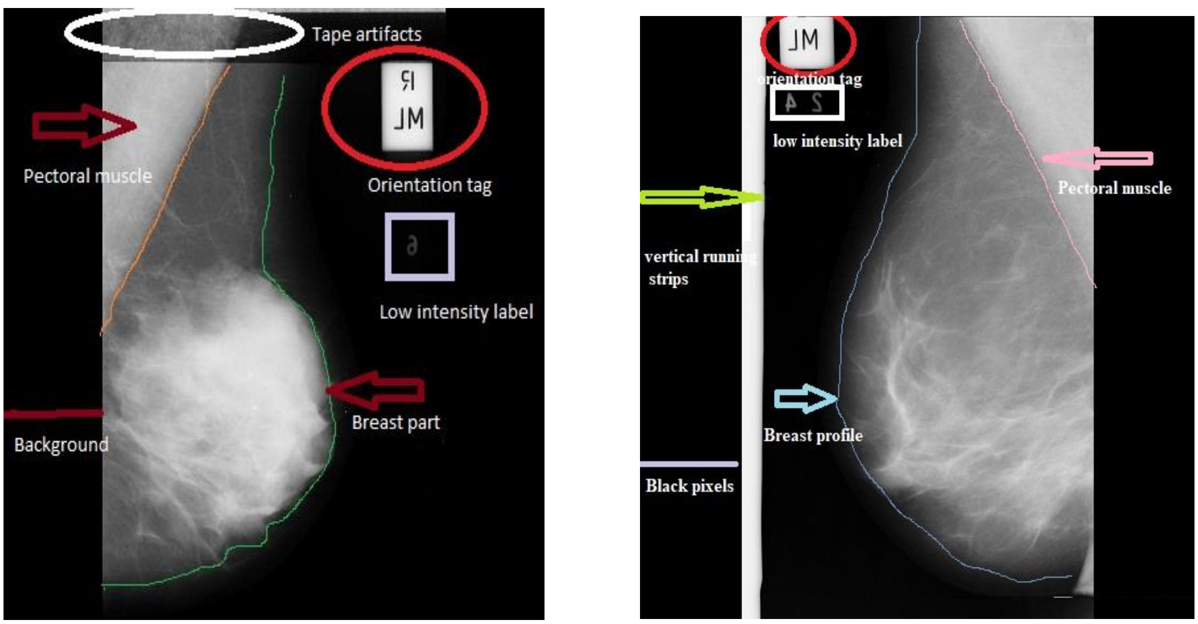
Step 3: Extract the largest breast profile using *bwareafilt* MATLAB function, it extracts or retain only the breast profile or object with the largest area. i.e. "BW2 = bwareafilt(BW,n) keeps the n largest objects. In the event of a tie for n-th place, only the first n objects are included in BW2." In our case n=1 that is the breast profile only.

Step 4: *bwmorph* MATLAB function with 'clean' and 'majority' parameters for removing isolated pixels (individual 1s that are surrounded by 0s) and smoothing the noise respectively. Morphological operation is study of objects that fix up or structuring once picture. BW2 = bwmorph(BW,operation) applies a specific morphological operation to the binary image BW. Depending on different operations, it can perform different structural rearrangement like removing interior pixels to leave an outline of the shapes, finding end points, removing spur pixels, removing isolated pixels and so on.

Step 5: Followed by morphological erosion and dilation using *imerode*. and *imdilate* MATLAB function. Dilation expands the connected sets of 1s of binary image whereas Erosion shrinks the connected sets of 1s of binary image,

Step 6: Filling the whole by using *imfill MATLAB* function with ‘holes’ parameters.

Step 7: Finally multiply the original image with image obtained from step 6 using MATLAB function *immultiply* to get breast image with reduced background and suppressed artifacts (Ibrahim et al. 2016).



(a) (b)

Figure 3. 3: Sample mini-MIAS database image that contains many background objects and artifacts

3.2.3 Pectoral muscle segmentation

3.2.3.1 Image segmentation

Disparity of an image in to its respective component or regions that have similarity according to predefined criteria is called Segmentation. Its main objective is to extract the ROI according to the interpolated criteria. Some of the image segmentation techniques that commonly used are: threshold based, Region based, Edge based, Watershed based and Cluster based segmentation. Intensity values, similarity and discontinuity are the two main principal approach of segmentation algorithm in gray images. In intensity value-based property, the segmentation algorithm partitions an image based on abrupt changes in intensity, such as edges or unsmothers in an image whereas the second approach segment an image into constituent regions of similarity based on predefined criteria

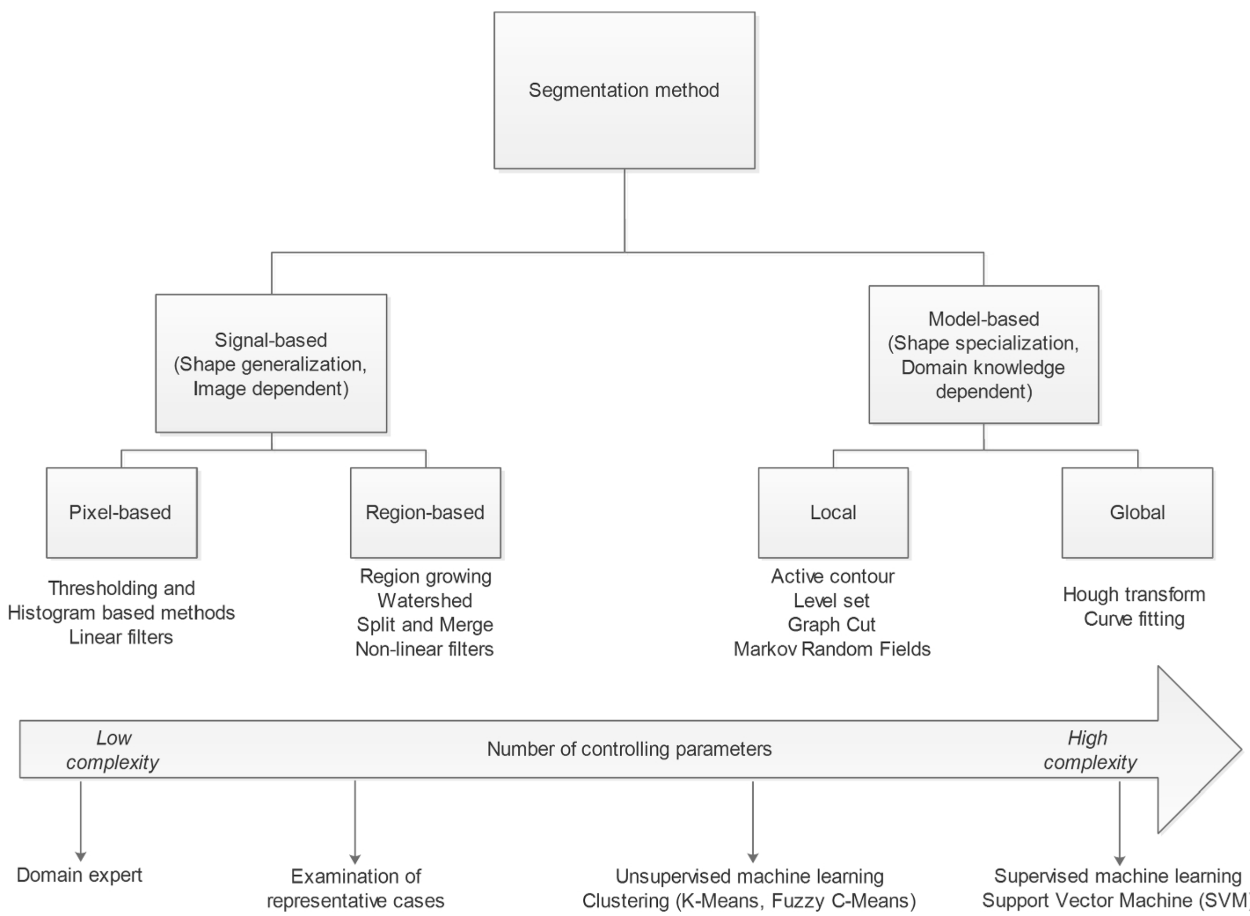


Figure 3. 4: classification of image segmentation techniques (Moghbel et al. 2019)

In mammography image capturing of MLO protentional view, the presence of pectoral muscle easily misleads the CAD system to false positive. CAD system could misclassify the pectoral muscle region (chest wall location portion) as fibro-glandular tissues (due to similar image characteristics and usually intensity value similarity). Moreover, breast radiologist and some CAD systems examines the portion of chest wall location (pectoral muscle) as sign of abnormal axillary lymph which is the clue in the appearance of breast carcinoma abnormalities and breast carcinoma is the most common area of cancers growing up (Moghbel et al. 2019).

3.2.3.2 *Region growing*

Segmentation technique, i.e. region growing, that is partitioning or grouping a pixel of an image into corresponding region which directly finds the region according to predefined similarity criteria. Forming grow region starts by the basic approach called 'seed' points. A seed point is starting point of the growth from the model of expected result to be segmented. Selecting a set of one or more seed point for region growing often is based on the nature of the problem and data available on the image. Seed or a set of seeds in seeded based region growing technique could be selected manually or automatically. In assigning pixels to specific region, connectivity property and formulating of stopping rule are crucial. Groping pixels ends when no more pixels suit the predefined similarity criteria for inclusion in the region, criteria of region growing methods are like intensity value, Color, texture. size, shape of the region being grouping etc.(Pitas 2000).

In this study, we use automated seeded region growing method for segmenting the pectoral method because of its efficiency and simplicity. In the proposed method, seeds are selected automatically based on the orientation of mammography image. Orientation of the image found by dividing image into half and counting the non-zero pixels whether it is right or left facing. As clearly explained in the below Figure of flow chart of automatic seeded based region growing method, the approach iteratively starts grown in to corresponding group from the automatic selected seed point by comparing all unallocated neighboring pixels to the region based on predefined measure of similarity criteria. The measure of similarity criteria for pectoral muscle segmentation were taken as the absolute difference between pixel's intensity value and region's mean. The pixel intensity threshold criteria depend on the knowledge in the histogram of image to be segmented. The selected threshold value for the algorithm is based on the "minimum area threshold" of digital mammography image. This process quit when the intensity difference value between region mean and new pixel becomes greater than a certain threshold, in contrast if the measure of similarity is less than certain threshold this way is segmented to region and start iteratively grown (Dirk-Jan Kroon , 2020).

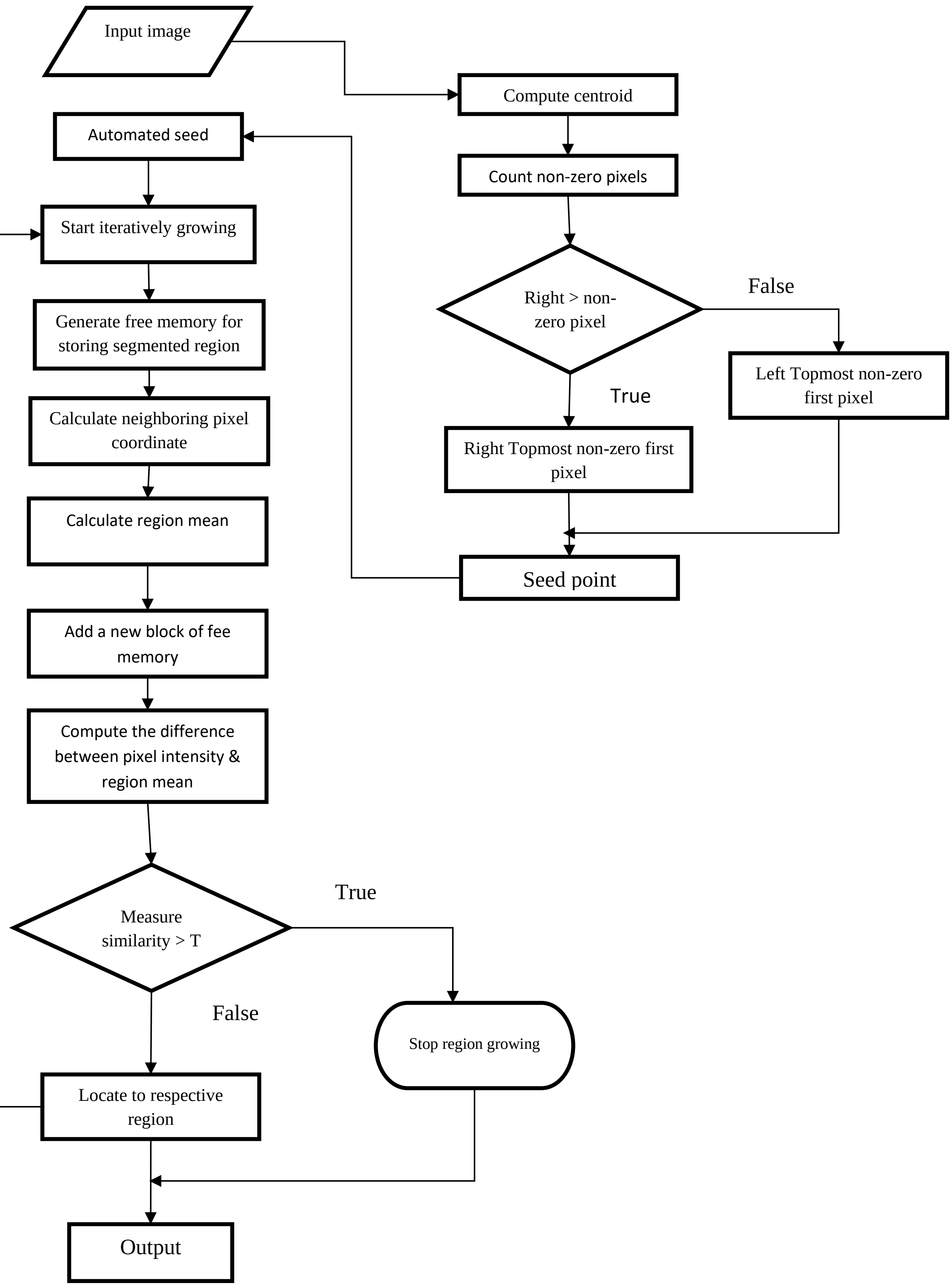

Input image
Compute centroid
Automated seed
Count non-zero pixels
Start iteratively growing
Right > non-zero pixel
False
Generate free memory for storing segmented region
Left Topmost non-zero first pixel
True
Calculate neighboring pixel coordinate
Right Topmost non-zero first pixel
Calculate region mean
Seed point
Add a new block of fee memory
Compute the difference between pixel intensity & region mean
True
Measure similarity > T
False
Stop region growing
Locate to respective region
Output

Figure 3. 5: Flow chart of automated seeded region growing method

3.2.4 Image enhancement

3.2.4.1 Contrast improvement

The quality of mammogram image enhancement stage is done using the2Dmedian filter and CLAHE technique. Adaptive histogram equalization (AHE) is a contrast enhancement method designed to be broadly applicable and having demonstrated effectiveness. It avoids excessive gray pixel merger and excessive bright local area of the image. However, slow speed and the over enhancement of noise it produces in relatively homogeneous regions are two problems (Zhu and Huang 2012; Tom and Wolfe 1983). Contrast limited adaptive histogram equalization is a modified and precious case of AHE (adaptive histogram equalization) technique that protect over saturation during image contrast enhancement specifically in homogeneous areas thereby the enhancement calculation is refined by imposing a user-defined maximum, i.e. clip limit, which prevent over enhancement of noise and minimize the edge-shadowing effect of unlimited AHE.

CLAHE operates on local area contrast of an image (tiles) rather than the entire image (global area) that prevents over contrast enhancement. MATLAB inbuilt function *adapthisteq*, for each tile individually, calculates the contrast transform function for image contrast enhancement. The histogram result found by contrast enhanced region approximately meets the histogram specified by 'Distribution' parameter value of CLAHE. The neighboring tiles are then connected using bilinear interpolation to reduce artificially produced boundaries. The *cliplimit* value is designed to appropriately limit the contrast value that helps in eliminate amplifying any noise that exist in the image.

The CLAHE method seeks to reduce from amplifying noise and edge-shadowing effect produced in homogeneous areas and was originally developed for medical imaging .

3.2.4.2 Noise reduction

Most of the mammography images consist of digitalization noise like vertical and horizontal lines. As researchers study indicate that median filter is one of the preferred nonlinear filters for removal of unknown noises in mammography images. Median filtering is a nonlinear filter commonly preferred for noise reduction of salt & pepper and spot noise in medical images. In this study, noise removal is carried out via 2D median filtering with 3-by-3 neighboring connectivity. Median filter is capable of preserving the sharpness of a picture edge information while removing differences between pixels in the pre-defined neighborhood. (Gallagher, Wise, and Processing 1981; George and Dhas 2017).

CHAPTER 4
RESULTS AND DISCUSSION

The proposed algorithm has been examined on 150 randomly selected databases from mini-MIAS that covers all type of breast tissue (like fatty, Dense & Glandular), different breast shape, size, deformities, asymmetries and abnormalities. The proposed algorithm was implemented on MATLAB 2019a software. Digital mammogram images consist of different background artifact noises (like high and low intensity labels, scanning artifacts and taping artifacts), horizontal and vertical running strips noise and pectoral muscles. The proposed method reduces irrelevant and unwanted background object of raw digital mammogram images. Initially digital images were binarized with threshold value of 0.1 then followed by morphological operation method for extracting the breast profile and to remove all unwanted small objects as shown in Figure 4.1.

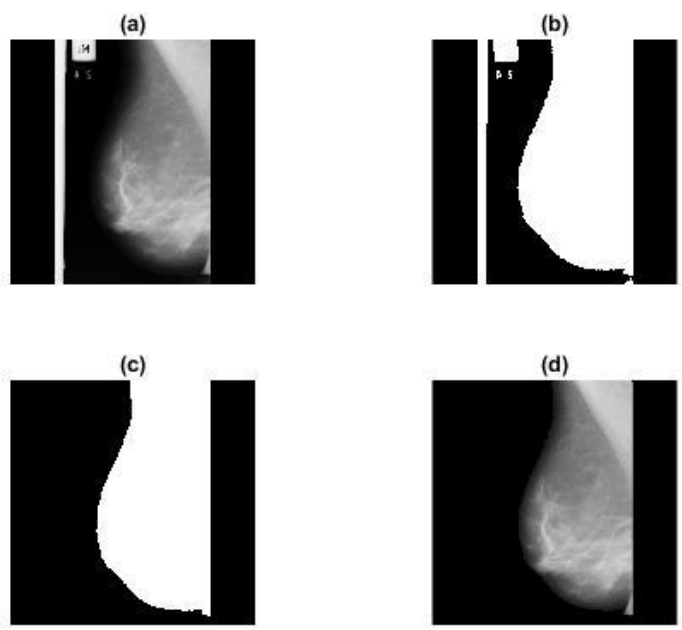


Figure 4. 1: shows the result of removal artifacts and separation of background, (a) original mini-MIAS image (b)binarization image with threshold 0.1 (c) morphological operation for biggest blob (breast profile) image (d) images (a) and (c) multiplying which results breast profile with no background artifacts and vertical running strip

Following removal of artifacts and background, the proposed algorithm automatically extracts the region of interest (ROI). Here, pectoral muscle is segmented using automatic seeded based region growing method as briefly explained in Figure 3.6. In the proposed algorithm, seeds are selected automatically based on the orientation of mammogram images. The MIAS database has either right facing or left facing MLO mammogram images. Thus, the seed point is selected from right Topmost or left Topmost non-zero first pixel for right oriented and left oriented image respectively. The orientation of mini-MIAS image are identified by dividing the image in to half (or by calculating the centroid of the image) and counting the non-zero pixels in which part are more existing. If it is right oriented, non-zero pixels consist more in right else left part consist of more non-zero pixels. Figure 4.2 show result of pectoral muscle segmentation on different breast tissue type, size, shape and abnormalities. The proposed algorithm for pectoral muscle segmentation has been tested on 150 randomly selected mini-MIAS database with all breast tissue type, shape, size and abnormalities.

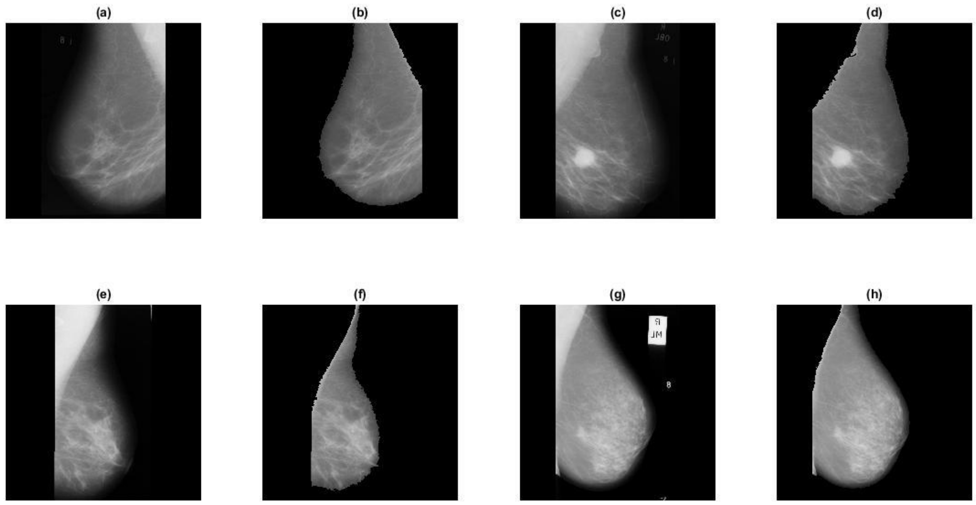


Figure 4. 2: sample result of pectoral muscle segmentation using automated seed-based region growing method: (a) (c) (e) and (g) are original mini-MIAS image, (b) (d) (f) and (h) are result of pectoral muscle segmentation

4.1 Accuracy Performance Evaluation

Accuracy performance evaluation is crucial step that defines how accurate the proposed algorithm is. There are many different ways of accuracy evaluation methods. Correctness and completeness are one of mostly used approaches. They are expressed by using TP, FN and FP parameters. TP, FN & FP defined as:

True Positive (TP): ROI pixels which are correctly identified by the method (Proper segmented)

False Negative (FN): ROI pixels which are missed as ROI by the method (under segmented)

False Positive (FP): pixels which are not in ROI are extracted as ROI pixels by the method (over segmented)

$$\text{Correctness} = \frac{TP}{TP+FP} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1.1)$$

$$\text{Completeness} = \frac{TP}{TP+FN} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1.2)$$

According to Several authors conclusion made, if the percentage of both correctness and completeness exceeds 95%, the algorithm can be considered accurate(Wirth et al. 2005).

The accuracy and efficiency of proposed algorithm for automatic pectoral muscle segmentation had been affected when the intensity of pectoral muscle and the intensity breast profile have almost similar intensity value, and when the pectoral muscle intensity are lower than the remaining breast part as shown figure 4.3. Among 150 randomly selected mini-MIAS databases that have not such difficulties, the proposed algorithm gives excellent segmentation accuracy and efficiency with 0.97 and 0.96 correctness and completeness value respectively. In contrast mammogram images that possess intensity of pectoral muscle with similarity or lower than the remaining breast part, the proposed algorithm for pectoral muscle segmentation did not properly segmented the region of interest pixels. Table 4.1 shows number of properly segmented (TP), over segmented (FP) and under segmented (FN) among the randomly selected mini-MIAS images for the proposed algorithm. The TP, FN and FP value are calculated by counting down the ROI pixels which are correctly identified by the method or proper segmented, ROI pixels which are missed as ROI by the method (under segmented) and pixels which are not in ROI are extracted as ROI pixels by the method (over segmented) respectively according to mammographic anatomy of pectoral muscle and pectoral muscle have slightly higher intensity compared the rest of breast tissue and appear to upper left/right corner of MLO view. Figure 4.3 shows sample mini-MIAS images with pectoral muscle intensity value similarity or lower than the remain breast profile.

Table4 1: shows number of properly segmented (TP), over segmented (FP) and under segmented (FN)among the randomly selected mini-MIAS images for the proposed algorithm

Numbers of Dataset used	150
Numbers of True Positive	141
Numbers of False Negative	5
Numbers of False Positive	4
Correctness	0.97
Completeness	0.96

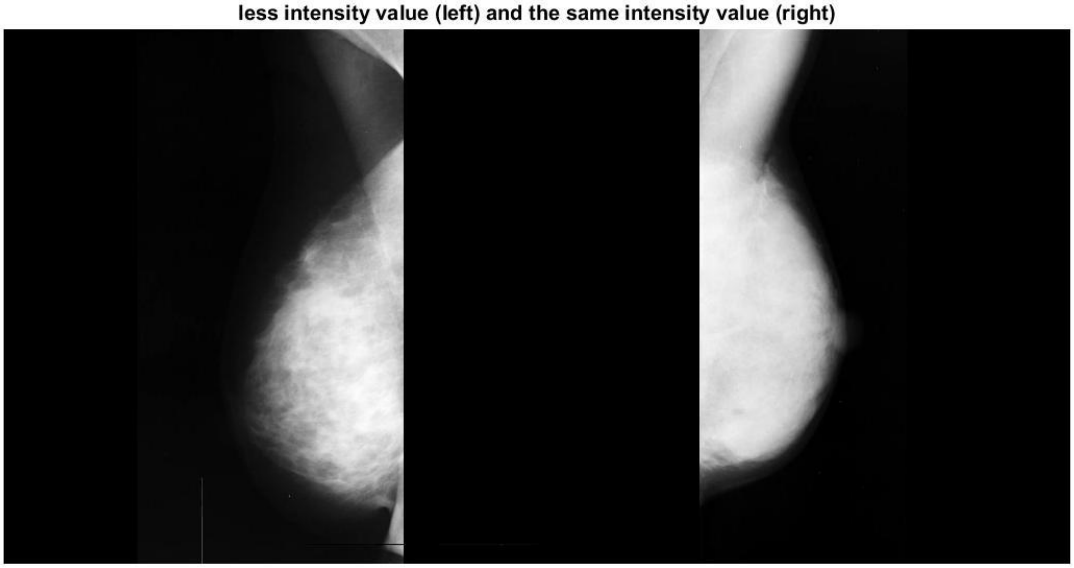


Figure 4. 3: shows sample mini-MIAS images with pectoral muscle intensity value similarity or lower than the remain breast profile

The image quality was enhanced using CLAHE and Median filter as shown in figure 4.5 the contrast limited adaptive histogram equalization for contrast enhancement of mammography image was done by MATLAB inbuilt Function *adapthisteq* with '*ClipLimit*' value and curved histogram or 'exponential' histogram shape for the image tiles, i.e. '*Distribution'*. The image quality evaluation is basic in medical imaging systems such as compression, transmission and enhancement (Makandar, Karibasappa, and Science 2010; Senthilkumar and Umamaheswari 2011; Δασκαλάκης et al. 2015)

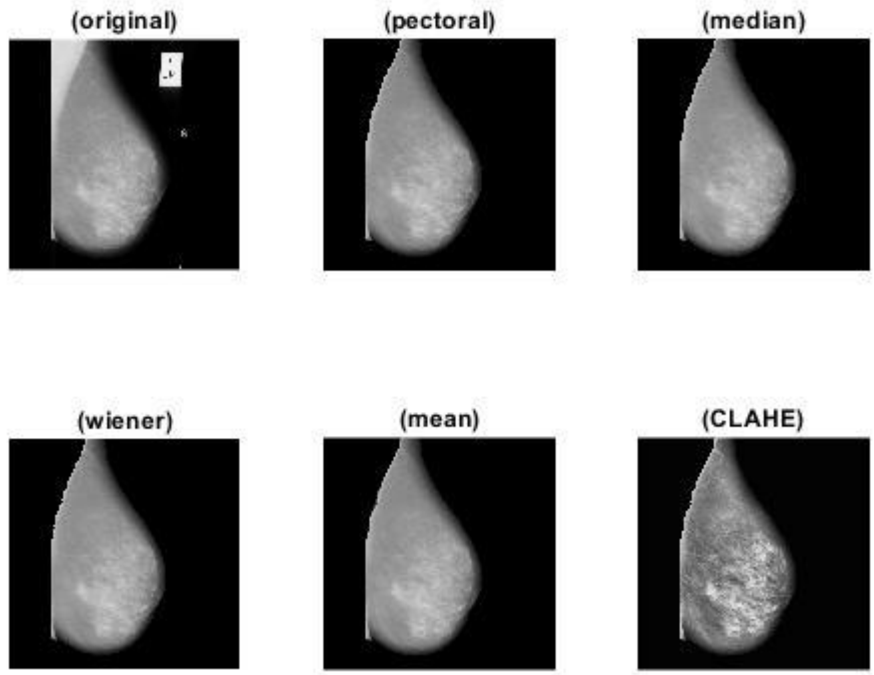


Figure 4. 4: shows result of image quality enhancement and different filters

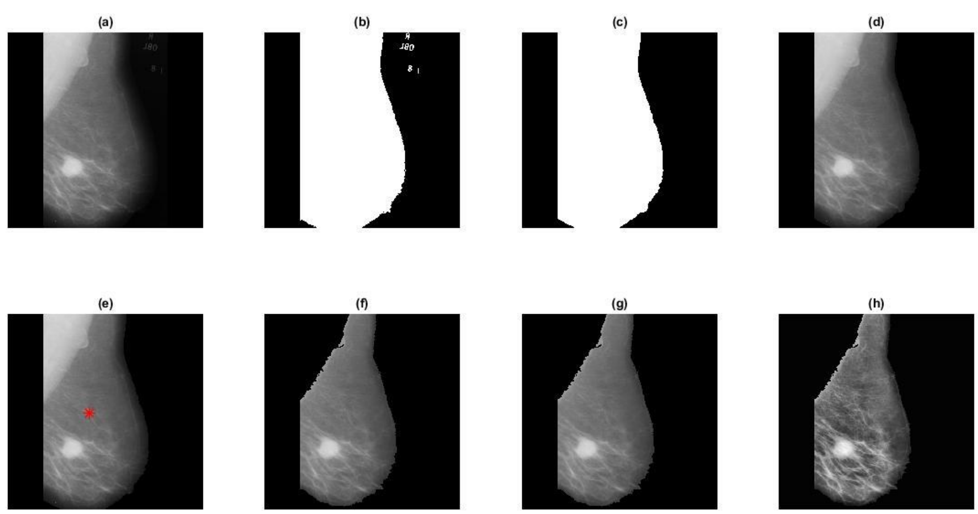


Figure 4. 5: shows experimental result of proposed method on mdb028 reference number of MIAS with Fatty (F)breast tissue type and Malignant (M): (a) original mini-MIAS image (b) binarization of image with threshold 0.1 (c) morphological operation (d) images (a) and (c) multiplying which gives image without background artifacts (e) centroid of the image for automatic seed selection (f) pectoral muscle segmented (g) median filter (h) result of CLAHE

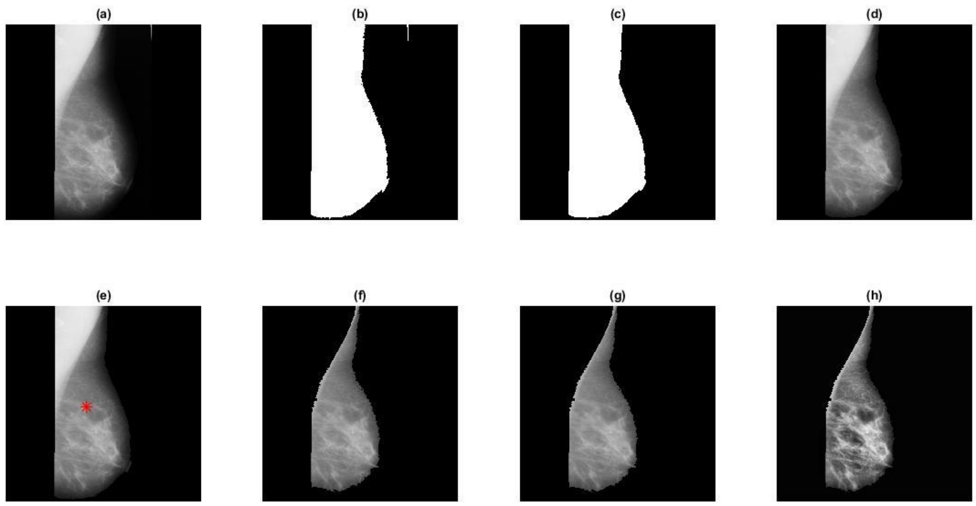


Figure 4. 6: shows experimental result of proposed method on mdb048 reference number of MIAS with Fatty-glandular(G) breast tissue type and Normal(N)

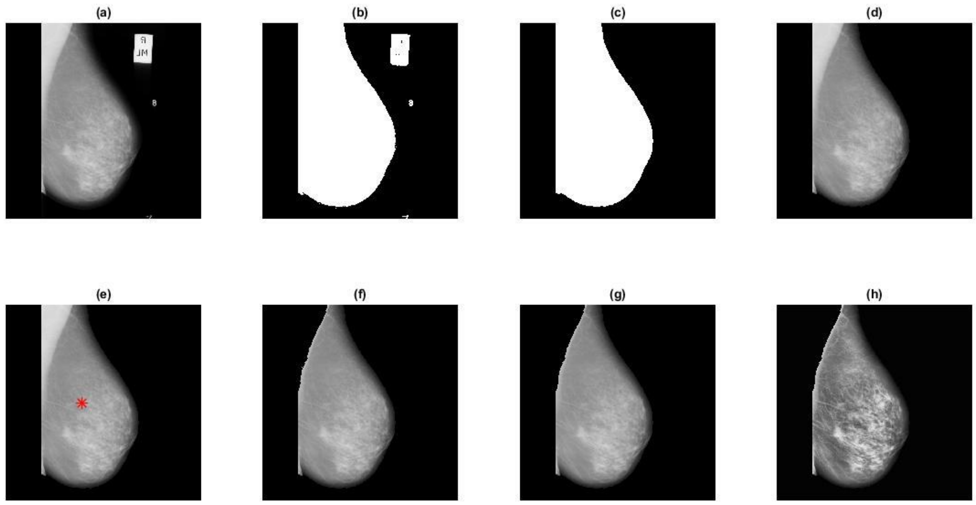


Figure 4. 7: shows experimental result of proposed method on mdb058 reference number of MIAS with Dense-glandular (D) breast tissue type and Malignant (M)

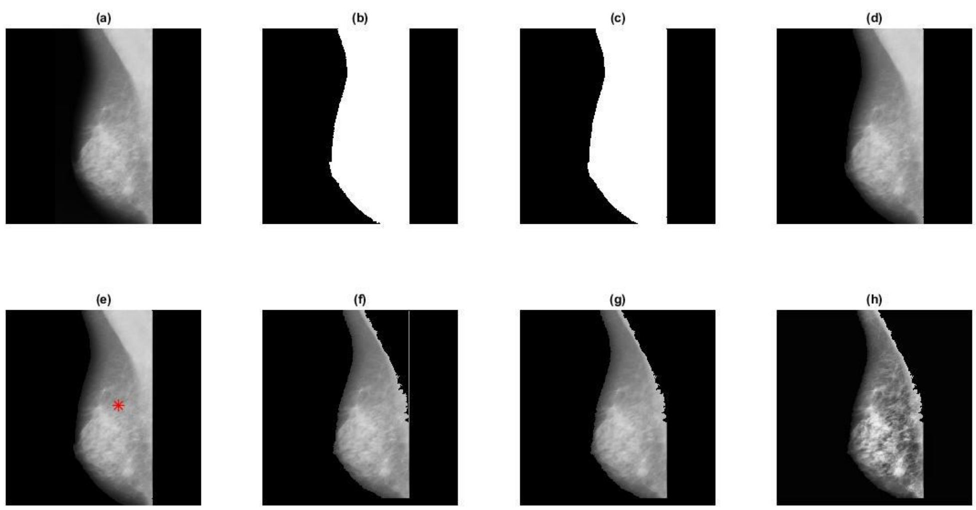


Figure 4. 8: shows experimental result of proposed method on mdb199 reference number of MIAS with Dense-glandular(D) breast tissue type and Benign(B)

Performance of image quality improvement for the proposed method was evaluated by quality assessment metrics i.e. MSE, PSNR& SNR on different images. Noise reduction of mammography images was compared with different noise filter methods, i.e. Median, mean and wiener filters (significant filter mask [3, 3]).

4.2 Quality assessment metrics

4.2.1 Performance of image quality based on MSE (Mean Square Error)

MSE is regarded as one of the parameters which determine an image quality. The mean-square error (MSE) and the peak signal-to-noise ratio (PSNR) are used to compare image compression quality. The MSE represents the cumulative squared error between the compressed and the original image, whereas PSNR represents a measure of the peak error. Lower value of MSE refers higher image quality or lower error to mean.

The mathematical expression of MSE is as follow:

$$MSE = \frac{1}{mn} \sum_{0}^{m-1} \sum_{0}^{n-1} ||f(i,j) - g(i,j)||^2 \text{……….(2.1)(Instruments 2013)}$$

Where: f represents the matrix data of original image, g represents the matrix data of degraded image in question , m represents the number of rows of pixels of the image and i represents the index of the raw, n represents the number of columns of pixels of image and j represents the index of that column.

In this study we calculated the MSE value for the filters such as median, wiener and mean for contrast limited adaptive histogram equalized image(George and Dhas 2017).MSE value is calculated by *immse* MATLAB inbuilt Function. Figure4.9 shows the obtained result of MSE.

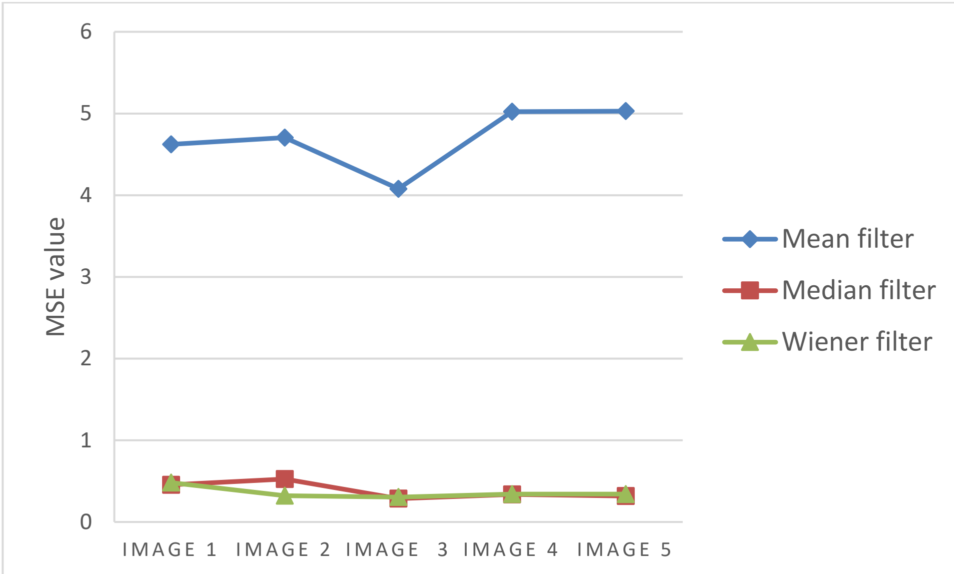


Figure 4. 9: MSE result on comparison of different filters

4.2.2 Performance of Image Quality Based on PSNR (Peak Signal to Noise Ratio)

PSNR are quantitative or empirical measure that compare the effect of image quality enhancement algorithms. “It is the ratio between the maximum possible power of a signal and power of distorting noise that affects the quality of its representation.” The PSNR block computes the peak signal-to-noise ratio, in decibels, between two images. This ratio is used as a quality measurement between the original and a compressed image. The higher the PSNR, the better the quality of the compressed, or reconstructed image.

The mathematical expression of PSNR is as follow:

$$PSNR = 20\log_{10}\left(\frac{MAXf}{\sqrt{MSE}}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.1)$$

Where, $MAXf$ is the maximum signal value that exists in original “known to be good” image.

$$MSE = \frac{1}{mn}\sum_0^{m-1}\sum_0^{n-1}||f(i,j) - g(i,j)||^2 \text{.........(2.2)(Instruments 2013)}$$

In addition to this, PSNR and SNR value are calculated by [*peaksnr, snr]* = *psnr(A, ref)*MATLAB Function. The higher the PSNR value indicates the better image quality has been reconstructed algorithm. Figure 4.10 shows the PSNR obtained result for median, wiener and mean filters with CLAHE.

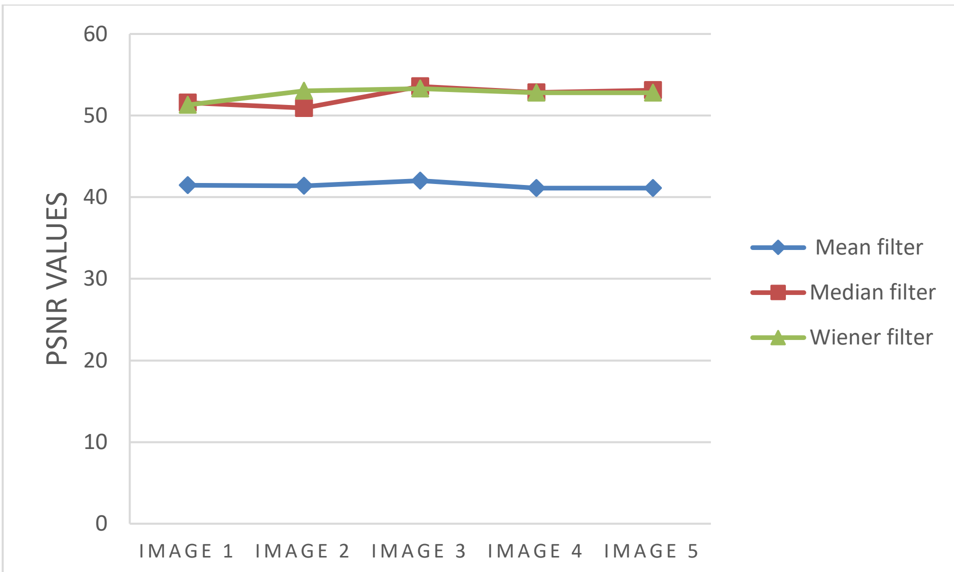


Figure 4. 10: PSNR result on comparison of different filters

4.2.3 Performance of Image Quality based on SNR (Signal to Noise Ratio)

SNR is quality parameters which measures the sensitivity of an imaging system. "SNR is defined as the proportion power of a signal to the power of background noise." The figure 4.11 below indicate the experiment result of mammogram images on SNR value for different filters with CLAHE.

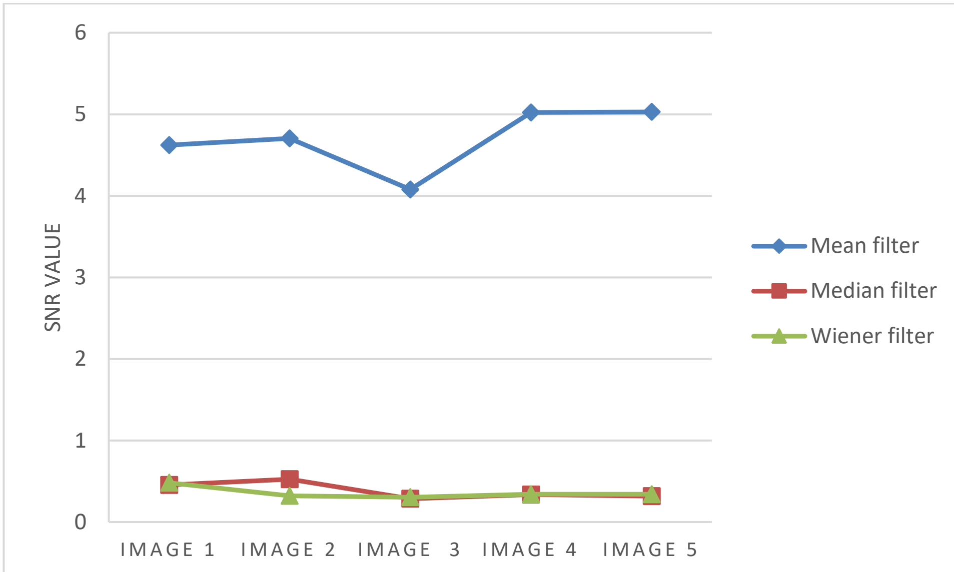


Figure 4. 11:SNR result on comparison of different filters

The strong reason for using CLAHE and Median filter for image quality enrichment is, its highest PSNR value, lower MSE value, less sensitive and its adequate capability of edge preserving comparing with mean filter and wiener filter for almost all the images tested as shown in figure 4.9, figure4.10, figure 4.11 and figure 4.4. The above quality assessment metrics graphs strongly prove that median filter is qualified for noise removal of mammography image.

In the present study, an efficient preprocessing algorithm for segmenting pectoral muscle and image quality enhancement of mammography image was developed by suitable combining in a the benefits of background artifact suppression(binarization method and morphological operation), image enhancement (median filter and CLAHE) and pectoral muscle segmentation (automated Seeded Region Growing technique) in order to improve the accuracy and efficiency of CAD system in early detection of breast cancer which is strongly believed that the accuracy of preprocessing determines the success of the remaining CAD steps like segmentation, feature extraction and classification.

By visual inspection of the original and the segmented images in Figure 4.2 that is tested on different breast tissue type, size, shape, deformity and abnormality, it can be observed that the proposed technique accurately extracts all breast type and category of pectoral muscle in MLO mammography MIAS image. This helps in accurately extraction of targeted ROI. The success of the proposed segmentation scheme is mostly due to its characteristic to perform individually.

The segmentation results of the proposed segmentation scheme were comparatively evaluated with correctness and completeness and was reported as 97% and 96% respectively. According to Several authors conclusion made, if the percentage of both correctness and completeness exceeds 95%, the algorithm can be considered accurate(Wirth et al. 2005).the proposed algorithm segment pectoral muscle accurately. As explained in Table 4.1 the accuracy of the proposed algorithm affected by two cases; when the intensity value of pectoral muscle is less than the remaining breast profile and when pectoral muscle and the breast profile have similar intensity value as shown in figure 4.3.out of this cases the proposed automated seeded region growing technique segments pectoral muscle exactly.

In summary, the proposed method sufficiently works on removal of background artifacts (like label, wedge, scanning artifacts and tape artifacts), horizontal and vertical running strips noise and pectoral muscle. Moreover, it preciously enhances the quality of image which helps in magnify small detail of mammography images. Furthermore, the proposed algorithm helps breast radiologist to diagnosis breast abnormalities more accurately and takes decision swiftly. It assists breast radiologist to detect breast abnormalities at early stage. It plays great role in minimize the possibility of false negative. Its benefits women for higher quality of life and economically treated thereby increase survival rate. It increases numbers of available treatment options, thus increase the probability of fully recovery and decrease mortality from breast cancer. It helps to interpret breast images correctly and conspicuously.

CHAPTER 5
CONCLUSION AND RECOMMENDATION

5.1 Conclusion

Numbers of new diagnosed breast cancer cases are still increasing from time to time. Hence, raising the requirement for timely screening and proper interpretation, this in turn helps in decreasing breast cancer related mortality. Therefore, early and timely breast cancer detection improves quality of life. Nowadays, mammography is gold stand and widely used x-ray imaging technique for early detection of breast cancer. However, mammogram images difficult to interpreted and usually exposed to background artifacts (like label, wedge, Tape artifact and scanning artifacts), and unknown noises in addition to its low x-ray dose radiation that show poor image quality. These all pose immense influences in detecting small abnormalities at primary level.

In this paper, an efficient preprocessing method of mammography image quality enhancement and accurate ROI extraction algorithm is designed. The proposed method removes irrelevant and less important background artifacts, unknown noises and pectoral muscle. our result shows that, median filter & CLAHE preciously improves mammography image quality and magnify fines detail of an image compare with mean filter and wiener filter. This helps much in detecting of breast cancer at primary stage. Automated seed-based region growing method was designed that successfully extracted a pectoral muscle. The proposed method has been examined on mini-MIAS database of mammographic breast images.

Collectively, the results found verify that the proposed algorithm definitively performs well in enrichment of mammography image quality and extraction of ROI for almost all breast tissue type, size, shape, asymmetry and abnormalities. This suits CAD system in increasing its accuracy and efficiency of breast cancer detection at early stage.

5.2 Recommendations

1. We call and encourage as the implementation of this thesis findings will improve the performance of breast radiologist and CAD system in early detection of breast cancers.

2. For future studies extended segmentation method & models in the area of deep learning for better ROI extraction accuracy performance and solving intensity- based segmentation limitations is planned and we also recommend.

REFERENCES

American cancer society. (2018). Mammograms: Limitation of mammograms. Retrieved October 3, 2019 from https://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection/mammograms/limitations-of-mammograms.html

American cancer society. (2019). American cancer society recommendations for the early detection of breast cancer: Breast cancer early detection and diagnosis. Retrieved March 5, 2020 from https://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection/american-cancer-society-recommendations-for-the-early-detection-of-breast-cancer.html#references

Barth, Volker. 2011. Diagnosis of breast diseases: integrating the findings of clinical presentation, mammography, and ultrasound (Thieme).

Becker, Sven 2015. 'A historic and scientific review of breast cancer: The next global healthcare challenge', J International Journal of Gynecology Obstetrics, 131: S36-S39.

Becker, Sven %J International Journal of Gynecology, and Obstetrics. 2015. 'A historic and scientific review of breast cancer: The next global healthcare challenge', 131: S36-S39.

Bilimoria, Karl Y, Andrew K Stewart, David P Winchester, and Clifford Y %J Annals of surgical oncology Ko. 2008. 'The National Cancer Data Base: a powerful initiative to improve cancer care in the United States', 15: 683-90.

Bird, Richard E, Terry W Wallace, and Bonnie C %J Radiology Yankaskas. 1992. 'Analysis of cancers missed at screening mammography', 184: 613-17.

Bland, Kirby I, Edward M Copeland, V Suzanne Klimberg, and William J Gradishar. 2017. The Breast E-Book: Comprehensive Management of Benign and Malignant Diseases (Elsevier Health Sciences).

Breast Cancer Research Foundation.(2018).breast cancer statistics and resources.from https://www.bcrf.org/breast-cancer-statistics-and-resources

Caouette-Laberge, L., and D. Borsuk. 2013. 'Congenital anomalies of the breast', Semin Plast Surg, 27: 36-41.

Casey, Holly D, Paul E Chasan, and Leland R %J Annals of plastic surgery Chick. 1996. 'Familial polythelia without associated anomalies', 36: 101-04.

De Silva, Nirupama K, and Mary L Brandt. 2006a. 'Disorders of the breast in children and adolescents, Part 2: breast masses', Journal of Pediatric Adolescent Gynecology, 19: 415-18.

De Silva, Nirupama K, and Mary L Brandt. 2006b. 'Disorders of the breast in children and adolescents, Part 1: Disorders of growth and infections of the breast', Journal of pediatric adolescent gynecology, 19: 345-49.

Dirk-Jan Kroon (2020). Region Growing (https://www.mathworks.com/matlabcentral/fileexchange/19084-region-growing), MATLAB Central File Exchange. Retrieved July 1, 2020.

Drake, Richard, A Wayne Vogl, and Adam WM Mitchell. 2009. Gray's Anatomy for Students E-Book (Elsevier Health Sciences).

Gallagher, N, Gı %J IEEE Transactions on Acoustics Wise, Speech,, and Signal Processing. 1981. 'A theoretical analysis of the properties of median filters', 29: 1136-41.

George, M Jayesh, and D Anto Sahaya Dhas. 2017. "Preprocessing filters for mammogram images: A review." In 2017 Conference on Emerging Devices and Smart Systems (ICEDSS), 1-7. IEEE.

Halalli, Bhagirathi, and Aziz %J Breast Imaging Makandar. 2018. 'Computer aided diagnosis-medical image analysis techniques': 85.

Heath, Michael, Kevin Bowyer, Daniel Kopans, P Kegelmeyer, Richard Moore, Kyong Chang, and S Munishkumaran. 1998. 'Current status of the digital database for screening mammography.' in, Digital mammography (Springer).

Ibrahim, Naglaa S Ali, Naglaa F Soliman, Mahmoud Abdallah, and Fathi E Abd El-Samie. 2016. "An algorithm for pre-processing and segmentation of mammogram images." In 2016

11th International Conference on Computer Engineering & Systems (ICCES), 187-90. IEEE.

Instruments, N. 2013. "Peak signal-to-noise ratio as an image quality metric." In.

International Agency for Research on Cancer (IARC) and World Health Organization (WHO). GLOBOCAN 2018: Breast. http://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheet.pdf, 2019

Kamal, R. M., S. T. Hamed, and D. S. Salem. 2009. 'Classification of inflammatory breast disorders and step by step diagnosis', Breast J, 15: 367-80.

Kerlikowske, Karla, Deborah Grady, Susan M Rubin, Christian Sandrock, and Virginia L %J Jama Ernster. 1995. 'Efficacy of screening mammography: a meta-analysis', 273: 149-54.

Lattanzio, Vincenzo, and Giovanni Simonetti. 2010. Mammography: Guide to Interpreting, Reporting and Auditing Mammographic Images-Re. Co. RM (From Italian Reporting and Codifying the Results of Mammography) (Springer Science & Business Media).

Leong, P. W., N. C. Chotai, and S. Kulkarni. 2018. 'Imaging Features of Inflammatory Breast Disorders: A Pictorial Essay', Korean J Radiol, 19: 5-14.

Lepori, D. 2015. 'Inflammatory breast disease: The radiologist's role', Diagn Interv Imaging, 96: 1045-64.

Lester, Susan C, Shikha Bose, Yunn-Yi Chen, James L Connolly, Monica E de Baca, Patrick L Fitzgibbons, Daniel F Hayes, Celina Kleer, Frances P O'Malley, David L %J Archives of pathology Page, and laboratory medicine. 2009. 'Protocol for the examination of specimens from patients with invasive carcinoma of the breast', 133: 1515-38.

Makandar, Aziz, and Bhagirathi %J International Journal of Computer Applications Halalli. 2016. 'Pre-processing of mammography image for early detection of breast cancer', 144: 11-15.

Makandar, Aziz Ur Rahaman, K %J Journal of Computer Science Karibasappa, and Mathamatical Science. 2010. 'Wavelet based medical image compression using SPHIT', 1: 769-75.

Malhotra, Gautam K, Xiangshan Zhao, Hamid Band, Vimla %J Cancer biology Band, and therapy. 2010. 'Histological, molecular and functional subtypes of breast cancers', 10: 955-60.

Mayo clinic. (2017).Mammogram. Retrieved July 31, 2019.from https://www.mayoclinic.org/tests-procedures/mammogram/about/pac-20384806

Moghbel, Mehrdad, Chia Yee Ooi, Nordinah Ismail, Yuan Wen Hau, and Nogol %J Artificial Intelligence Review Memari. 2019. 'A review of breast boundary and pectoral muscle segmentation methods in computer-aided detection/diagnosis of breast mammography': 1-46.

Pitas, Ioannis. 2000. Digital image processing algorithms and applications (John Wiley & Sons).

Pohlodek, Kamil. 2014. Introduction to breast diseases (Faculty of Medicine: Comenius University Bratislava).

RadiologyInfo.org for patients. (2018). Mammograppy. Retrieved April 12, 2019 from https://www.radiologyinfo.org/en/info.cfm?pg=mammo

Rampun, Andrik, Karen López-Linares, Philip J Morrow, Bryan W Scotney, Hui Wang, Inmaculada Garcia Ocaña, Grégory Maclair, Reyer Zwiggelaar, Miguel A González Ballester, and Iván %J Medical image analysis Macía. 2019. 'Breast pectoral muscle segmentation in mammograms using a modified holistically-nested edge detection network', 57: 1-17.

Salzman, Brooke, Elizabeth Collins, and Lauren %J American family physician Hersh. 2019. 'Common breast problems', 99: 505-14.

Science photo library. (2020). Breast cancer stages, illustration. From https://www.sciencephoto.com /media/704648/view/breast-cancer-stages-illustration

Senthilkumar, B, and G %J European Journal of Scientific Research Umamaheswari. 2011. 'A novel edge detection algorithm for the detection of breast cancer', 53: 51-55.

Shamberger, R. C., K. J. Welch, and J. Upton, 3rd. 1989. 'Surgical treatment of thoracic deformity in Poland's syndrome', J Pediatr Surg, 24: 760-5; discussion 66.

Tom, Victor, and Gregory Wolfe. 1983. Adaptive Histogram Equalization And Its Applications (SPIE).

Wineski, Lawrence E. 2018. Snell's Clinical Anatomy by Regions (Lippincott Williams & Wilkins).

Wirth, Michael, Dennis Nikitenko, J %J ICGST International Journal on Graphics Lyon, Vision, and Image Processing. 2005. 'Segmentation of the breast region in mammograms using a rule-based fuzzy reasoning algorithm', 5: 45-54.

Zhu, Youlian, and Cheng %J Physics Procedia Huang. 2012. 'An adaptive histogram equalization algorithm on the image gray level mapping', 25: 601-08.

Δασκαλάκης, Αντώνης, Διονύσης Α Κάβουρας, Παναγιώτης Μπουγιούκος, Σπυρίδων Κωστόπουλος, and Παντελής Γεωργιάδης. 2015. 'An efficient clahe-based, spot-adaptive, image segmentation technique for improving microarray genes' quantification'.

APPENDIX 1: Ethical Approval letter

NEAR EAST UNIVERSITY

ETHICAL APPROVAL DOCUMENT

Date:12//08/2020

To the Graduate School of Applied Sciences

For the thesis project entitled as "PREPROCESSING OF MAMMOGRAPHY IMAGE FOR

IMPROVING ACCURACY & EFFICIENCY OF CAD IN EARLY DETECTION OF BREAST CANCER", the researchers declare that they did not collect any data from human/animal or any other subjects. Therefore, this project does not need to go through the ethics committee evaluation.

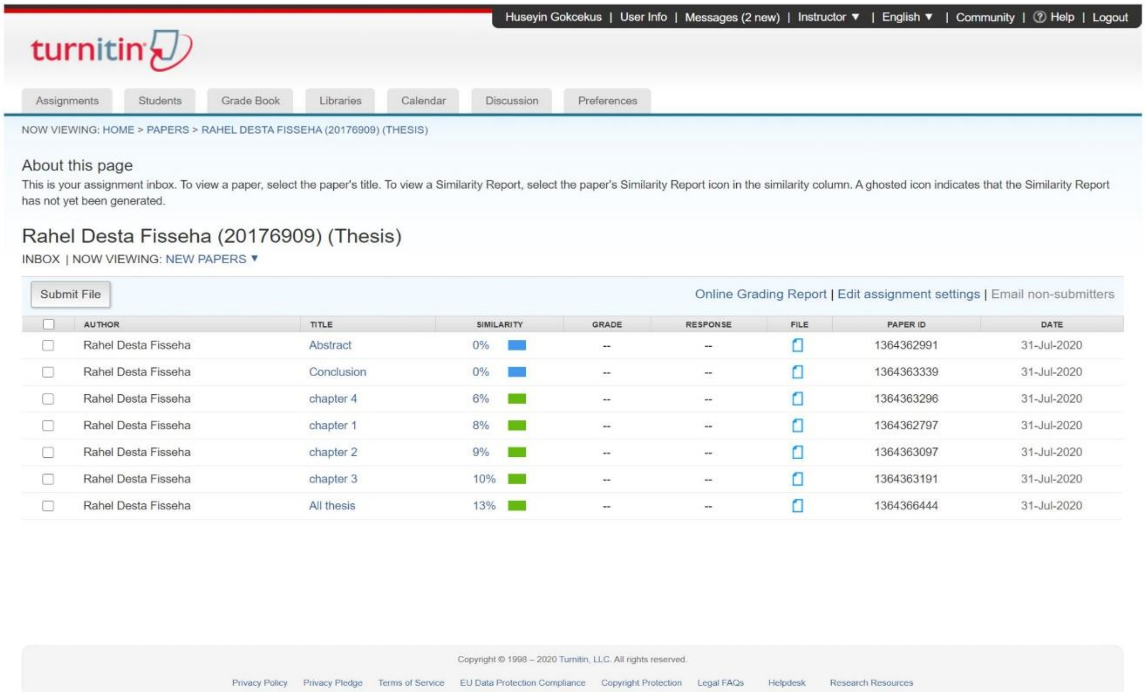
Title: PROF. DR.

Name Surname: AYŞE GÜNAY KİBARER

Signature:

Role in the Research Project: Supervisor

APPENDIX 2: Similarity Report -TURNITIN (Consultant Signed)



Huseyin Gokcekus | User Info | Messages (2 new) | Instructor ▼ | English ▼ | Community | ⓘ Help | Logout

turnitin

Assignments | Students | Grade Book | Libraries | Calendar | Discussion | Preferences

NOW VIEWING: HOME > PAPERS > RAHEL DESTA FISSEHA (20176909) (THESIS)

About this page

This is your assignment inbox. To view a paper, select the paper's title. To view a Similarity Report, select the paper's Similarity Report icon in the similarity column. A ghosted icon indicates that the Similarity Report has not yet been generated.

Rahel Desta Fisseha (20176909) (Thesis)

INBOX | NOW VIEWING: NEW PAPERS ▼

Submit File

Online Grading Report | Edit assignment settings | Email non-submitters

AUTHOR	TITLE	SIMILARITY	GRADE	RESPONSE	FILE	PAPER ID	DATE
Rahel Desta Fisseha	Abstract	0%	--	--		1364362991	31-Jul-2020
Rahel Desta Fisseha	Conclusion	0%	--	--		1364363339	31-Jul-2020
Rahel Desta Fisseha	chapter 4	6%	--	--		1364363296	31-Jul-2020
Rahel Desta Fisseha	chapter 1	8%	--	--		1364362797	31-Jul-2020
Rahel Desta Fisseha	chapter 2	9%	--	--		1364363097	31-Jul-2020
Rahel Desta Fisseha	chapter 3	10%	--	--		1364363191	31-Jul-2020
Rahel Desta Fisseha	All thesis	13%	--	--		1364366444	31-Jul-2020

Prof. Dr. Ayşe Günay Kibarer

Head of the Department