

I. DATE OF PUBLICATION XXXX 00, 0000, DATE OF CURRENT VERSION XXXX 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

# Improved Threshold Based and Trainable Fully Automated Segmentation for Breast Cancer Boundary and Pectoral Muscle in Mammogram Images

Dilovan Asaad Zebari<sup>1</sup>, Diyar Qader Zeebaree<sup>2</sup>, Adnan Mohsin Abdulazeez<sup>3</sup>, Habibollah Haron<sup>4</sup>, Haza Nuzly Abdull Hamed<sup>5</sup>

<sup>1</sup>Center of Scientific Research and Development Nawroz University, Duhok, Kurdistan Region, Iraq

<sup>2</sup>Research Center of Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq

<sup>3</sup>Presidency of Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq

<sup>4,5</sup>School of Computing, Faculty of Engineering, University Teknologi Malaysia, Johor, Malaysia

Corresponding author: Diyar Qader Zeebaree (e-mail: [dqszebaree@dpu.edu.krd](mailto:dqszebaree@dpu.edu.krd), [dilovan.majeed@nawroz.edu.krd](mailto:dilovan.majeed@nawroz.edu.krd)).

This work was supported by the Duhok Polytechnic university, under Internal Research Grant

**ABSTRACT** Segmentation of the breast region and pectoral muscle are fundamental subsequent steps in the process of Computer-Aided Diagnosis (CAD) systems. Segmenting the breast region and pectoral muscle are considered a difficult task, particularly in mammogram images because of artefacts, homogeneity among the region of the breast and pectoral muscle, and low contrast along the region of breast boundary, the similarity between the texture of the Region of Interest (ROI), and the unwanted region and irregular ROI. This study aims to propose an improved threshold-based and trainable segmentation model to derive ROI. A hybrid segmentation approach for the boundary of the breast region and pectoral muscle in mammogram images was established based on thresholding and Machine Learning (ML) techniques. For breast boundary estimation, the region of the breast was highlighted by eliminating bands of the wavelet transform. The initial breast boundary was determined through a new thresholding technique. Morphological operations and masking were employed to correct the overestimated boundary by deleting small objects. In the medical imaging field, significant progress to develop effective and accurate ML methods for the segmentation process. In the literature, the imperative role of ML methods in enabling effective and more accurate segmentation method has been highlighted. In this study, an ML technique was built based on the Histogram of Oriented Gradient (HOG) feature with neural network classifiers to determine the region of pectoral muscle and ROI. The proposed segmentation approach was tested by utilizing 322, 200, 100 mammogram images from mammographic image analysis society (mini-MIAS), INbreast, Breast Cancer Digital Repository (BCDR) databases, respectively. The experimental results were compared with manual segmentation based on different texture features. Moreover, evaluation and comparison for the boundary of the breast region and pectoral muscle segmentation have been done separately. The experimental results showed that the boundary of the breast region and the pectoral muscle segmentation approach obtained an accuracy of 98.13% and 98.41% (mini-MIAS), 100%, and 98.01% (INbreast), and 99.8% and 99.5% (BCDR), respectively. On average, the proposed study achieved 99.31% accuracy for the boundary of breast region segmentation and 98.64% accuracy for pectoral muscle segmentation. The overall ROI performance of the proposed method showed improving accuracy after improving the threshold technique for background segmentation and building an ML technique for pectoral muscle segmentation. More so, this paper also included the ground-truth as an evaluation of comprehensive similarity. In the clinic, this analysis may be provided as a valuable support for breast cancer identification.

**INDEX TERMS** Breast cancer, Digital mammogram, Threshold technique, ML technique, Breast segmentation, Pectoral muscle segmentation.

## I. INTRODUCTION

Cancer is a deadly disease that affects humans globally. Cancer has been studied since the 1,900s and has been extensively recognized as an incurable disease [1]. Cancer involves the development and progressive growth of abnormal cells in the human body. Breast cancer, which is one of the earliest cancer types discovered in the world, was first documented in 1,600 BC in Egypt [2, 3]. Breast cancer is the second primary cause of death among women based on the statistics provided by the American Cancer Society in 2017 [4,5]. Hence, breast cancer cells should be detected at an early stage to decrease the mortality rate among women [6,7]. Based on the statistics provided by GLOBOCON in 2018, 2.08 million (24.2%) breast cancer cases have been recorded among all cancer cases diagnosed [8].

Mammography capturing of breast cancer is a standard procedure used to identify cancer cells at an early stage. However, the analysis of hundreds of mammogram images every day is impractical for radiologists; the task is time-consuming and exhausting, leading to false positives or false negatives. Mammograms are images that are hard to interpret because they include labels, pectoral muscles, and scratches. These artefacts contain high-intensity grey values, their visible appearance is close to abnormal images, thereby misleading presenting segmentation techniques and hindering them from segment accurate pathology-bearing regions. Thus, the suppression of these artefacts is a fundamental step [9, 10]. Given that the visual appearance of all images is remarkably close to each other, segmentation of the breast region is a crucial and important stage in Computer-Aided Diagnosis (CAD) systems [11, 12]. CAD is a designed approach that can minimize observational oversights, it has the possibility to improve the subjectivity of conventional analysis of histopathology images. The second reader concept of CAD is becoming a common system due to the reliability, consistency, and velocity of the system. Segmenting the breast region in a CAD system is considered a crucial phase utilized to accelerate the next processes with preserving the important anatomical information. Therefore, the region of breast segmentation and pectoral muscle is considered a challenging task in CAD systems, particularly in scanned mammogram images because of artefacts such as duct tape, light leakages, tags, and imperfections in the scanning process. More so, another challenge is a low contrast of the line of the breast skin (the line between the region of breast and background) and homogeneity among the region of the breast region and pectoral muscle [13, 14].

Accurate identification of the breast boundary of a mammogram presents two major problems. The contrast of the region close to the breast boundary decreases progressively due to imbalanced compression of breast tissue during acquisition. In the context of digitized mammographic images, the visibility of the breast boundary further reduces during digitization because of additional

noise. As a result, the breast skin line has low visibility, and the detection of the boundary of the breast becomes challenging. In addition, the non-uniform background contains high-intensity regions, such as labels, annotations, and frames as well as unexposed regions that adversely affect the segmentation of the breast area [15].

This study proposes a novel method for extracting breast boundary accurately and eliminating artefacts and noise. Extracting Region of Interest (ROI) in medical images is considered a crucial step. Usually, the main meaning of ROI is the significant or meaningful regions of the image. Generally, utilizing ROI has the ability to avoid irrelevant regions in the image as well as to obtain a rapid image processing method. For the ROI in an image to be accurately located, the proposed model uses wavelet transformation to suppress noise through the elimination of high-frequency bands and the preparation of an enhanced mammogram. The proposed model involves a process that is summarized as binaries of the input mammogram based on the proposed adaptive thresholding method. This step produces a binarized image that consists of a number of objects, such as breast image and artefacts. After that, disconnect the binarized objects by using the erosion process of morphological operations to obtain the number of independent binarized objects. The largest object, which is the region of breast, is retained within the image, and other small objects are deleted. The spiky boundary close to the edges of the image is isolated, and the rest of the region is smoothed through morphological dilation. The mask binary that consists of the background and the breast mammogram is obtained. The last step involves the use of the binary mask as a pointer to obtain the original value of the pixel of the mammogram. The ROI and pectoral muscle are obtained from the elimination of the background of mammograms and landmarks. The missing border is identified, and the ROI is segmented from pectoral muscle through post-segmentation by using a trainable Machine Learning (ML) algorithm. The interaction between experts with ML methods assists in leading to ameliorate the outcome of decision-making support systems. Due to having several factors that can affect the segmentation process such as the similarity between textures of both ROI and pectoral muscle, inhomogeneous ROI, missing border of ROI, and irregular ROI. These aspects suppose that in this study it is not sensible to depend only on the threshold to identify ROI. The threshold value depends on the intensity while by using ML we depend on blocks which are considered as local information not only one value that is considered as a poor value. This is the main reason which motivated us to use ML in pectoral muscle segmentation. To deal with these challenges we used the HOG feature which can capture the texture of the image as well as the specific and accurate information of the edge of the texture and neural network used as a classifier.

The rest of the paper is organized as follows. Section II presents the related work of the study. Section III describes the proposed segmentation method and is classified into two main sections that tackle the automatic method of segmentation based on the threshold and the proposed trainable segmentation. In Section IV, experimental results are discussed in detail. Section V concludes the study.

## II. RELATED WORK

In Computer-Aided Diagnosis (CAD) systems based upon mammography images, the breast features should be quantified automatically in order to provide for experts the clinical evidence. The mammogram segmentation is considered a key part of the CAD system, which can be used to estimate the breast skin line and the boundary of the pectoral muscle to identify breast counters. Mammogram segmentation of current studies has concentrated on the segmenting the boundary of breast and pectoral muscle. Such studies are mostly divided into five categories, namely, thresholding technique, morphology-based, active contour, region growing, and texture-based according to their segmentation techniques.

Thresholding techniques that are global or adaptive are commonly utilized to obtain the boundary of skin line in mammograms due to the remarkable intensity variation among foreground tissues and mammograms background. In contrast, due to the low contrast among the region of breast and pectoral muscle the thresholding techniques to obtain the pectoral muscle has limited. Moreover, the adaptability of the threshold value must be designed to identify regions with pectoral muscle [16]. [17] utilized the conjunction of a global thresholding technique based upon reducing the measures of fuzziness and implemented the detection of Sobel edge to determine the boundary of breast. The approach proposed by [13] is similar, except that the original image has been enhanced based on adaptive contrast enhancement before determining the value of the threshold. More so, to obtain overlapping masks the study proposed by [18] used several thresholding values, where the mean of gray level is the last value of threshold that located within the smallest and largest masks. Moreover, the combination between local thresholding technique and the algorithm of minimum cross-entropy thresholding has been utilized by [19] to determine the boundary of breast [20] proposed a segmentation method based upon region growing by initializing 40 points along the mask boundary through thresholding. The main drawbacks of this method are dynamic homogeneity, significant time requirement, and coarse contour. Methods based morphology [18, 21] natural shape features were utilized to construct complicated models which are fit the objects of the region of breast. The major drawback of these models driven techniques is that all the complex shapes that are shown in mammogram cannot be covered by generalized shape. On the other hand, another method widely utilized to segment the region of breast by initializing a breast boundary

and allowing the initialized breast boundary to method the actual boundary of breast based upon reducing energy functions. However, some active counter approaches based on edge [21, 22] deals with mammogram images only for the boundary skin line and pectoral muscle detection in the region of breast. Methods based upon texture, by using texture filters extracts texture from the image such as wavelet [23] or Gabor filter [24] and determine breast boundary among objects through remarkable changes in texture. A hybrid approach proposed by [25] by combining region-based active contour and a model-based approach; the method obtained very good results in experiments whereas had low accuracy in segmenting the boundary of pectoral muscle with complex contours.

Many segmentation methods have been proposed developed for the boundary of breast and pectoral muscle. However, only a few methods that use all the images in the mini-MIAS database have been evaluated quantitatively. The proposed algorithm of [26] is based on edge detection and scale-space concepts for breast region segmentation. A multi-level Otsu threshold an automatic algorithm was proposed to segment the breast region [17]. The introduced model of [27] is a fully automated pipeline based on a gradient weight map to estimate the breast boundary; the pectoral breast was detected based on the unsupervised pixel-wise labeling. Moreover, the performance accuracy of the proposed pectoral muscle segmentation model was compared with the method developed in [28], where an intensity-based technique was introduced for pectoral muscle detection. The  $3 \times 2$  mask filter was applied to enhance pectoral muscles, and the threshold was used to determine the boundary points of pectoral muscles. All detected points were connected to determine the boundary of pectoral muscles. The technique in [29] used the morphological opening to eliminate labels and annotations. In this technique, the Sliding Window Algorithm (SWA) was employed to remove pectoral muscles [30] employed region growing, thresholding, and k-means clustering to segment pectoral muscles at the first phase. An approach-based Machine Learning (ML) was employed to segment pectoral muscles.

Existing methods present several limitations. For the estimation of breast boundary, most thresholding techniques consider all grey levels in the image. The major problem of this technique is that it does not consider the non-homogeneity of the background of the image. As such the main problem resulting in under-segmentation is that the low contrast regions of the breast are considered as image backgrounds. By considering all grey levels in the image, the chosen value of the threshold is affected by artefacts (e.g. duct tape and tag). Edge-based active contour models are implemented on the original image to determine the final breast boundary. However, in many images, the boundary skin line (breast boundary) is unclear or obscured due to noise influence in the original image. For the segmentation

of pectoral muscle, by estimation approaches, the straight-line of pectoral muscle boundaries cannot be detected with curved shapes. Meanwhile, because of the homogeneity between pectoral muscle and breast region both techniques thresholding and region growing are failed in the segmentation tasks between them.

### III. Proposed Method

A mammogram is commonly used in imaging systems in the field of gynecology to detect ROI for diagnosing breast cancer. A mammogram is also used to monitoring specific diseases. The automatic analysis of a digitalized mammogram by using a computer requires its segmentation into different anatomical regions. Segmentation is important because it limits abnormalities in search zones to the relevant breast region without unnecessary interference from the image background.

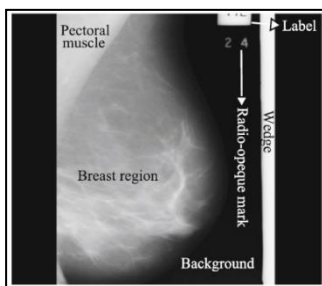


Figure 1: Original Mammogram Case mdb051 from mini-MIAS.

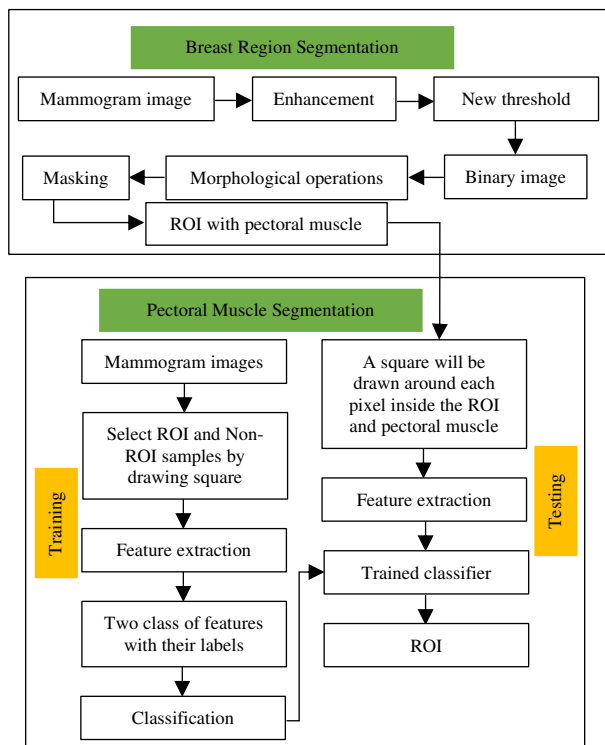


Figure 2: Fully Automatic Segmentation System.

Typically screening in mammogram images involves capturing two different views, namely, Mediolateral Oblique (MLO) view which is taking from angled or oblique while Cranial-Caudal (CC) view which is taking from above. The view of MLO is preferred through routine screening of mammogram images due to imaging more tissue of the breast in the armpit as well as in the upper outside quadrant. Mammogram typically includes artefacts and annotations, which could adversely affect the analysis. Pectoral muscle appears in the MLO view of mammogram images as high-intensity and a triangular region on the upper posterior margin. In mammograms, the kinds of noise that can be observed include low- and high-intensity label and tape artefacts (Figure 1). This study proposed a novel method to extract ROI from mammogram images, the general proposed method illustrated in (Figure 2). First, to enhance mammogram through wavelet transform techniques. Second, to estimate the breast region using a new threshold technique. Finally, a machine learning technique has been built to segment pectoral muscle from ROI.

The following sections show detailed illustrations of the proposed model for the automatic segmentation of ROI from mammogram images to identify abnormal cases.

#### A. IMAGE ACQUISITION

Whole images of the Mammographic Image Analysis Society (Mini-MIAS) were used to evaluate the proposed segmentation method. This database comprises 322 mammograms, which are freely available to the public for use in scientific research. All 322 mammogram images were collected from 161 pairs of MLO views from the right and left views. These mammogram images were obtained from an imaging process of film screen conducted by a program of the national breast screening in the UK. The mammogram images are divided depending on intensity into different categories, including glandular, fatty, and dense, which consists of 104, 106, and 112 mammograms, respectively. This database contains two major categories, namely, abnormal and normal mammograms. The former contains 207 benign and 115 malignant mammograms, with a total of 322 [7, 31].

The dataset from INbreast has been collected at a university hospital in Portugal (Centro Hospitalar de S. Joao [CHSJ], Breast Centre, Porto) under a convention with the Portuguese National Committee of Data Protection. This dataset was composed of 115 patients (cases). From each of 90 patients, 4 mammogram images have been collected with affected of both breasts (right and left) while 50 mammogram images from remain cases (25 patients) have been collected with mastectomy patients. Therefore, 410 normal, benign, and malignant cases of mammogram images were collected including MLO and CC views. This database is available publicly which contains 410 mammogram

images including asymmetries, masses, calcifications, and distortions [32].

Breast Cancer Digital Repository (BCDR) database is one of the newer SFM datasets. BCDR contains 1125 studies from both views MLO and CC totaling 3703 mammogram images. Each of them is provided by 8-bit TIFF images and with 720×1168 pixels resolution. Mammogram images in the BCDR database which are formatted with 8-bit TIFF are not difficult to work with them in the process. This dataset can be used as a benchmark dataset for CAD models. Figure 3 shows some samples of mammogram images from each dataset [32].

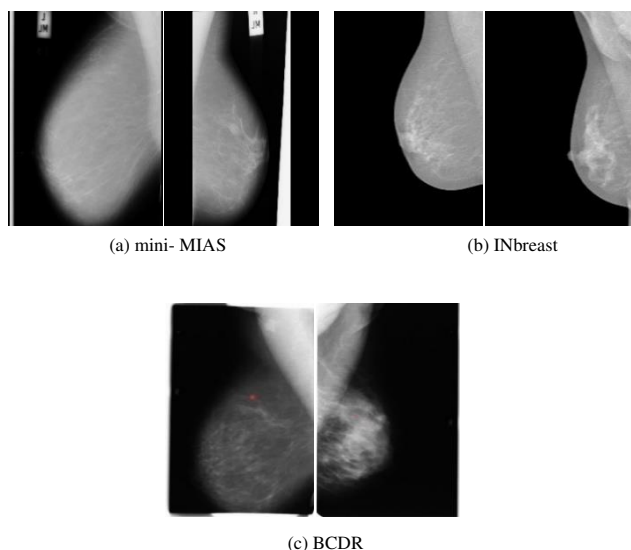


Figure 3: Sample Images from Used Datasets.

Nowadays, most of the datasets of mammogram images do not contain any ground-truth or annotations from the expert radiologist. This makes the quantitative evaluation for mammogram images quite difficult [27]. For the mini-MIAS dataset, the ground-truth has been generated by annotating the boundary of the pectoral muscle for each counter [33]. This process has been done with the help of a clinician which is supervised closely by an expert radiologist. Regarding INbreast [34] ground-truth was generated, by providing the annotation of the dataset under supervised by an expert radiologist. Finally, the boundaries of the pectoral muscle of mammogram images from BCDR have been provided with the assistance of an experienced observer which is confirmed by an expert radiologist [35].

## B. IMAGE ENHANCEMENT

The mammogram image contrast should be improved to display specific regions, such as ROI. The mammogram images have low contrast, which should be enhanced without removing the details. The use of the original image for segmentation can suffer from segmentation problems that are over and under segmentation. Over segmentation is obtain apart from the background as an ROI, and the segmentation method will result in under-segmentation in the whole ROI.

Over- and under-segmentation problems occur mainly because ROI involves a poor edge and is not clear for human vision. This problem will reduce the diagnostic accuracy because the missing part of the ROI might change the final decision. To overcome this problem, the present work proposes an algorithm to highlight the ROI and image background and create a large difference between them through wavelet decomposition. Enhancing the mammogram is very important to highlight the ROI and clarify the border of the ROI. This process will help the segmentation method to extract the whole ROI from the background.

Wavelet Transform (WT) is considered as a common example frequency domain. In the transformation domain, WT can be utilized as data filtering. The basic idea of utilizing WT as a filter is that this technique has a capability in image separation into four different groups. The first phase of the disintegration step is the place the information of any image into four different sub-band groups low-low (LL), HH (high-high), low-high (LH), and high-low (HL) groups. Each group of sub-bands presents specific information according to the image. The diagonal information, even data, vertical data, and the low-frequency data of the image are represented by HH, HL, LH, and LL, respectively [36]. The following phase of the procedure of disintegration incorporates further segmentation of the LL sub-band as it is shown in Figure 4.

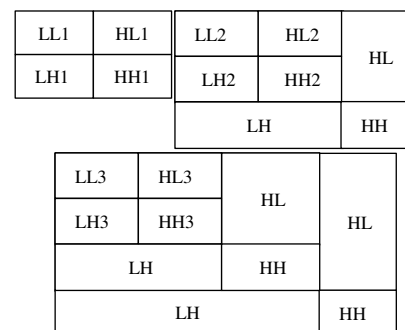


Figure 4: Multiresolution Scheme after Several Levels of Wavelet Transform.

Wavelet has a different property from Fourier Transform which depends on time-frequency. In other words, wavelet has the ability to separate high frequencies from low frequencies with preserving the location of pixels. This technique assists in discovering high frequencies and analyzing them in a better way. Based on the existing frequencies noise can be removed from the image because as it has been observed that noise always in high frequencies. Due to this reason, this study has been motivated to use the wavelet technique to remove sub-bands that hold noises and highlight ROI.

The enhancement of the mammogram images is crucial because it enables the reduction of noise while preserving salient tissue boundaries. This phase facilitates the accurate

identification of structure boundaries, enhanced visualization of the position of structures, and quantification of the morphology. Noise is the primary challenge associated with the analysis and segmentation of mammograms, but the essential techniques depend on the objective of pre-processing. This study reduced the noise by wavelet transform. In theory, noise is contained by bands with high frequency and is mostly located in the HH band. The wavelet transform approach eliminates bands within one level of decomposition, and one component with high frequency is eliminated at least which are HL, LH, and HH bands. A close examination of the mammograms showed that they were processed with HH, LH, and HL bands with different options, where some white spots were found. These spots were not considered as part of the original mammogram, thereby adversely affecting the metric results. The high-frequency component of the horizontal edges is located in LH, while the examined mammogram possesses more vertical images than the horizontal edges, although the majority of the lines are diagonal. Therefore, if this band is eliminated, then much noise will be eliminated, while this not related much with the information of edge. To ascertain the level of decomposition that is required for the elimination of noise while maintaining the mammogram image, we tested various bands at the first level. As shown in Figure 5, the best-produced mammogram is generated when HH, LH, and HL bands were eliminated, where extremely high contrast was observed between the background and the region of the breast and artefacts were detected. This study removed HH, LH, and HL bands before implementing a new threshold technique to binarize the mammogram image.



**Figure 5:** Eliminating HL-LH-HH Bands based on Wavelet Filter Using Different cases of mini-MIAS images.

### C. BREAST REGION SEGMENTATION

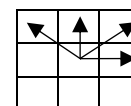
This stage of the proposed segmentation method focuses on the development of an automatic method to improve thresholding and distinguish among various kinds of breast cancers (benign and malignant). The image must first be segmented prior to the automatic extraction of texture features. Nevertheless, the most important determinant of accurate segmentation is the mammogram quality because the segmentation can be made more challenging because of the presence of artefacts and noise. The implication of these quality impacts can be missed boundaries as a result of having these artefacts and the capturing direction of low contrast between the ROI. Therefore, the proposed method

extracts the breast from the mammogram background, determines artefacts, and eliminate them.

#### 1) NEW THRESHOLD TECHNIQUE

Figure 1 illustrates the rough segmentation of a pectoral region and the presence of artefacts (i.e. wedges and labels). The artefacts are removed by employing the described technique that establishes a cut-off search limit within a breast region, thereby minimizing the occurrences of false positives in detecting ROIs. Therefore, the segmentation of a mammogram is essential because it is critical in detecting suspicious regions with breast cancer. Mammogram segmentation mainly aims to distinguish ROI from the background and tissues that surround the ROI. Segmenting ROI is a difficult task because the homogeneity characteristics of ROI are entrenched with an uneven background of breast tissue, causing the discriminative task of distinguishing suspected regions in medical images a substantial endeavour [37, 38]. Thus, techniques of ROI segmentation in mammograms should be developed prior to emerging key features to inform medical practitioners of the presence of breast cancer. For accurate segmentation of the breast region from mammograms, the present work proposes a new thresholding-based segmentation approach for image binarization and elimination of artefacts. The proposed method used different features including entropy, mean, and median.

Grey-level Co-occurrence Matrix (GLCM) is a robust statistical tool that is used to extract a group of texture features from images. Entropy and mean are two features that are derived from GLCM. The main aim of this study is to build a model to segment ROI from un-wanted RIO. Thus, in this study, we used GLCM as a technique to support our proposed method and to propose a new threshold value based on GLCM. GLCM is a robust statistical tool feature group for extracting second-order texture information from images. GLCM exhibits how the pixel brightness in an image occurs. A matrix is built up at a distance  $d=1$  and at angles in degrees (0,45,90,135), Figure 6 shows the directions of GLCM. Initially, GLCM considered a group of features that rely on statistics in the second order. These features can be utilized in terms of uniformity and homogeneity to reflect for correlation degree of the overall average between every two pairs of image pixels in various aspects. The distance separation between image pixels is considered as a key factor which can influence the discrimination abilities of GLCM. When we pick a value of 1 as a distance, the correlation degree could be reflected between image adjacent pixels whereas when we increase the value of distance, the correlation degree could be reflected between image distant pixels.



**Figure 6:** Four directions of adjacency for calculating the GLCM features.

The GLCM characterizes the spatial allocation of gray levels in the ROI which has been selected. An element at position (i,j) of the GLCM indicates the joint probability density of the occurrence of gray levels i and j in a specified orientation  $\theta$  and specified distance d from each other (Figure 6). Therefore, for various  $\theta$  and d values, various GLCMs are generated. Figure 7 illustrates how a GLCM with  $\theta = 0^\circ$  and d = 1 is generated. The number 4 in the co-occurrence matrix represents that there are 4 occurrences of a pixel with gray level 3 immediately to the right of the pixel with gray level 6.

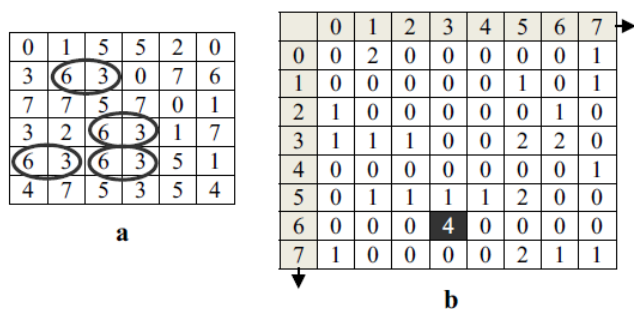


Figure 7: Construction principle of co-occurrence matrix: (a) Initial image, (b) co-occurrence matrix ( $\theta = 0^\circ$  and d = 1).

Mean is a very important measure in digital image processing and is used in spatial filtering and noise reduction. The mean value is defined in Equation 1. Entropy refers to the statistics used to quantify arbitrariness and is widely employed to distinguish and extract the statistical texture features of an image, Equation 2 showed the value of entropy. The significance of entropy in texture feature is reflected in the existence of numerous state-of-the-art literature that proposes effective classifiers for mammograms to accurately distinguish between abnormal and normal masses. Entropy is the measure of randomness (or uncertainty) in an image and the information transmitted. The concept was introduced by Claude Shannon and is called Shannon's entropy. The maximum, Renvi, Tsallis, spatial, minimum, conditional, cross, relative, and fuzzy entropies are used for image segmentation, image registration, image compression, image reconstruction, and edge detection in grey-level images [39]. Based on our investigation using only one of (mean, median, and entropy) will suffer from over-segmentation or under segmentation. Thus, Mean, median, and entropy are calculated based on sub-regions which means calculating from local information, not from global information (whole image). From each region, different objects have been obtained using each of mean, median, and entropy. When one of mean, median, and entropy suffer from over or under segmentation, it can be solved based on another one, as illustrated in Figure 8. Therefore, a more effective segmentation technique has been proposed based on the combination of them which leads to obtaining better results.

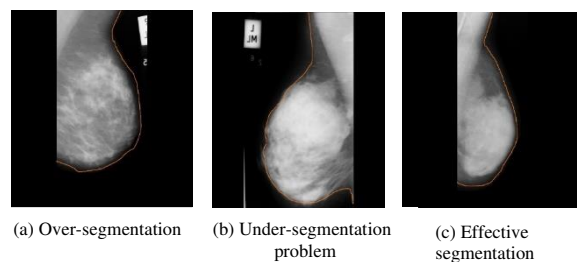


Figure 8: An Example of Breast Segmentation.

$$Mean = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} ip(i, j) \quad (1)$$

$$Entropy = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} [p(i, j) \log(p(i, j))] \quad (2)$$

where denoting by: Ng the number of gray levels in the image, p(i,j) the normalized co-occurrence matrix.

The process of dividing pixels into two different groups is called binarization, wherein white is designated as the breast region including pectoral muscles and black is designated as the background of the mammogram. The correct extraction possibility of the breast region from the background can be acquired through robust binarization. Many techniques have been performed for image binarization, but thresholding has sufficient accuracy with high-speed processing to segment greyscale images. In this process, the image's greyscale is transformed into image binary, where the value of each pixel is either zero or one. The black colour represents zero, and the white colour represents one. The binary mammogram can be processed better than the greyscale image, resulting in easy further processing. The initial principle of transforming the original mammogram into a binary mammogram involves the selection of a strong threshold value. The mammogram pixels are then compared with the threshold value to convert the greyscale mammogram to a binary mammogram that consists of white and black pixels.

The difficulty of binarization lies in selecting the strong threshold value, which can differentiate between the region of the breast, artefacts, pectoral muscles, and background. Large variations exist between mammogram images where their background colour is darker than the others, and also some images contain artefacts while others do not. Therefore, a strong threshold value that can work on all mammogram images is difficult to determine. In this regard, the segmentation task was carried out using the proposed threshold technique of texture features. An adaptive threshold value for mammogram binarization was calculated based on median, mean, and entropy. These features were extracted for each pixel based on its local information. Each pixel was presented by three values, which were used to binarize the image. The proposed method is based on three features that are extracted from each window around the pixel. At the beginning of the process, the image pixels were scanned successively. For each pixel, a window around the

pixel was obtained, and the three values were extracted from the window. Accordingly, three images were obtained, and one of them consists of median, mean, and entropy values. Each image was subjected to thresholding by taking the mean of each produced image. In this adaptive technique, the mammogram was divided into blocks of pixels. Each pixel was compared with three values of mean, median, and entropy, and the value of the selected pixel was set as 1 when it is larger than mean, median, and entropy; otherwise, it was set to 0. Each pixel contains three decisions (includes 0 or 1). The major decision was used to decide if the pixel is 1 or 0. The final decision was determined by the pixel with the highest number of votes. When the major decision was obtained, a single image with binary values can be produced. Figure 9 presents the proposed threshold technique for the binarization of mammograms.

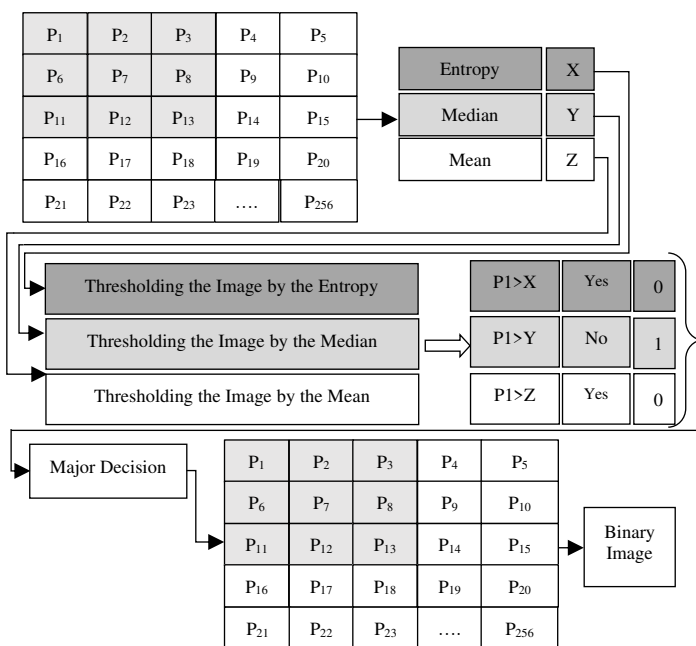


Figure 9: The Proposed Threshold Value for Mammogram Binarization.

An adaptive threshold technique was utilized to binarize the mammogram. In this process, every 8-bit grey value of the mammogram was converted into a 1-bit value, with 1 for breast region or pectoral muscle and 0 for mammogram background. Both regions of the breast and pectoral muscles were highlighted with white, whereas the mammogram background was highlighted with black (Figure 10). This process will improve the contrast among regions of the breast, pectoral muscle and mammogram background, and facilitate their extraction.



Figure 10: Obtained Binary Mammogram Based on the Proposed Threshold.

## 2) MORPHOLOGICAL OPERATION OVER BINARY IMAGE

The proposed thresholding method was applied to produce a binary mammogram. The binary image consists of a white region considered as the breast region and some landmarks with a considerable number of some other objects. Small objects were eliminated through morphological operations. The structure or shape of an object in a binary mammogram is affected by binary morphology operations. These operations are mostly performed during pre- and post-processing or even in the extraction of the characteristics of objects (regions) in a mammogram [40]. Binary morphological operations mainly involve dilation and erosion. Dilation involves the growing of some objects in mammogram binary images, whereas erosion is the operation through which objects in the binary mammogram images are thinned. The two operations are controlled by element structuring. Equations 3 and 4 can be used to express dilation and erosion mathematically:

$$A \oplus B = \{z | (\hat{B})_z \cap A \neq \emptyset\} \quad (3)$$

$$A \ominus B = \{z | (B)_z \cap A^c \neq \emptyset\} \quad (4)$$

where the binary image is denoted by  $A$ ,  $A^c$  denotes the supplement of  $A$  and  $(\hat{B})_z$  is an element of the  $B$  structure after being reflected and translated  $z$ . The dilation of  $A$  by  $B$ , which is represented as the set of all segregation  $z$ , where the overlapping occurs in  $A$  and  $\hat{B}$  depending on at least one element. The image  $A$  erosions occur depending on the structure of element  $B$ , which is considered as a set of entire structuring element origin positions, where no overlap occurs between the translated  $B$  and the background of image  $A$ . Opening and closing are two relevant morphological operations that are determined by integrating dilation and erosion. The object's contour can be smoothed through the opening, while some unwanted small objects are reduced. The opening can be implemented using the erosion process then followed by the process of dilation. Closing is a process through which cramped gaps are closed, and the small holes are filled to smoothen the objects. Closing is fulfilled by the process of dilation then followed by the process of erosion.



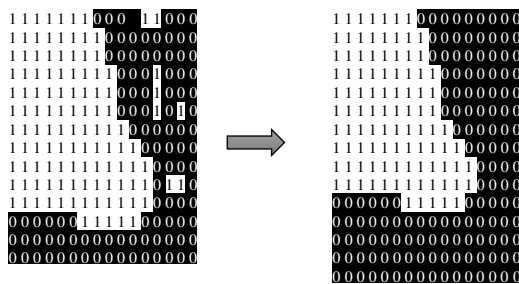


Figure 11: The Process of Morphological Operations.

Generally, the performance of the binary process of morphological operations and spatial filtering is similar. Both processes of a segmented binary mask are significant in pre-processing and post-processing. Two factors are considered when such processes are used. First, parameters including orientation and size should be carefully set to obtain the best result. Apart from the intended improvements in some regions in the image, they may be affected negatively by the uninformed application of those operations on the entire image. Figure 11 illustrates the morphological operation.

A binary mammogram can be represented by a matrix of pixels, which are represented by ones (white colour) and zeros (black colour). The breast region and pectoral muscle are represented by one-pixels, whereas the background is represented by zero-pixels. For preserving the essential features of the breast region, this process is used to reduce and simplify the shape. After applying this process, the topology of the original region is retained while most of the unwanted pixels are converted to zero. The result of morphological operations is presented in Figure 12.



Figure 12: Detecting and Removing Small Objects via Morphological Operations.

### 3) MASKING PROCESS

The last process of breast region segmentation is masking to retrieve the original pixel values. In digital image processing, masking involves changing the colour of certain areas of a picture or transferring these areas onto another background. Firstly, this process requires the clipping of relevant areas. The locations of the breast region pixels are used to generate the mask and ensure that any pixel value of the breast region is not missed. The intensity of the pixel values of the background region is assigned to zero. Along the segmented breast region border, a hard edge is created.

Artificial hard edge leads to the accumulation of undue saliency along the breast border. Figure 13 shows the masking process is utilized to eliminate the background from the breast region. The binary mammogram is transformed again to the greyscale mammogram with the same pixel values of the region of the breast. This process will help to select the cancer area also pervasion out of the cancer cells. However, an area called the pectoral muscle remains, which could affect the next stage of CAD processing. Thus, post-segmentation is used to segment ROI from pectoral muscles. Figure 14 shows an example of the process of extracting the breast region from the background.

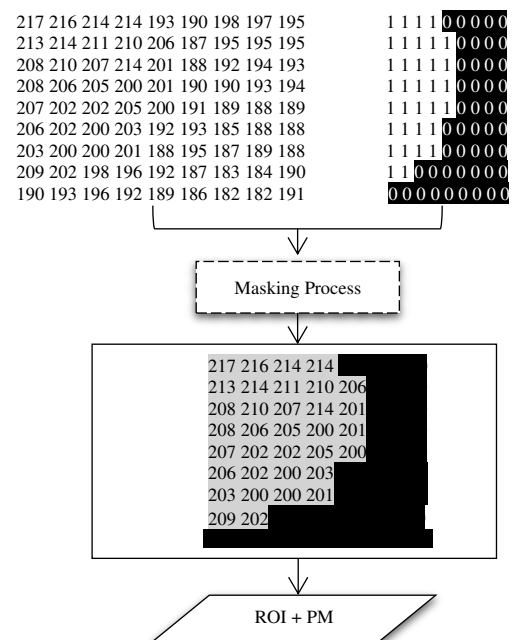


Figure 13: The Numerical Representation of Masking Process.

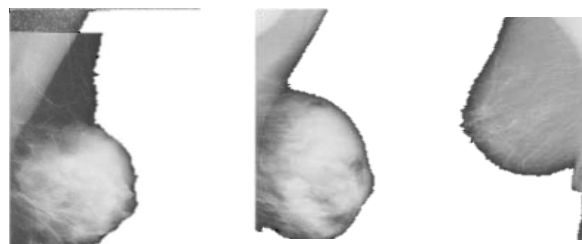


Figure 14: Retrieving Original Pixel Values from Original Mammogram via Masking Process.

### 4) PECTORAL MUSCLE SEGMENTATION

Histogram of Oriented Gradients (HOG) is a global feature descriptor that depends on the allocation of intensity gradients or the orientations of edge. This tool aims to quantify oriented gradient in confined image segments.

Based on the pixel value or the light amount, the shape and appearance of the image can be described. Alteration directional in intensity is considered a gradient or the image colour. Important image information can be extracted by utilizing the HOG standard deviation, mean, entropy, and variance features that can be determined with HOG. Moreover, HOG is used to identify objects in digital images [41].

In the processing of mammograms, breast cancer detection results are prone to biases when the segmented pectoral region is viewed from top to bottom in the MLO view in mammograms due to procedures associated with segmenting the pectoral region of a dominant dense region that contains wedges. Accordingly, the pectoral region is often segmented with restriction to search only in the suspicious region that is concentrated on the breast's soft tissue. This process could significantly increase the accuracy outcomes for the detection of abnormal masses in mammograms. Thus, the proposed threshold technique for image binarization is followed by filtering unwanted objects based on the training and testing models to extract the pectoral region (Figure 10). The detection of breast cancer in medical images relies heavily on one of the most critical parameters, which is texture. This parameter is important in identifying ROIs and objects encompassing different types of images. The texture is also essential in classifying, detecting, and segmenting based on colour and intensity. The analysis of texture involves the extraction of features from the treated image through the proposed technique. The texture of an image comprises a collection of pixels or closely associated pixels.

In a previously proposed segmentation model, the image is binarized to obtain the ROI with other ROI because of the similarity between textures. However, the breast region is segmented from the mammogram background and artefacts. This model aims to isolate the ROI from the unwanted ROI (pectoral muscle). This stage is difficult because of the overlap between ROI and unwanted ROI (pectoral muscles). Machine learning methods and HOG features are used models to isolate and extract the ROI. HOG can extract the features of the pixels based on the information of the neighbor pixels. HOG can calculate the gradient of the region, and ROI has a special texture that can help in identifying the ROI from unwanted ROI. A trainable model, which consists of training and testing models, is built to segment ROI from unwanted ROI (pectoral muscle). In training, several mammogram images were selected manually from the mini-MIAS dataset.

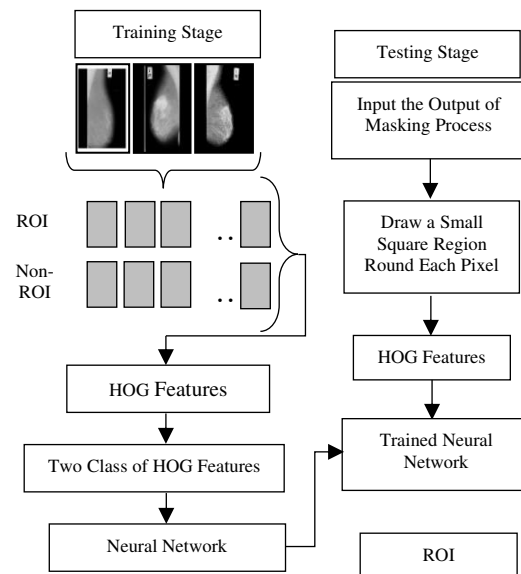


Figure 15: Trainable Segmentation Model.

Depending on the selected mammograms, several blocks from both ROI and unwanted ROI (pectoral muscle) of different samples were selected. The blocks were named ROI and non-ROI (pectoral muscle). The proposed model used 25 samples of mammograms for training. The training images consist of ROI and pectoral muscle, 20 regions were selected from each sample 10 regions inside ROI which are called ROI area whereas 10 regions outside of ROI which means from pectoral muscle region. Finally, 500 regions were selected from 25 images, including 250 ROI and 250 non-ROI regions. In this model, a single descriptor feature was employed to describe ROI and unwanted ROI and to identify the ROI effectively.

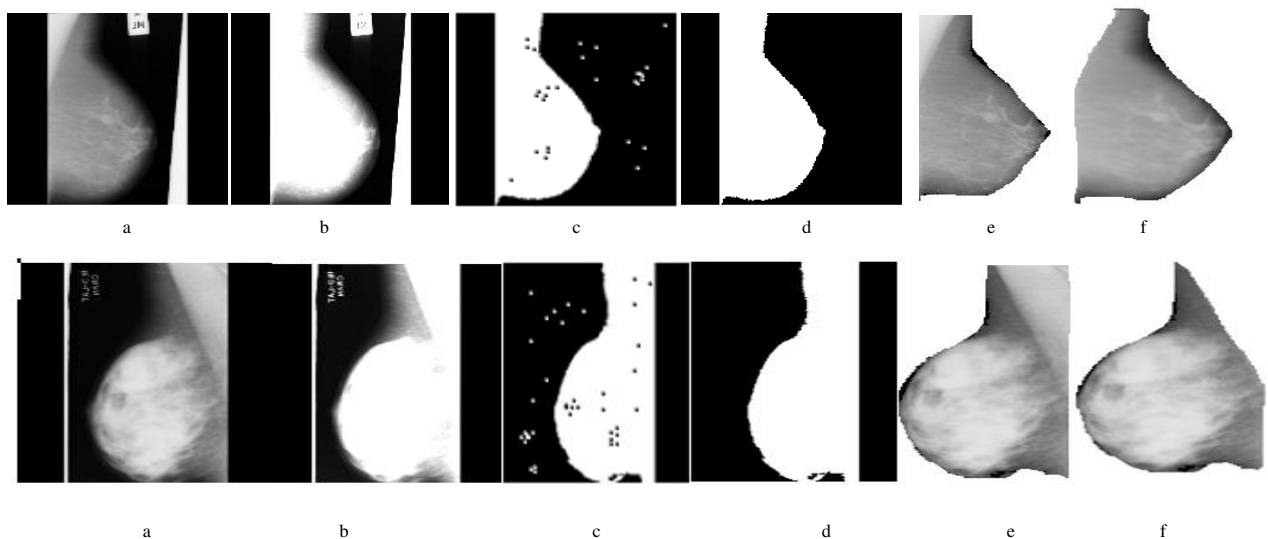
For each block, a collection of HOG descriptor features was extracted. The labeled features from blocks on the selected samples were trained based on Neural Network (NN) classifier. In classification, a two-layer backpropagation neural network has been used. The structure of the used backpropagation classifier is consists of two hidden layers and one output layer. After completing the training model and selecting each mammogram, the testing model was employed as an input for the proposed method. All pixels of the input mammogram were scanned individually. For each mammogram pixel, a small square region was built in the input mammogram, and this region has the same size as the window with the pixel as the center. Thereafter, HOG features were extracted from the region and fed into the trained NN to classify ROI and unwanted ROI (pectoral muscle). When the region was within an ROI, the central pixel was named ROI; otherwise, it was named as unwanted ROI (pectoral muscle). After labeling the pixels of the mammogram, the region of the “inner pixels” was considered as the segmented ROI. Figure 16 illustrates an example of the pectoral muscle segmentation.



**Figure 16: Pectoral Muscle Removal Based on the Proposed Trainable Segmentation.**

Figure 17 shows the outcomes in the succession of the fully proposed segmentation model steps on different sample images of the mini-MIAS database. Figure 17 (a) displays the first sample image with a label-marker in the upper right

corner. Figure 17 (b) presents the output of the enhanced and highlighted images before segmentation. Figure 17 (c) shows the outcome of the proposed new threshold value for the conversion of the image into a binary image. Figure 17 (d) shows the process of extracting the breast region from its background and from exhibits the final label-marker and the other artefacts suppressed image. Another step was implemented [Figure 17 (d)], where morphological operations are employed to disconnect objects of the binary image from each other by using erosion. The largest object (region of breast) was preserved, and all small objects were eliminated after calculating them. Figure 17 (e) indicates the masking process to retrieve the original pixel value of the region of the breast. Figure 17 (f) shows the extracted ROI from the pectoral muscle.



**Figure 17: Fully Automatic Segmentation Steps for Two different mini-MIAS Images.**

#### IV. EXPERIMENTAL RESULTS

The proposed study is validated through different numbers of experiments. The experiment results are performed utilizing MATLAB (2020b) with the Core-i7 processor, RAM 32 GB, and Windows-10 operating system. In the analysis of mammogram breast images, regardless of the segmenting pectoral muscle, the ROI concentrated on the breast region segmentation. Noise, background including artifacts, and pectoral muscle should be eliminated in sequence. The breast region in the mammogram is known as the region among the line through the background whereas the ROI is defined as the area breast line and pectoral muscle. Thus, this study evaluates the strategy that the segmentation of the region of the breast including pectoral muscle (ROI plus pectoral muscle), and segmenting ROI from the pectoral muscle (obtaining ROI). The proposed method is measured based on the previous steps as the performance of segmenting ROI from the background and pectoral muscle, respectively.

The proposed model was evaluated using 322 MLO mammogram of mini-MIAS, 200 MLO mammogram of INbreast, and 100 MLO mammogram of BCDR databases giving a total of 622 MLO mammogram. The validated has been done by evaluating the performance at four different stages. Firstly, to evaluate the performance of the proposed segmentation method, accuracy, sensitivity, and specificity were used. Secondly, automatic measurements were compared with the manual ones produced by an expert. Thirdly, the automatic measurements were evaluated in terms of accuracy in detecting early cancer cases. Finally, the proposed fully automatic segmentation was compared with traditional techniques and with recent previous studies as benchmarking [42, 43].

In the strategy of evaluation, five metrics were used to evaluate the performance of the proposed segmentation method. Sensitivity is used to deal only with positive cases (cancer cases); it presents the proportion of the detected positive cases over the actual positive cases, the higher the sensitivity, the lower the false-negative rate. Sensitivity

(Sen) can be calculated by implementing Equation (5). Specificity (Spe) is used to deal only with negative cases (healthy cases); it reflects the proportion of the detected negative cases over the actual negative cases, the higher the specificity, the lower the false-positive rate. Specificity can be calculated by implementing Equation (6). Accuracy (Acc) (classification rate) measure denotes the correctness of the proposed detection method. Accuracy can be used to deal with all cases; it indicates the precision of predict results. Accuracy can be calculated by implementing Equation (7). Jaccard Index (Jac) it is also called the Jaccard similarity coefficient, this considers as a statistic utilized in comprehending the resemblance between image sets. The measurement confirms the resemblance between finite sets of samples. In formal, it can be defined as the intersection size divided by the union size of the sets of samples. Thus, the similarity and difference among the results of ROI segmentation and the ground truth which is calculated by implementing Equation (8). The dice coefficient is also called a dice similarity coefficient, it is a statistical measurement that can assess the resemblance between two different sets of data. it is used to fully assess the proposed segmentation performance within the similarity between two sets of data have been evaluated based on the dice coefficient which has been calculated using Equation (9) [27].

$$\text{Sensitivity} = TP / (TP + FN) \quad (5)$$

$$\text{Specificity} = TN / (FP + TN) \quad (6)$$

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (7)$$

$$\text{Jaccard Index (Jac)} = \frac{|ROI_{pm} \cap ROI_{gt}|}{|ROI_{pm} \cup ROI_{gt}|} \quad (8)$$

$$\text{Dice} = 2 \frac{|ROI_{pm} \cap ROI_{gt}|}{|ROI_{pm}| + |ROI_{gt}|} \quad (9)$$

Where True Positive (*TP*) ill cases have been diagnosed correctly. False Positive (*FP*) ill cases have been identified incorrectly. True Negative (*TN*) healthy cases have been identified correctly. False Negative (*FN*) healthy cases have been identified incorrectly.  $ROI_{pm}$  was the area of ROI segmentation using the proposed method, and  $ROI_{gt}$  was the ROI of the ground truth.

### A. CLASSIFICATION RESULTS

Automatic ROI segmentation is simple and shares the same texture colour as ROI. The background of the image and the pectoral muscle of the MLO views in mammograms are considered difficult tasks. This study proposed a fully automatic segmentation technique to crop ROI. In general, expert measurements are important for the evaluation of the performance of any segmentation technique. A comparison with the automatic segmentation technique should be conducted to show the closeness and correlation between them. The datasets of the mini-MIAS, INbreast, and BCDR databases were used to examine the proposed technique.

**Table 1.** COMPARISON BETWEEN THE PROPOSED SEGMENTATION METHOD AND THE MANUAL SEGMENTATION BY USING THE TEXTURE FEATURE

| Features                    | Dataset   | Cancer Subtype |           | Number of Images |     |    |    | Percentage (%) |             |             |
|-----------------------------|-----------|----------------|-----------|------------------|-----|----|----|----------------|-------------|-------------|
|                             |           | Benign         | Malignant | TN               | TP  | FN | FP | Accuracy       | Sensitivity | Specificity |
| FD for manual segmentation  | mini-MIAS | 207            | 115       | 85               | 186 | 30 | 21 | 84.16          | 86.11       | 80.18       |
| FD for proposed solution    |           | 207            | 115       | 78               | 176 | 37 | 31 | 78.88          | 82.62       | 71.55       |
| LBP for manual segmentation |           | 207            | 115       | 81               | 179 | 34 | 28 | 80.74          | 84.03       | 74.31       |
| LBP for proposed solution   |           | 207            | 115       | 75               | 176 | 40 | 31 | 77.95          | 81.48       | 70.75       |
| FD for manual segmentation  | INbreast  | 127            | 73        | 52               | 119 | 21 | 8  | 85.5           | 85          | 86.66       |
| FD for proposed solution    |           | 127            | 73        | 47               | 112 | 26 | 15 | 79.5           | 81.15       | 75.8        |
| LBP for manual segmentation |           | 127            | 73        | 50               | 115 | 23 | 12 | 82.5           | 83.33       | 80.64       |
| LBP for proposed solution   |           | 127            | 73        | 49               | 113 | 24 | 14 | 81             | 82.48       | 77.77       |
| FD for manual segmentation  | BCDR      | 50             | 50        | 44               | 35  | 6  | 15 | 79             | 85.36       | 74.57       |
| FD for proposed solution    |           | 50             | 50        | 40               | 35  | 10 | 15 | 75             | 77.77       | 72.72       |
| LBP for manual segmentation |           | 50             | 50        | 43               | 34  | 7  | 16 | 77             | 82.92       | 72.88       |
| LBP for proposed solution   |           | 50             | 50        | 41               | 33  | 9  | 17 | 74             | 78.57       | 70.68       |

Depending on the texture feature, breast cancer subtype has been identified. ROI should be used to extract texture features to obtain specific information related only to the ROI from the mammogram. Several studies used texture features to detect the risk of breast cancer and classify it as benign or malignant. In this paper, Local Binary Pattern (LBP) texture and Fractal Dimension (FD) texture features were selected. Firstly, ROI was cropped manually from mammograms, LBP, and FD features were extracted, and Artificial Neural Network (ANN) classifier was utilized for diagnosing the case. Subsequently, ROI was segmented from a mammogram based on the proposed segmentation model. FD and LBP features were calculated and fed to the ANN classifier. This process showed the closeness between the proposed segmentation model and the manual segmentation. Table 1 presents the diagnosis results of both manual cropping and cropping based on the automatic proposed model. The remarkable closeness between the manual crop and the proposed model was shown, indicating that the most important region was cropped from the mammogram by using the proposed model.

The performance evaluation of the final trained method for different datasets has evaluated on both train and test sets. The prediction based on the different confusion matrices. Figure 18 illustrates the classification results utilizing the

proposed method. As it is depicted if Figure 18 (b), the proposed method accurately predicts 254 instances out of 322 cases in the test set for the mini-MIAS dataset. Based on the confusion matrix of the proposed method using a neural network, from 115 breast cancer images of malignant cases 78 images the rate of 67.82% are identified correctly whereas 37 images the rate of 32.18% are incorrectly diagnosed. However, from 207 images of benign cases of breast cancer, 176 cases in the rate of 85.02% are diagnosed correctly while 31 cases in the rate of 14.98% were misdiagnosed. For INbreast dataset, Figure 18 (c) shown that the proposed method accurately predicts 162 cases out of 200 cases in the test set. It is illustrated that from 73 malignant cases 49 cases with a rate of 67.12% are diagnosed correctly whereas 24 cases with a rate of 32.88% are diagnosed incorrectly. On the other hand, from 127 benign cases, 113 instances with a rate of 88.97% are identified correctly while 14 benign cases with a rate of 11.03% were misdiagnosed. Furthermore, for the BCDR dataset 100 images were used with 50 cases as malignant and 50 cases as benign. From Figure 18 (d) it is shown that 40 and 35 cases were identified correctly with a rate of 80% and 70% from malignant and benign cases, respectively. In contrast, 10 and 15 cases with a rate of 20% and 30% cases were misdiagnosed.

|              |          | Predicted class          |                          |          |
|--------------|----------|--------------------------|--------------------------|----------|
|              |          | Positive                 | Negative                 |          |
| Actual class | Positive | TP                       | FN                       | TP rate  |
|              | Negative | FP                       | TN                       | TN rate  |
|              |          | Positive predicted value | Negative predicted value | Accuracy |

(a) Confusion Matrix

| Output class | Target class |                  |                  |
|--------------|--------------|------------------|------------------|
|              | Malignant    | Benign           |                  |
| Malignant    | 78<br>TP     | 37<br>FN         | 67.82%<br>32.18% |
| Benign       | 31<br>FP     | 176<br>TN        | 85.02%<br>14.98% |
|              |              | 71.55%<br>28.45% | 82.62%<br>17.38% |
|              |              |                  | 78.88%<br>21.12% |

(b) mini-MIAS Dataset

| Output class | Target class |                  |                  |
|--------------|--------------|------------------|------------------|
|              | Malignant    | Benign           |                  |
| Malignant    | 49<br>TP     | 24<br>FN         | 67.12%<br>32.88% |
| Benign       | 14<br>FP     | 113<br>TN        | 88.97%<br>11.03% |
|              |              | 77.77%<br>22.23% | 82.48%<br>17.52% |
|              |              |                  | 81%<br>19%       |

(c) INbreast Dataset

| Output class | Target class |                  |                  |
|--------------|--------------|------------------|------------------|
|              | Malignant    | Benign           |                  |
| Malignant    | 40<br>TP     | 10<br>FN         | 80%<br>20%       |
| Benign       | 15<br>FP     | 35<br>TN         | 70%<br>30%       |
|              |              | 72.72%<br>27.28% | 77.77%<br>22.23% |
|              |              |                  | 75%<br>25%       |

(d) BCDR Dataset

**Figure 18: Confusion Matrix of the Proposed Method based on Neural Network.**

### B. MANUAL VERSUS MEASUREMENT

Both the length and width of the ROI based upon the domain expert the manual versus measurements were assessed. The correlation between the measurements of the diameters of the ROI produced and the measurements generated by the expert. The correlation among both measurements' manual and automatically based on linear regression for correlation by using the R2 is shown in Equation 8.

$$R^2 = \left( \frac{N(\sum xy) - (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum x)^2][N \sum y^2 - (\sum y)^2]}} \right)^2 \quad (10)$$

The angle of Regression Line (ARL) was checked in the first evaluation. The scatter plot of Figure 19 (a and b) shows the width and length of ROI for the automatic manual versus assessments. The close correlation among both manual and automatic assessment in many cases are shown.

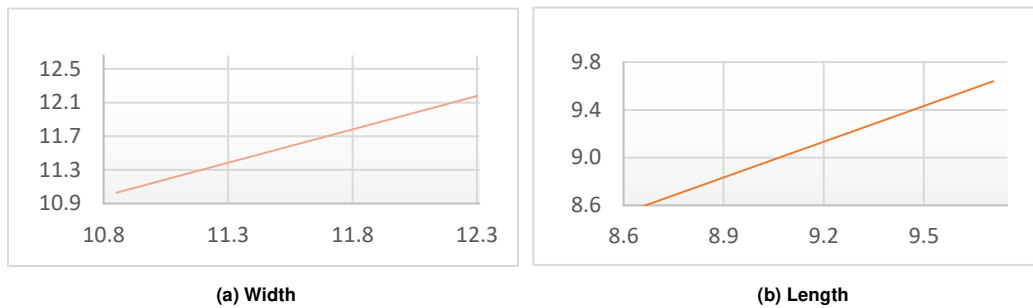


Figure 19: The Correlation between Manual and Automatic Versus assessments.

In contrast, after evaluating the correlation among both assessments manual and automatic, their closeness was tested by utilizing the Bland Altman (BA) evaluation method. The method of scatter plot, Bland and Altman were invented this method, characterizes both of agreement and disagreement among two parameters measured quantitatively. BA is utilized to compute the agreement amount by constructing the limits of agreement among measurements. Calculating the statistical limits can be done by two quantitative measurements which are mean and standard deviation, the variations in both of them are considered a major method of calculating. The scatter plot method consists of two axes, where the variation among two

paired measurements (A-B) is represented by Y-axis whereas the average of these measures ((A+B)/2) illustrates by the X-axis. The variation in the paired assessments was plotted against the mean of the two quantitative assessments. The technique of BA shows that two assessments are close when 95% of the data points locate within  $\pm 2$  of the variation of the mean [44]. This property is based upon the normal distribution theory. Moreover, the results confirm the conclusion, that is, the width and length measurements of the proposed model increases the accuracy. The final estimation for both width in (A) and length in (B) is shown in the BA analysis in Figure 20.

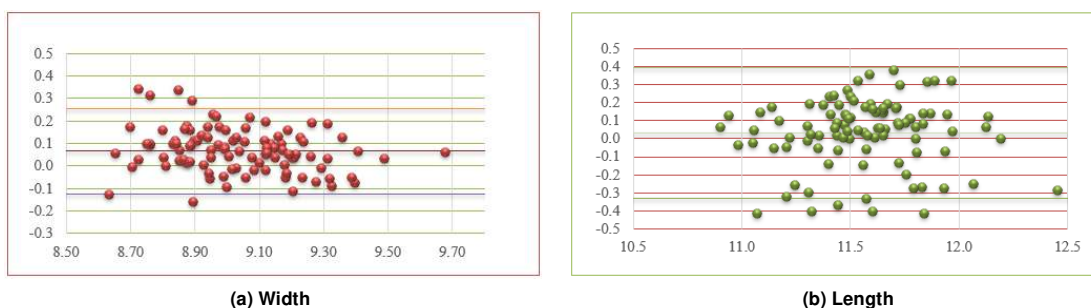


Figure 20: The Analysis of Bland Altman for Manual and Automatic Versus.

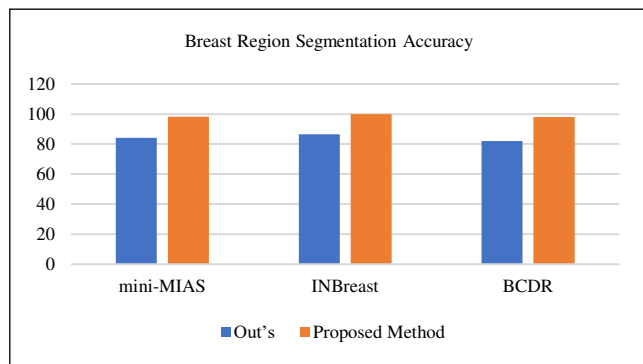
### C. BREAST BOUNDARY SEGMENTATION RESULTS

The percentage of the successful segmentation of the ROI was calculated manually by using the Human Visual System (HVS). The proposed system involves two stages, including the binarization of the image to obtain a binary image with true positive and then filtering out non-ROI (false positive). The first stage is employed to identify the TP and minimize the number of FP. The second step aims at selecting the right

ROI (TP). This section involves the evaluation of both stages.

In the binarization stage, the Out's threshold-based framework applied on 322 images of mini-MIAS datasets, this method successfully obtained the TP objects beside the FP with 271 out of 322 images and 51 images unsuccessfully. Thus, this method achieved a TP rate of 84.16%. However, by using the new threshold model, the rate increased to 98.13%, where 316 of 322 images

succeeded, where six did not succeed (Figure 21). For INbreast dataset, this study used 200 MLO mammogram images, based on Out's threshold 173 TP objects were obtained successfully out of 200 images. Thus, this technique achieved 86.5% accuracy. In contrast, by applying the new threshold method 200 TP objects have been obtained successfully out of 200. Therefore, the accuracy has been increased to 100%. Finally, 100 MLO mammogram images have been used to evaluate the performance of the proposed method. Based on Out's threshold, the BCDR dataset obtained 82% accuracy. The proposed method, for BCDR images TP objects beside FP with 98 images were successfully obtained whereas only 2 FP objects obtained. Thus, the accuracy rate has increased to 98%.



**Figure 21: The Accuracy of the Proposed Threshold Technique.**

Table 2 shows the results for the mini-MIAS, INbreast, and BCDR databases, which indicated that the proposed method yields very good accuracy for breast region segmentation. The best results achieved when we segment region of breast from background including artifacts in the INbreast database with accuracy 100%, sensitivity 100%, specificity 100%, jaccard 98.61%, and dice 98.7%. Similarly, for BCDR database we obtain specificity 100% as well. The proposed segmentation method obtained 99.8% accuracy, 99.71% sensitivity, jaccard 99.31%, and dice 99.1% for the BCDR database. For mini-MIAS, we achieve accuracy 98.13%, sensitivity 98.9%, specificity 98.8%, jaccard 98.01%, and dice 99.01%. On average to summarize the results our method obtained accuracy 99.31%, sensitivity 99.54%, specificity 99.41%, jaccard 98.67%, and dice 99.14% through different databases. The line of breast boundary through background in INbreast dataset was visible and clear compared to other databases. Due to this reason, the proposed method obtained higher results. Overall, the evaluation results indicated that the proposed threshold model is robust for the region of breast segmentation from the background. More so, it has been shown that the proposed method has the ability to generalize across different datasets.

To ensure the effectiveness of the proposed method, we used traditional techniques for comparison with the proposed trainable ROI segmentation method in the MLO view of mammogram images. This study consists of two main stages

to propose a fully automatic segmentation. A new threshold technique was proposed followed by morphological operations to segment the breast region from the background. This stage of the proposed segmentation method obtained 98.13%, 100%, and 98% accuracy for mini-MIAS, INbreast, and BCDR, respectively. Otsu's thresholding technique with some of the previous studies was used for comparison. The comparison of the study directly is difficult because of the variation in such issues such as the number of employed images and evaluation methods. In contrast, our results have been summarized in Table 4 region of breast segmentation and Table 4 pectoral muscle segmentation by presenting some of the previous studies in the literature. Studies that have used the datasets which are used by our study have been covered here. More so, in the literature, there are many developed studies that used qualitative evaluation by an expert. Furthermore, some of the studies in the literature used some private datasets to evaluate their study. Obviously, it has been shown in Table 4 the proposed threshold technique outperformed the previous studies. Our method obtained higher results than previous methods across all used datasets which are mini-MIAS- INbreast, and BCDR. This due to cause the previous studies focused on such information which may suffer from over or under segmentation problem whereas our method used three different texture (mean, median, and entropy) which can overcome the over and under segmentation problem. Thus, this made our proposed method more robust and flexible. The segmentation accuracy between the proposed technique and Otsu's thresholding and the recent techniques are illustrated in Table 2. Otsu's thresholding was evaluated on 322 mammograms before evaluating the proposed model. The proposed method for breast region segmentation from background outperformed the methods in recent studies. As has been illustrated in Table 2, the high region of breast segmentation accuracy obtained indexed in INbreast, mini-MIAS, BCDR databases, respectively. Higher sensitivity and specificity were achieved in INbreast, BCDR, and mini-MIAS databases, respectively. Finally, best jaccard and dice were obtained indexes in BCDR, INbreast, and mini-MIAS databases, respectively.

**Table 2.** ACCURACY (%) COMPARISON OF THE SEGMENTATION OF BEAST REGION FROM BACKGROUND

| Algorithms          | Segmentation approach                                 | Dataset   | Total no. of images | Acc          | Sen          | Spe          | Jac          | Dice         |
|---------------------|---|-----------|---------------------|--------------|--------------|--------------|--------------|--------------|
| Otus's thresholding | Thresholding  | mini-MIAS | 322                 | 84.1         | ---          | ---          | ---          | ---          |
| Otus's thresholding | Thresholding  | INbreast  | 200                 | 86.5         | ---          | ---          | ---          | ---          |
| Otus's thresholding | Thresholding  | BCDR      | 100                 | 82           | ---          | ---          | ---          | ---          |
| [16]                | Thresholding  | mini-MIAS | 300                 | 98           | ---          | ---          | ---          | ---          |
| [17]                | Thresholding  | mini-MIAS | 320                 | 98           | ---          | ---          | ---          | ---          |
| [26]                | Active counter  | mini-MIAS | 65                  | 97           | ---          | ---          | ---          | ---          |
| [45]                | Active counter  | mini-MIAS | 25                  | 97           | ---          | ---          | ---          | ---          |
| [46]                | Active counter  | mini-MIAS | 84                  | 96           | ---          | ---          | ---          | ---          |
| [47]                | Thresholding  | mini-MIAS | 322                 | <b>98.4</b>  | 98.7         | 97.6         | 97.6         | 98.8         |
|                     | and active  | INbreast  | 208                 | 99.2         | 99.8         | 98.7         | 97.7         | <b>98.9</b>  |
|                     | counter   | BCDR      | 100                 | <b>99.9</b>  | 99.2         | 99.2         | <b>98.4</b>  | 99.2         |
| [54]                | Contrast Enhancement and Intensity based Thresholding | mini-MIAS | 322                 | 92.55        | ---          | ---          | ---          | ---          |
| Ours                | Thresholding  | mini-MIAS | 322                 | 98.13        | <b>98.9</b>  | <b>98.4</b>  | <b>98.12</b> | <b>99.01</b> |
|                     |   | INbreast  | 200                 | <b>100</b>   | <b>100</b>   | <b>100</b>   | <b>98.61</b> | 98.7         |
|                     |   | BCDR      | 100                 | 99.8         | <b>99.73</b> | <b>99.84</b> | 99.31        | <b>99.71</b> |
| <b>Overall</b>      |   |           | 622                 | <b>99.31</b> | <b>99.54</b> | <b>99.41</b> | <b>98.67</b> | <b>99.14</b> |

#### D. PECTORAL MUSCLE SEGMENTATION RESULTS

In this stage, the binary image that includes the TP and FP objects was obtained. The FP object should be removed to obtain specific information from the TP objects and obtain an automatic solution. After applying the proposed threshold method, the output obtained includes the ROI connected with another object (unwanted-ROI or FP object). Therefore, this study proposed a method to obtain only the segmented object (both TP and FP) and mask it with the original image to obtain the original grey value (original image information). The proposed method requires samples for the training stage from both the border of the ROI and the FP object. This process will help in obtaining a powerful model that can recognize the ROI border.

To show the strength of using HOG features for the pectoral muscle segmentation system using mammogram images, the performance of the proposed segmentation method was evaluated initially based on the HOG feature, scale-invariant feature transform (SIFT), and speeded up robust features (SURF). From mini-MIAS, INBreast, and BCDR databases 50 images were taken randomly to test our proposed segmentation method. Four different metrics have been used to calculate quantitative performance measures. These metrics are extensively utilized in the previous works to test the performance of the segmentation methods which are Structural Similarity Index (SSIM), Probabilistic Rand Index (PRI), Variation of Information (VoI), and Global Consistency Error (GCE). SSIM is considered a quality evaluation algorithm used for indexing the similarity among the segmented and ground-truth images. Different components can be compared using SSIM which are

contrast, luminance, and structure among the image X (segmented) and image Y (ground-truth) using a local window, SSIM can be performed using Equation (11). However, the number of portions of pair pixels among different images whose labels are harmonious from segmented and ground-truth can be calculated using PRI. Based on averaging via a set of images of ground-truth to compute for scale differences in human perception. The range of PRI value is among zero and one, where the more similarity indicated by the higher value of PRI. More so, the distance among two segmentations of manual and automatic can be measured using a nonnegative metric named VOI. This measurement can be done based on the information variation between both manual and automatic segmentation. VOI is able to compute the distance among two clusters depending on the mutual information and entropy. The VOI can be performed using Equation (12), where the lower value of VOI indicates to greater similarity. The GCE assesses the range to the image which has been segmented is able to view as a refinement of images from the ground-truth. The segmentation process is consistent when a segment is a group of pixels and each pixel is in a region of refinement when the segment (S) is a useful subset of segment (S'). In this condition, the local error is equal to 0 whereas when there is no relationship among the two segments are overlapped in an inconsistent method. Equation (13) is performed to calculate the local refinement error among two segments which are segmented image (S) and ground-truth image (S'). The GEC range is between zero and one, a lower value of GEC is considered better [48].



The results acquired from the breast cancer segmentation using mammogram images shown in Table 3 and Figure 22. HOG, SIFT, and SURF features were used in the evaluation using mammogram images from different databases. The overall average of each feature using four quantitative metrics is performed using the same samples of mini-MIAS, INbreast, and BCDR databases. Based on the results which have been obtained it has been investigated that the HOG feature outperformed SIFT and SURF features. Thus, to build a robust and effective ML system mammogram breast cancer segmentation HOG feature has been exploited.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (11)$$

Where  $\mu_x$  and  $\mu_y$  are considered the mean intensities of  $x$  and  $y$ , the standard deviations of  $x$  and  $y$  has been represented by  $\sigma_x^2 + \sigma_y^2$ .  $\sigma_{xy}$  indicates the measure of covariance of  $x$  and  $y$ .  $C_1 = (K_1L)^2$ ,  $C_2 = (K_2L)^2$  are small constants utilized to preserve stability when either  $\mu_x^2 + \mu_y^2$  and  $\sigma_x^2 + \sigma_y^2$  are very near to 0. The dynamic range of the pixels of values represented by  $L$ , and  $K_1, K_2 < 1$ .

$$VOI(S, S') = H(S) + H(S') - 2I(S, S') \quad (12)$$

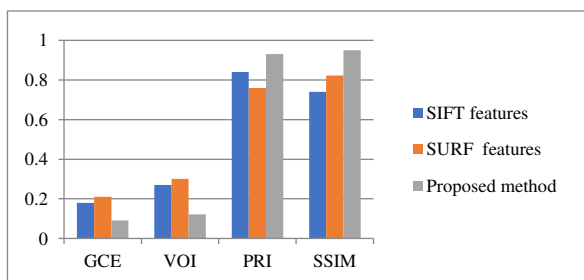
Here  $S$  is segmented image and  $S'$  represents ground-truth image. The range of VOI is among 0 and  $\infty$ .  $H$  indicates the entropy and manual information represented by  $I$ .

$$GCE(S, S') = \frac{1}{n} \min \{ \sum_i E(S, S', p_i) \sum_i E(S', S, p_i) \} \quad (13)$$

Where  $S$  and  $S'$  represent two segments, given pixel  $p_i$  represents the segments which contain  $p_i$  in two segments  $S$  and  $S'$ .

**TABLE 3.** THE EVALUATION OF THE PROPOSED SEGMENTATION MODEL USING DIFFERENT FEATURES

| Features        | GCE  | VOI   | PRI  | SSIM  |
|-----------------|------|-------|------|-------|
| SIFT            | 0.18 | 0.27  | 0.84 | 0.74  |
| SURF            | 0.21 | 0.3   | 0.76 | 0.823 |
| Proposed method | 0.09 | 0.121 | 0.93 | 0.95  |



**Figure 22: The Comparative Analysis of the Proposed Segmentation Model Using Different Features.**

Moreover, in the previous stage, the proposed system has succeeded with only 316, 200, and 98 images from mini-MIAS, INbreast, and BCDR databases. To detect ROI, the second phase of the proposed method has been implemented using mini-MIAS, INbreast, and BCDR databases. For mini-MIAS by filtering out non-ROI to avoid the FP, out of 316 images, 311 TP were accurately detected. Thus, the use of the trainable model achieved accuracy 98.41%, sensitivity 98.7%, specificity 99.7%, jaccard 96.3%, and dice 98.8% segmentation rate of ROI. Regarding INbreast dataset applied on the 200 images, the proposed study produces values of accuracy 98.01%, sensitivity 97.03%, specificity 99.2%, jaccard 94.01%, and dice 97.8%. Besides, the proposed study achieved accuracy 99.5%, sensitivity 99.02%, specificity 100%, jaccard 97.7%, and dice 98.9% for BCDR dataset. On average to summarize the results our method obtained accuracy 98.64%, sensitivity 98.25%, specificity 99.63%, jaccard 96%, and dice 98.5% through different databases. Pectoral muscle segmentation was more difficult and more complex in INbreast dataset due to the texture similarity between the pectoral muscle ROI which causes to obtain lower results compared to other datasets. On the other hand, the segmentation of pectoral muscle in BCDR is less complex and more visible which leads to outperformed compared INbreast and mini-MIAS datasets.

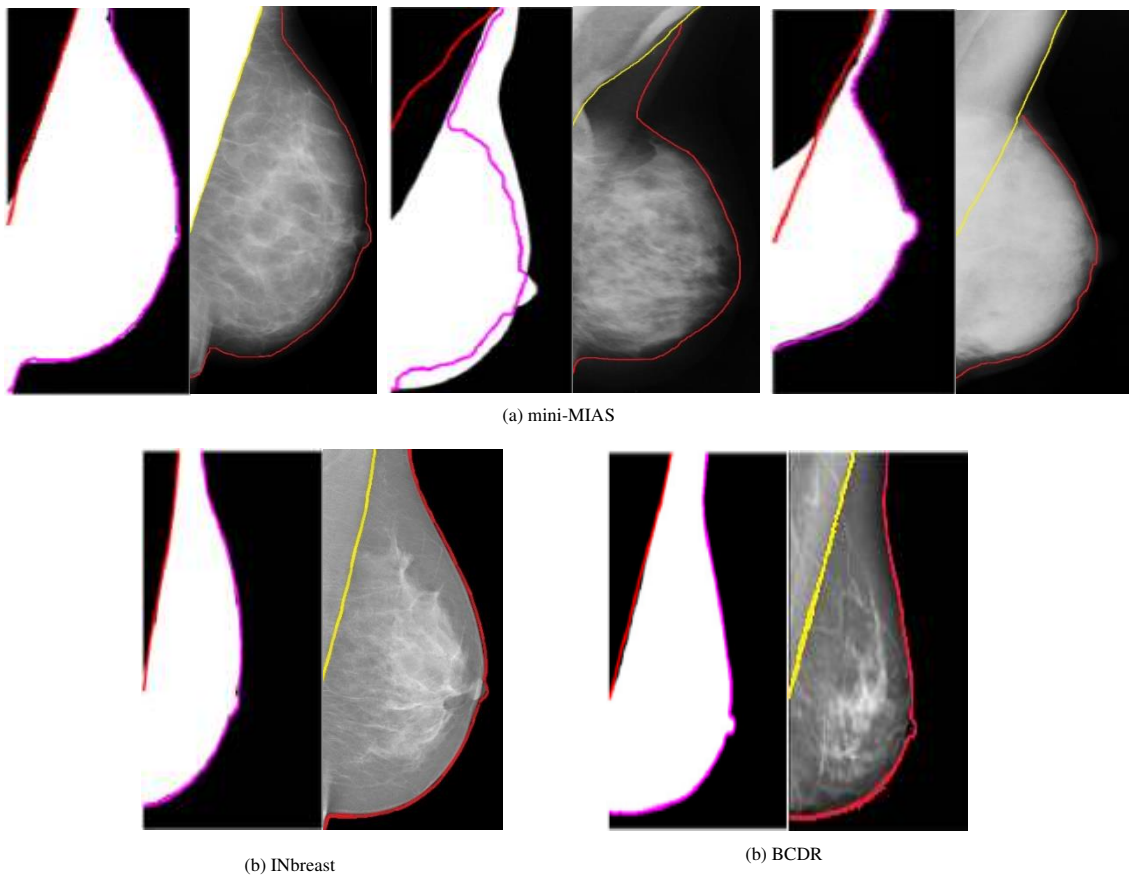
Furthermore, a trainable segmentation method was proposed to segment ROI from unwanted ROI (pectoral muscle). To show the effectiveness of the proposed method, we used region-growing, thresholding, and K-means clustering traditional techniques for comparison with the proposed trainable ROI segmentation. The segmentation accuracy between the proposed, traditional, and recent techniques is illustrated in Table 4. The results of region growing, thresholding, and K-means clustering were obtained from [30]. Three recent techniques were also used to verify whether the proposed technique is better than recent techniques. An overall comparison with three traditional techniques and three recent previous techniques was carried out, and the proposed trainable segmentation technique achieved the highest segmentation accuracy. Among all previous techniques mentioned above, two techniques used the same number of datasets as the proposed technique, whereas four techniques used a small number of mammogram images. Based on the comparison, the proposed technique is an efficient technique for ROI segmentation in the MLO view of mammograms.

**Table 4.** ACCURACY (%) COMPARISON OF THE SEGMENTATION OF ROI FROM UN-WANTED ROI (PECTORAL MUSCLE)

| Algorithms         | Segmentation approach       | Dataset   | Total no. of images | Acc          | Sen          | Spe          | Jac          | Dice        |
|--------------------|-----------------------------|-----------|---------------------|--------------|--------------|--------------|--------------|-------------|
| Region growing     | Region growing              | mini-MIAS | 318                 | 85.53        | ---          | ---          | ---          | ---         |
| Thresholding       | Thresholding                | mini-MIAS | 318                 | 72.64        | ---          | ---          | ---          | ---         |
| K-means clustering | K-means clustering          | mini-MIAS | 318                 | 66.67        | ---          | ---          | ---          | ---         |
| [28]               | Thresholding                | mini-MIAS | 320                 | 87.19        | ---          | ---          | ---          | ---         |
| [29]               | Region intensity            | mini-MIAS | 322                 | 91.30        | ---          | ---          | ---          | ---         |
| [30]               | Machine learning            | mini-MIAS | 322                 | 93.71        | ---          | ---          | ---          | ---         |
| [21]               | Region growing              | mini-MIAS | 322                 | 92.8         | ---          | ---          | ---          | ---         |
| [44]               | Region growing              | mini-MIAS | 322                 | 95.7         | ---          | ---          | ---          | ---         |
| [47]               | Canny edge detection        | mini-MIAS | 322                 | 98.1         | 89.6         | 98.9         | 92.1         | 97.8        |
|                    |                             | INbreast  | 208                 | 99.1         | 89.6         | 99.7         | 84.6         | 89.6        |
|                    |                             | BCDR      | 100                 | 99.4         | 92.7         | 99.9         | 85.8         | 91.9        |
| [27]               | Pixel-wise clustering       | mini-MIAS | 322                 | 97.08        | ---          | ---          | 94.89        | 96.4        |
|                    |                             | INbreast  | 201                 | 97.38        | ---          | ---          | <b>95.96</b> | 97.6        |
|                    |                             | BCDR      | 100                 | 97.91        | ---          | ---          | 96.22        | 97.66       |
| [49]               | Deep learning               | mini-MIAS | 322                 | <b>99.3</b>  | 98.2         | 99.5         | 94.6         | 97.5        |
|                    |                             | INbreast  | All                 | <b>99.6</b>  | 95.2         | 99.8         | 92.6         | 95.6        |
|                    |                             | BCDR      | 100                 | <b>99.9</b>  | <b>99.6</b>  | 99.6         | 96.9         | <b>98.8</b> |
| [51]               | Region growing              | MIAS      | ---                 | <b>94.4</b>  | <b>89.62</b> | 99.99        | ---          | ---         |
| [52]               | Geometrical properties      | INbreast  | 60                  | ---          | <b>95.6</b>  | ---          | ---          | ---         |
| [53]               | Fully convolutional network | MIAS      | 644                 | 96           | ---          | ---          | ---          | 94.5        |
|                    |                             | INBreast  | 416                 | 95           | ---          | ---          | ---          | 94          |
| Ours               | Machine learning            | mini-MIAS | 316                 | 98.41        | <b>98.7</b>  | <b>99.7</b>  | <b>96.3</b>  | <b>98.8</b> |
|                    |                             | INbreast  | 200                 | 98.01        | <b>97.03</b> | <b>99.2</b>  | 94.01        | <b>97.8</b> |
|                    |                             | BCDR      | 98                  | 99.5         | 99.02        | <b>100</b>   | <b>97.7</b>  | 98.9        |
| <b>Overall</b>     |                             |           | 614                 | <b>98.64</b> | <b>98.25</b> | <b>99.63</b> | <b>96</b>    | <b>98.5</b> |

Figure 23 shows the segmentation example for mammogram images in the databases that have been used based on the proposed study and ground truth. Regarding ground truth for mammogram images in the mini-MIAS database have manual segmentation. This study focused on the ground truth for breast boundary and pectoral muscle estimation that has been provided by [50]. A clinician annotated the estimation of both of them under the supervision of an expert radiologist. Figure 23 (a) illustrates the segmentation example from mini-MIAS images. The breast boundary and pectoral muscle for proposed segmentation have been shown in red and yellow lines, respectively. However, the breast boundary and pectoral muscle segmentation have been shown in magenta and red lines, respectively. The masked images are used as ground truth segmentation whereas the segmentation of the proposed study used grayscale images. The first pair of Figure 23 (a) is considered as a good segmentation result which obtained 97.01% jaccard and 99.3% dice. The second of Figure 23 (a) segmented the boundary of the breast incorrectly due to the over and under segmentation problem, respectively. The last pair of Figure 23 (a) segmented pectoral muscle incorrectly due to homogeneity between the texture of the pectoral muscle and region of the breast. According to the mammogram images in the INbreast database, the pectoral

muscle has been annotated by an expert radiologist whereas the boundary of the breast has been annotated by one of the authors [34]. On the other hand, Figure 23 (b) shows example results for the ground truth and proposed study utilizing the INbreast database. Based on the results that have been obtained showed that the proposed study is robust in breast boundary segmentation for INbreast datasets. However, the proposed study achieved average results because of the big average texture homogeneity between region of breast and region of the pectoral muscle. However, the pectoral muscle and the boundary of the breast have been annotated by an author for mammogram images in the BCDR database [36]. On the other hand, Figure 23 (c) shows example results for the ground truth and proposed study utilizing the BCDR database. Based on the results that have been obtained showed that the proposed study is robust in both breast boundary and pectoral muscle segmentation. More so, the proposed study outperformed on pectoral muscle segmentation compared to mini-MIAS and INbreast databases. However, the proposed study achieved lower results than INbreast for the boundary of breast segmentation. As a result, it has been investigated that the pectoral muscle segmentation is always considered a difficult task for all databases compared to breast boundary segmentation based on the achieved Jaccard and dice results.



**Figure 23: Example of Segmentation Results for Different Samples: (a) mini-MIAS Database (b) INbreast Database (c) BCDR Database**

In the segmentation stage, a trainable model was proposed to extract ROI from the original image. Mammogram images suffer from noise and it has low quality, and those two points affect the image segmentation task. Our proposed study has reduced the effect of those two limitations by enhancing mammograms before segmentation. However, in some cases, the ROI has very poor or no border. Therefore, our proposed model cannot obtain the missing border. Furthermore, a trainable model was proposed to extract the ROI from the pectoral muscle. Mammogram images suffer from the similarity between the texture of ROI and the texture of pectoral muscle, irregular ROI, inhomogeneous ROI, and the missing border of ROI. These factors affect the segmentation process, thereby confusing the object detection model in identifying ROI.

## VII. CONCLUSION

Segmentation helps in eliminating or segmenting the wanted area of the image from the unwanted for further processing. Breast region extraction is useful because of the search-zone limitation of the abnormalities of the breast. Furthermore, pectoral muscles always appear in the MLO

view of mammogram images. Identification and segmentation of this region are crucial considering the overlap between pectoral muscles and ROI. This study mainly aims to design and develop a fully automated segmentation technique that can detect the breast region and eliminate pectoral muscles from mammogram images. Thus, we have proposed an efficient model based on a new thresholding technique and machine learning system. Firstly, we proposed an enhancement method based on wavelet transform to detect breast boundary in mammogram images. Secondly, ROI and unwanted ROI (pectoral muscle) have been segmented from background and artefacts through proposing a new threshold technique. Finally, a machine learning system has built to segmented ROI from unwanted ROI (pectoral muscle) for discriminating between benign or malignant. Research related to the segmentation of ROI from the MLO view of mammogram images is limited. This study highlights that mammogram segmentation is still an open research problem. The proposed solution can overcome various segmentation challenges of the MLO view of mammogram images. Moreover, the proposed segmentation method is effective and outperforms previous methods.

## REFERENCES

- [1] D. N. Ponraj, M. E. Jenifer, P. Poongodi, and J. S. Manoharan, "A survey on the preprocessing techniques of mammogram for the detection of breast cancer," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 2, no. 12, pp. 656-664, 2011.
- [2] M. A. Mohammed, B. Al-Khateeb, A. N. Rashid, D. A. Ibrahim, M. K. A. Ghani, and S. A. Mostafa, "Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images," *Computers & Electrical Engineering*, vol. 70, pp. 871-882, 2018.
- [3] O. I. Obaid, M. A. Mohammed, M. K. A. Ghani, A. Mostafa, and F. Taha, "Evaluating the performance of machine learning techniques in the classification of Wisconsin Breast Cancer," *International Journal of Engineering & Technology*, vol. 7, no. 4.36, pp. 160-166, 2018.
- [4] L. Panigrahi, K. Verma, and B. K. Singh, "Ultrasound image segmentation using a novel multi-scale Gaussian kernel fuzzy clustering and multi-scale vector field convolution," *Expert Systems with Applications*, vol. 115, pp. 486-498, 2019.
- [5] V. P. Singh, S. Srivastava, and R. Srivastava, "Automated and effective content-based image retrieval for digital mammography," *Journal of X-ray Science and Technology*, vol. 26, no. 1, pp. 29-49, 2018.
- [6] D. Q. Zeebaree, H. Haron, and A. M. Abdulazeez, "Gene selection and classification of microarray data using convolutional neural network," in *2018 International Conference on Advanced Science and Engineering (ICOASE)*, 2018, pp. 145-150: IEEE.
- [7] J. Virmani and S. Thakur, "Classification of Breast Tissue Density Patterns Using SVM-Based Hierarchical Classifier," in *Software Engineering*: Springer, 2019, pp. 185-191.
- [8] F. Gulzar, M. S. Akhtar, R. Sadiq, S. Bashir, S. Jamil, and S. M. Baig, "Identifying the reasons for delayed presentation of Pakistani breast cancer patients at a tertiary care hospital," *Cancer management and research*, vol. 11, p. 1087, 2019.
- [9] M. Mustra, M. Grgic, and R. M. Rangayyan, "Review of recent advances in segmentation of the breast boundary and the pectoral muscle in mammograms," *Medical & biological engineering & computing*, vol. 54, no. 7, pp. 1003-1024, 2016.
- [10] E. E. Sterns, "Relation between clinical and mammographic diagnosis of breast problems and the cancer/biopsy rate," *Canadian journal of surgery*, vol. 39, no. 2, p. 128, 1996.
- [11] R. Highnam and M. Brady, "Mammographic image analysis kluwer academic," ed: Dordrecht, 1999.
- [12] M. Mustra and M. Grgic, "Robust automatic breast and pectoral muscle segmentation from scanned mammograms," *Signal processing*, vol. 93, no. 10, pp. 2817-2827, 2013.
- [13] D. Q. Zeebaree, H. Haron, A. M. Abdulazeez, and D. A. Zebari, "Machine learning and Region Growing for Breast Cancer Segmentation," in *2019 International Conference on Advanced Science and Engineering (ICOASE)*, 2019, pp. 88-93: IEEE.
- [14] P. Filipczuk, M. Kowal, and A. Obuchowicz, "Automatic breast cancer diagnosis based on k-means clustering and adaptive thresholding hybrid segmentation," in *Image processing and communications challenges 3*: Springer, 2011, pp. 295-302.
- [15] I. A. Lbakhir, R. Es-Salhi, I. Daoudi, and S. Tallal, "A new mammogram preprocessing method for Computer-Aided Diagnosis systems," in *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*, 2017, pp. 166-171: IEEE.
- [16] K. Czaplicka and J. Włodarczyk, "Automatic breast-line and pectoral muscle segmentation," *Schedae Informaticae*, vol. 20, 2011.
- [17] D. Raba, A. Oliver, J. Martí, M. Peracaula, and J. Espunya, "Breast segmentation with pectoral muscle suppression on digital mammograms," in *Iberian conference on pattern recognition and image analysis*, 2005, pp. 471-478: Springer.
- [18] M. Masek, Y. Attikiouzel, and C. deSilva, "Skin-air interface extraction from mammograms using an automatic local thresholding algorithm," *ICB, Brno, CR*, pp. 204-206, 2000.
- [19] Z. Chen and R. Zwiggelaar, "A combined method for automatic identification of the breast boundary in mammograms," in *2012 5th International Conference on BioMedical Engineering and Informatics*, 2012, pp. 121-125: IEEE.
- [20] C. Zhou *et al.*, "Review of recent advances in segmentation of the breast boundary and the pectoral muscle in mammograms," *Med Phys*, vol. 28, no. 6, pp. 1056-1069, 2001.
- [21] D. J. Williams and M. Shah, "A fast algorithm for active contours and curvature estimation," *CVGIP: Image understanding*, vol. 55, no. 1, pp. 14-26, 1992.
- [22] S. Lobregt and M. A. Viergever, "A discrete dynamic contour model," *IEEE transactions on medical imaging*, vol. 14, no. 1, pp. 12-24, 1995.
- [23] P. Casti *et al.*, "Estimation of the breast skin-line in mammograms using multidirectional Gabor filters," *Computers in biology and medicine*, vol. 43, no. 11, pp. 1870-1881, 2013.
- [24] J. Chakraborty, S. Mukhopadhyay, V. Singla, N. Khandelwal, and P. Bhattacharyya, "Automatic detection of pectoral muscle using average gradient and shape based feature," *Journal of digital imaging*, vol. 25, no. 3, pp. 387-399, 2012.
- [25] N. Karssemeijer, "Automated classification of parenchymal patterns in mammograms," *Physics in medicine & biology*, vol. 43, no. 2, p. 365, 1998.
- [26] R. Martí, A. Oliver, D. Raba, and J. Freixenet, "Breast skin-line segmentation using contour growing," in *Iberian conference on pattern recognition and image analysis*, 2007, pp. 564-571: Springer.
- [27] P. Shi, J. Zhong, A. Rampun, and H. Wang, "A hierarchical pipeline for breast boundary segmentation and calcification detection in mammograms," *Computers in biology and medicine*, vol. 96, pp. 178-188, 2018.
- [28] P. Vikhe and V. Thool, "Intensity based automatic boundary identification of pectoral muscle in mammograms," *Procedia Computer Science*, vol. 79, pp. 262-269, 2016.
- [29] A. Shrivastava, A. Chaudhary, D. Kulshreshtha, V. P. Singh, and R. Srivastava, "Automated digital mammogram segmentation using dispersed region growing and sliding window algorithm," in *2017 2nd international conference on image, vision and computing (ICIVC)*, 2017, pp. 366-370: IEEE.
- [30] V. Shinde and B. T. Rao, "Novel approach to segment the pectoral muscle in the mammograms," in *Cognitive Informatics and Soft Computing*: Springer, 2019, pp. 227-237.
- [31] S. K. Suguna, R. Ranganathan, J. Sangeetha, S. Shandilya, and S. K. Shandilya, "Application of Nature-Inspired Algorithms in Medical Image Processing," in *Advances in Nature-Inspired Computing and Applications*: Springer, 2019, pp. 61-100.
- [32] M. A. Al-antari and T.-S. Kim, "Evaluation of Deep Learning Detection and Classification towards Computer-aided Diagnosis of Breast Lesions in Digital X-ray Mammograms," *Computer Methods and Programs in Biomedicine*, p. 105584, 2020.
- [33] P. SUCKLING J, "The mammographic image analysis society digital mammogram database," *Digital Mammo*, pp. 375-386, 1994.
- [34] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast: toward a full-field digital mammographic database," *Academic radiology*, vol. 19, no. 2, pp. 236-248, 2012.
- [35] M. G. Lopez *et al.*, "BCDR: a breast cancer digital repository," in *15th International conference on experimental mechanics*, 2012, vol. 1215.
- [36] R. Ramani, N. S. Vanitha, and S. Valarmathy, "The pre-processing techniques for breast cancer detection in mammography images," *International Journal of Image, Graphics and Signal Processing*, vol. 5, no. 5, p. 47, 2013.
- [37] G. Kom, A. Tiedeu, and M. Kom, "Automated detection of masses in mammograms by local adaptive thresholding," *Computers in Biology and Medicine*, vol. 37, no. 1, pp. 37-48, 2007.
- [38] S.-C. Tai, Z.-S. Chen, and W.-T. Tsai, "An automatic mass detection system in mammograms based on complex texture features," *IEEE journal of biomedical and health informatics*, vol. 18, no. 2, pp. 618-627, 2013.

- [39] R. Chaieb and K. Kalti, "Feature subset selection for classification of malignant and benign breast masses in digital mammography," *Pattern Analysis and Applications*, vol. 22, no. 3, pp. 803-829, 2019.
- [40] M. Karnan and K. Thangavel, "Automatic detection of the breast border and nipple position on digital mammograms using genetic algorithm for asymmetry approach to detection of microcalcifications," *Computer methods and programs in biomedicine*, vol. 87, no. 1, pp. 12-20, 2007.
- [41] K. C. Tatikonda, C. M. Bhuma, and S. K. Samayamantula, "The Analysis of Digital Mammograms Using HOG and GLCM Features," in *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2018, pp. 1-7: IEEE.
- [42] A. Rampun, L. Zheng, P. Malcolm, B. Tiddeman, and R. Zwiggelaar, "Computer-aided detection of prostate cancer in T2-weighted MRI within the peripheral zone," *Physics in Medicine & Biology*, vol. 61, no. 13, p. 4796, 2016.
- [43] D. Q. Zeebaree, H. Haron, A. M. Abdulazeez, and D. A. Zebari, "Trainable Model Based on New Uniform LBP Feature to Identify the Risk of the Breast Cancer," in *2019 International Conference on Advanced Science and Engineering (ICOASE)*, 2019, pp. 106-111: IEEE.
- [44] I. K. Maitra, S. Nag, and S. K. Bandyopadhyay, "Technique for preprocessing of digital mammogram," *Computer methods and programs in biomedicine*, vol. 107, no. 2, pp. 175-188, 2.
- [45] M. A. Wirth and A. Stapinski, "Segmentation of the breast region in mammograms using active contours," in *Visual Communications and Image Processing 2003*, 2003, vol. 5150, pp. 1995-2006: International Society for Optics and Photonics.
- [46] R. Ferrari, A. Frere, R. Rangayyan, J. Desautels, and R. Borges, "Identification of the breast boundary in mammograms using active contour models," *Medical and Biological Engineering and Computing*, vol. 42, no. 2, pp. 201-208, 2004.
- [47] A. Rampun, P. J. Morrow, B. W. Scotney, and J. Winder, "Fully automated breast boundary and pectoral muscle segmentation in mammograms," *Artificial intelligence in medicine*, vol. 79, pp. 28-41, 2017.
- [48] S. Al-Fahdawi *et al.*, "A fully automatic nerve segmentation and morphometric parameter quantification system for early diagnosis of diabetic neuropathy in corneal images," *computer methods and programs in biomedicine*, vol. 135, pp. 151-166, 2016.
- [49] A. Rampun *et al.*, "Breast pectoral muscle segmentation in mammograms using a modified holistically-nested edge detection network," *Medical image analysis*, vol. 57, pp. 1-17, 2019.
- [50] A. Oliver, X. Lladó, A. Torrent, and J. Martí, "One-shot segmentation of breast, pectoral muscle, and background in digitised mammograms," in *2014 IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 912-916: IEEE.
- [51] ESENER, İ. I., Ergin, S., & YÜKSEL, T. (2018). A novel multistage system for the detection and removal of pectoral muscles in mammograms. *Turkish Journal of Electrical Engineering & Computer Sciences*, 26(1), 35-49.
- [52] Toz, G., & Erdogmus, P. (2018). A single sided edge marking method for detecting pectoral muscle in digital mammograms. *Engineering, Technology & Applied Science Research*, 8(1), 2367-2373.
- [53] Ali, M. J., Raza, B., Shahid, A. R., Mahmood, F., Yousuf, M. A., Dar, A. H., & Iqbal, U. (2020). Enhancing breast pectoral muscle segmentation performance by using skip connections in fully convolutional network. *International Journal of Imaging Systems and Technology*.
- [54] Suganthi, G. V., Sutha, J., Parvathy, M., & Devi, C. D. (2020). Pectoral Muscle Segmentation in Mammograms. *Biomedical and Pharmacology Journal*, 13(3).