Developing an Improved Method to Remove Pectoral Muscle for Better Diagnosis of Breast Cancer in Mammography Images

Golnoush Abaei[1], Zahra Rezaei[2], Usama Qasim Mian[3],
Yasir Azhari Abdalgadir Abdalla[4], Nitin Mathew[5], Leong Yi Gan[6]
School of Computing Technologies, RMIT University, Melbourne, Australia[1]
Department of Computer Engineering, Marvdasht Branch, Islamic Azad University, Marvdasht, Iran[2]
School of Information Technology, Monash University Malaysia, Malaysia[3, 4, 5, 6]

Abstract*—Mammography is a non-invasive method to study breast tissues for abnormalities. Computer-aided diagnosis (CAD) can automate the process of diagnosing malignant and benign tumors accurately. However, accurate results can be hampered by the presence of the pectoral muscle, which has a similar opacity to the breast tissue area. Detecting and removing pectoral muscles is not trivial due to various factors, and there are artifacts present near the pectoral muscle that can hamper proper segmentation. Given the significance of the topic, it is crucial to devise an accurate method for automatically detecting the muscle area in a mammography image and eliminating it from the rest of the image. This process of removing the pectoral muscle from the breast image can aid in precise segmentation and diagnosis of the tumor area, ultimately leading to faster diagnosis and better outcomes for patients. This study examined two segmentation algorithms, Level Set and Region Growing, for segmenting the pectoral muscle. An Improved Region Growing-based (IRG) algorithm was also proposed and showed promising results in automatically segmenting the pectoral muscle. All algorithms were tested on the MIAS dataset, and radiologists evaluated the results, showing an accuracy rating of up to 83% for IRG. The results indicated that IRG outperformed Level Set considerably due to many optimizations and modifications. IRG can be used as part of the preprocessing unit of an automated cancer diagnosis system.

Keywords*—*Breast cancer; preprocessing pectoral muscle segmentation; level set algorithm; region growing algorithm

I. INTRODUCTION

According to the World Health Organization (WHO), there were an estimated 2.3 million women diagnosed with breast cancer, and 685,000 died in 2020. Along with lung cancer, breast cancer is the most prevalent cancer worldwide, representing 12.3% of total diagnosed cancers in 2018. However, if the disease is caught early, treatments can be highly effective, with five-year survival probabilities of 90% and greater in advanced economies [1]. Unfortunately, this rate significantly drops to 66% in India and 40% in South Africa [2]. Mutebi and Anderson [1] mention that early detection is crucial for effective treatments and more so in the developing world since advanced-stage cancer treatments can be very costly and require advanced medical procedures along with a trained medical workforce to provide that treatment. These resources are not widely available in developing countries; hence, patients in these countries must get their cancers diagnosed early. To detect breast cancer early, WHO recommends yearly cancer screening for women at higher risk for developing this disease, which includes factors such as genetics, age, smoking, and drinking, among others. In addition, getting these screenings requires a trained radiologist to examine the mammography image carefully. Trained radiologists are scarce in many rural areas and often, an appointment with a specialist can be expensive for a big majority in the developing world. Furthermore, radiologists are also prone to intra/inter-observer variability errors.

One of the solutions to this problem is Computer-aided diagnosis (CAD) systems, which are highly recommended to assist radiologists in detecting breast tumors and outlining their borders. CAD systems usually use algorithms that include thresholding, region-based techniques, and edge detection techniques [3]. Mammography is an inexpensive and non-invasive method through which one can diagnose breast cancer in its early stages. As these images need interpretation by a radiologist, this may develop some problems due to fatigue, repetition, and the need for a great deal of attention to detail and other factors. Mammography can show changes in the breast up to two years before a physician can feel them. Computer-aided detection and diagnosis are considered to be one of the most promising approaches that may improve the efficiency of mammography.

In breast CAD, accurate breast segmentation is a crucial preprocessing step to speed up the subsequent processes without losing important anatomical information. However, breast and pectoral muscle segmentation is challenging, especially in scanned mammograms. This mostly happens due to the presence of some artifacts, such as duct tape and tags or it might be caused because of low contrast along the breast skin line and homogeneity between pectoral and breast tissues. Although many methods have been proposed for removing breast boundary and pectoral muscle and segmenting them, only a few have been evaluated quantitatively using all the images in the MIAS (Mammographic Image Analysis Society) database [4]. When performing mammography, the muscle area in the image can often appear similar to the tumor, making accurate segmentation difficult. To address this issue, removing the muscle area before the segmentation stage is recommended. This study presents an improved method for accurately and quickly detecting and removing the pectoral muscle area from the mammography image, thereby preparing it for segmentation. Unlike existing methods, this method considers the separation of the muscle area during segmentation, avoiding the challenges posed using the information given about the tumor area, which may be unavailable in some datasets.

This empirical study is based on two well-known segmentation methods and proposed an improved robust approach based on a set of adjustments to these methods on the MIAS database [5] to assist CAD in its preprocessing step. This paper has studied and implemented two different types of popular algorithms, namely Level Set and Region Growing. Many novel findings and implementations of these algorithms have been discussed due to the direct comparison that has been done between these two algorithms, and their findings are compared along with suggestions on what type of mammogram each algorithm performs better or worse. Moreover, an Improved Region Growing algorithm (IRG) was proposed, which introduced the concept of dynamic thresholding. The threshold values were used based on testing to find what number of iterations would make the segmentation jump out in the breast tissue. This modification improved the accuracy of the algorithm; furthermore, it also provided insights on how this method can be used and further improved using machine learning techniques to predict the best threshold value for each image. This research also includes a web-based application that is based on the IRG algorithm, which can be used by anyone to access the API (Application Programming Interface), which can be integrated into classification systems as a preprocessing step. This facility will reduce the time required for future research projects focusing on classification since they can directly call the API and get the segmented images instead of writing segmentation algorithms themselves.

This paper is organized as follows. Section II contains the literature review of the proposed method to segment the pectoral muscle regions. This is followed by Section III, which contains the methodology, including design and implementation. Section IV highlights the results and discussion and finally, the paper is concluded in Section V.

II. Literature Review

Accurately removing breast pectoral muscle is challenging for researchers due to the difference in size, shape, and position of the breast in a mammogram [6]. As indicated by [7], there are four main types of problematic images within the most popular datasets used by researchers: MIAS [5] and DDSM [8]. These include images with tape artifacts, images with axillary fold, images with the invisible contour of the pectoral muscle, the level, and images with no pectoral muscle at all. Due to these issues, many different algorithms have been proposed to segment the pectoral muscle effectively from the mammogram. These classifications are visualized in Fig. 1.

The following section presents previous works proposed by other researchers for pectoral muscle segmentation. To make it easier for readers to follow this section, papers are categorized based on whether they belong to Intensity-based, Edge-based, or Deep learning-based algorithms.

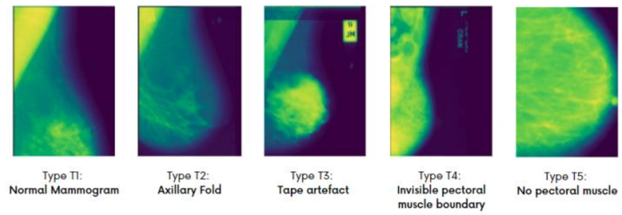


Fig. 1. Classification of the type of mammograms from the MIAS dataset

A. *Intensity-based Algorithms*

Intensity-based algorithms make use of the difference in the intensity in the mammography image. The pectoral muscle has a high intensity compared to the breast tissue. These techniques use these properties to segment the pectoral muscle out of the mammogram, assuming the pectoral muscle will have a higher intensity [9]. Such techniques include region growing, thresholding, and watershed.

Watershed algorithms are derived from mathematical morphology that segments the images into homogenous regions first introduced in 1978 [10]. Region growing uses the difference in the intensity of the pixels within an image. Region growing uses a starting seed value; from that, adjacent pixels are added depending on the homogeneity criteria assigned. Pixels are continually compared to the ones within the region and added until all adjacent pixels are too dissimilar [11].

Vikhe and Thool [12] used a thresholding-based segmentation technique using contrast enhancement and got an acceptable rate of 96.56%, verified by a certified radiologist. However, the result of this technique was affected by images with invisible contours between the pectoral muscle and the breast tissue.

Taifi et al. [13] used a watershed transformation technique to extract the pectoral muscle from the mammogram. The result was promising, with 90-99% accuracy and 86-99% for precision. There were, however, instances where this algorithm over-segmented the mammogram.

Gómez et al. [14] proposed a region-growing method with seed and threshold methods. The initial seed value used was (10,10). This algorithm’s novelty is from its method of calculating the image threshold. The result was 91.92% for the mini-MIAS dataset. As with other intensity-based methods, this method struggles with the invisible contour between the pectoral muscle and the breast tissue.

Hazarika and Mahanta [15] proposed another region-growing-based segmentation method to remove pectoral muscle in mediolateral oblique view mammograms. This method uses a triangular region to estimate the area of the pectoral muscle, after which the region-growing algorithm is applied with automatic starting seed selection. Later, the segmented region is refined to increase the accuracy of the output. This method had an acceptable rate of 86.67% based on 150 images of the mini-MIAS dataset; the accuracy rate is sufficiently accurate.

The threshold-based segmentation methods, including the region growing, have performed relatively well; however, there are a few limitations attached to this method. For example, these methods over-segment or under-segment the images with unclear pectoral muscle and breast tissue boundaries, and they do not consist of any spatial information of the image [9].

B. Edge-based Algorithms

Edge-based segmentation methods use changes in the brightness of the images to identify different regions and segment out the pectoral muscle based on the sharp differences in the brightness of the breast tissue. Straight-line modelling of the pectoral muscle boundaries is used in this method to segment the breast muscle out of the mammography image.

Rampun et al. [4] used canny edge detection and contour growing techniques to segment the image. However, this method fails when the canny method cannot detect any edge between the regions and an invalid selection of the initial contour is made. Moreover, this method also overestimates the pectoral muscle boundary when artifacts are present in the image.

Level Set is also an edge detection technique introduced in 1988 by Osher and Sethian, which represents curves or surfaces as a level set of a higher dimensional hyper-surface [16]. The level set is suited to handle problems in which the evolving interfaces can develop sharp corners and cusps and change in topological features, topology, and images with a relatively high level of noise.

Li et al. [17] proposed a new level of set-based methodology on numerous medical imaging photographs. Comparing their results to a smooth model, the researchers concluded that their proposed method outperforms the smooth model in accuracy, efficiency, and robustness.

Zhou et al. [18] used a correntropy-based level set to segment the pectoral muscle in the mammogram and concluded that this method is considerably less time-consuming and complex and gives excellent results compared to the other state-of-the-art methods.

Anitha and Peter [19] proposed a Kernel-Based Fuzzy Level Set (KFLS) to preprocess the image and to segment the image into several clusters based on the breast structure. The results based on this method showed a high percentage of sensitivity and accuracy of 93.32 and 94.31, respectively.

The line/edge detection methods have advantages when handling noisy data, as they can easily be adjusted to noise. However, it requires large storage and more computational requirements.

C. Deep Learning-based Algorithms

Due to the variability of the shape, size, and type of breast in the mammograms, there has been a lot of interest in utilizing deep learning-based algorithms to segment the pectoral muscle.

Wang et al. [20] first applied some image normalization to the input image and then trained the model using 2000 digital images with a dice-similarity coefficient of 0.8879 based on 825 of those images. Ali et al. [21] are also using this technique with Gaussian and median filters. The accuracy rate achieved was around 97%. A similar technique is applied by Kim et al. [22]. They trained their model on the 322 images of the min-MIAS dataset with an accuracy rate of 95.88% accuracy.

Rampun et al. [7] used a Convolutional Neural Network (CNN) inspired by a holistically nested edge detection network to automatically model the characteristics of the pectoral muscle.

A hybrid breast cancer classification technique was proposed by [23]. Three deep learning models, including ResNet50, Inception-V3, and AlexNet, were used for feature extraction. The Term Variance feature selection is used to select the best features. The multiclass support vector machine was applied to classify the MIAS dataset.

While deep learning models report a higher accuracy rate, they require large datasets to be trained and a very high computational power for the neural networks to get trained. A large dataset for mammograms is not readily available, so this is a hurdle for deep learning techniques at this stage.

D. Further Readings and Research Potential

Wavelet-based algorithms are also used for muscle segmentation and are not discussed in the literature review but can be helpful for some readers. This technique uses a short-term Fourier transform. The spatial frequency of the image is identified by the wavelets. Ferrari et al. [24] proposed a wavelet technique; their method was tested on 84 MLO mammograms from the min-MIAS dataset with 0.58% false positive and 5.77% false-negative percentages, which indicates a good percentage of accurate segmentations.

The Wavelet technique has an advantage in the fact that all information required to segment the image is given by the wavelet decomposition. However, this technique does result in some lost information about the image in the process. There can be further improvements in the mentioned techniques for example, the choice of the initial seed in a region growing currently does not consider the shape and size of the pectoral muscle, which can be used to increase the accuracy rate further.

III. Methodology

This section covers the methodology of the major components in the study, namely the preprocessing steps, the Level Set and Region Growing algorithms, and some details about the development of the Web Application. The overall structure of the proposed method is shown in Fig. 2. The process of removing the pectoral muscle from a mammogram is broken down into three steps – preprocessing, pectoral muscle segmentation, and performance evaluation.

A. Preprocessing

Before a mammogram is fed into a segmentation algorithm, it will be preprocessed. Details of all the preprocessing steps can be found here. The first two preprocessing steps are the removal of empty space (bar) and automatic left flipping. This needs to be done because both the Level Set and Region Growing algorithms require the pectoral muscle region to be on the top-left corner of the mammogram. Otherwise, the pectoral muscle region cannot be detected. Fig. 3 and Fig. 4 show the outcomes of the Level Set algorithm when the bar removal and automatic flipping preprocessing steps were not applied.

Another preprocessing step is the application of contrast on the mammogram. This increases the brightness differences between the pixels in the breast region and the pectoral muscle region, allowing both the segmentation algorithms to detect the pectoral muscle boundary more easily. Fig. 5 shows how the Level Set algorithm fails to detect the boundary of the pectoral muscle region (overshooting) when contrast is not applied on a mammogram.

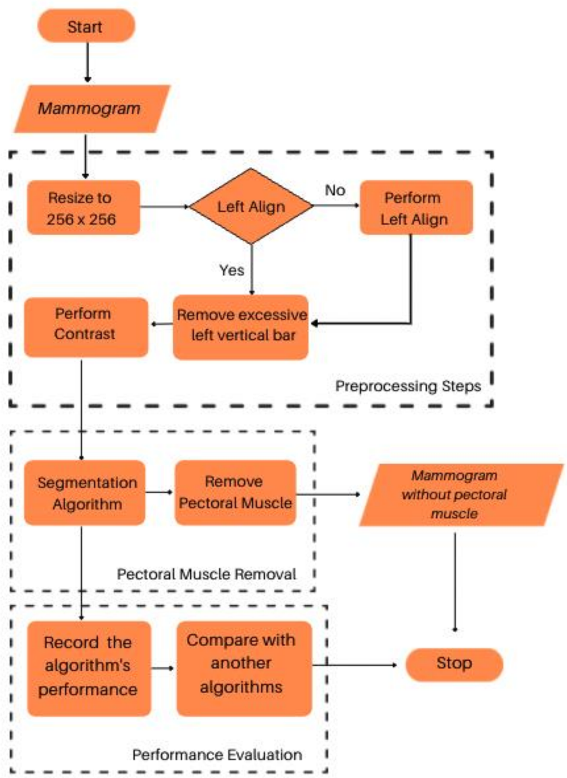


Fig. 2. The overall structure of the proposed method.

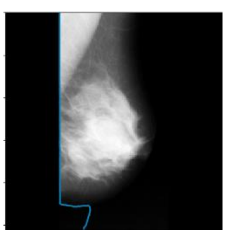


Fig. 3. Boundary found by the Level Set algorithm (denoted by the blue line) if the empty space on the left is not removed.

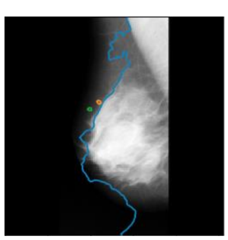


Fig. 4. Boundary found by the Level Set algorithm (denoted by the blue line) if the image is not automatically flipped to the left.

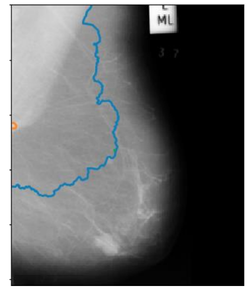


Fig. 5. Boundary found by the Level Set algorithm (denoted by the blue line) if contrast is not applied to the mammogram.

B. Segmentation

1) Level set algorithm: The Level Set image segmentation algorithm is widely used in the field of image processing and is now frequently employed in image segmentation. It is one of the active contour models that can handle complex topologies and capture boundaries and is specially used in images with intensity inhomogeneity, such as medical images [25] [26]. Researchers often prefer this method because it is flexible, easy to understand, and easy to employ. Level set models are based on the evolution of the zero level [16]. Let's assume that the closed interface $t:[0,\infty]\rightarrow Rn$ is an initial circle in 2-D/3D space when $t = 0$. In order to complete the evolution, a zero-level set function (φ) was constructed. In the general case, let 0 be a closed, disjoint, $(N - 1)$ dimensional initial hypersurface. Hence, φ can be defined by

$$xt,t=\mp r \quad (1)$$

where, r is the distance from x to the hyper-surface 0. φ is positive if x is outside of 0 and negative if it is inside. Since motion can be seen as:

$$xt,t=0 \quad (2)$$

Find the partial derivative of both sides of formula Eq. 2, by the chain rule, can get:

$$\varphi t+\Delta .xt \quad (3)$$

The speed of evolution is one of the crucial parameters; hence, it is defined as below:

$$V=xt.n \quad (4)$$

where, $n=\Delta\varphi\Delta\varphi$ is a normal vector or mean curvature, the final curve evolution equation is

$$\varphi t=V.\Delta\varphi \quad (5)$$

Level set methods can usually be divided into two main categories, which are edged-based and region-based.

C. Improved Region Growing (IRG) Algorithm

The Region-Growing Algorithm is an intensity-based segmentation algorithm. The algorithm consists of three main steps. First, a particular pixel is assigned as the initial seed, and a fixed threshold value is chosen. Second, the seed is then compared with its adjacent pixels at each iteration. If the difference between the initial seed and the adjacent pixel is within the threshold value, the adjacent pixel will merge with the initial seed to form a region. Lastly, the algorithm will terminate when all the adjacent pixels are too dissimilar.

An additional step was implemented to improve the accuracy of the original algorithm in removing the pectoral muscle region. This involves checking if similar adjacent pixels are still found after running the algorithm for a predefined maximum number of iterations. If so, the algorithm will restart with a lower threshold value, in other words, stricter criteria for grouping similar pixels. This approach prevented the algorithm from removing the breast region; see more in the Segmentation Section. The flowchart of the Improved Region Growing (IRG) algorithm is shown in Fig. 6.

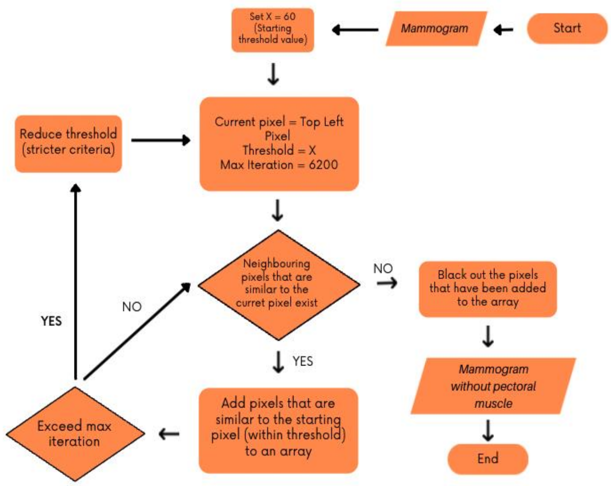


Fig. 6. Flow chart of the Improved Region Growing (IRG) algorithm to remove the pectoral muscle region from a mammogram

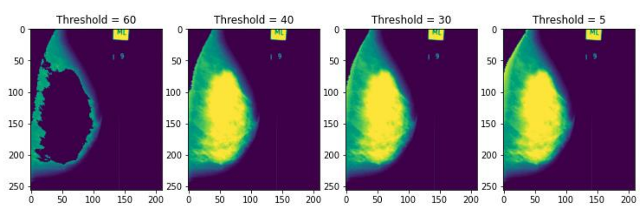


Fig. 7. The result of the improved region-growing algorithm varies with different thresholds used.

In the IRG algorithm, the threshold value is an important parameter as it determines the criteria, whether strict or loose, when grouping the pixels in the mammogram. Different thresholds would result in different results, as shown in Fig. 7. Segmentation for the mammogram (mdb003.pgm) works best when the threshold is set to 40. In this research, it was observed that different images require a different threshold to run well. Fig. 8 shows the various optimal thresholds needed by the different mammograms. Therefore, fixing one predefined threshold value before the algorithm runs would result in some mammograms not being segmented well. To solve this problem, the algorithm was modified by adding an additional step. The step involves restarting the algorithm with a lower threshold value (stricter criteria) once the algorithm has run for a predefined number of iterations. The logic behind this change is that if the algorithm runs for more than the predefined number of iterations, this means that the boundary found by the algorithm has crossed the pectoral muscle boundary, resulting in overshooting. By doing this, all mammograms will be segmented using the optimal threshold value.

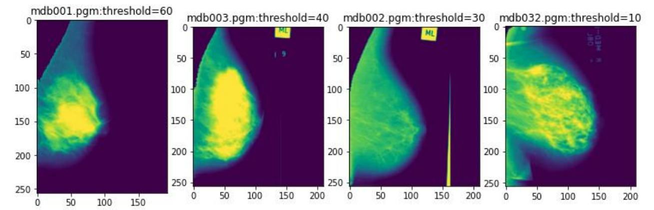


Fig. 8. Different performance results based on a variety of threshold values.

IV. Results and Discussion

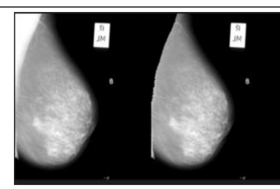
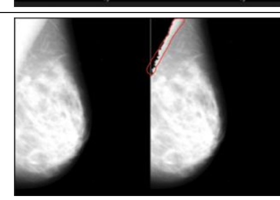
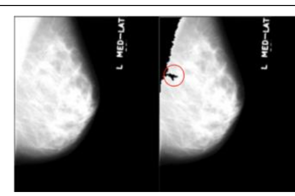
A. Evaluation of results by professional radiologists

To make sure the algorithm has removed the pectoral muscle from the mammogram completely and correctly, especially mammograms in which the pectoral muscle region is not easily seen, two radiologists were invited to evaluate the results. Both the segmentation algorithms are run with all 322 mammograms in the MIAS dataset. This section will first discuss the evaluation criteria and how the images in the MIAS dataset are classified into different categories. Then, the performance of the Level Set and Region Growing algorithm will be discussed.

B. Performance Evaluation

Each segmented mammogram can either be evaluated as acceptable or unacceptable. Table I describes the evaluation criteria. Pectoral removal is acceptable if the segmentation algorithm removes the entire pectoral muscle region from the mammography images. Pectoral removal is not acceptable if the algorithm does not remove the entire pectoral muscle region (see second row of Table I) or removes the entire muscle and some breast regions that may contain a tumor (see third row of Table I).

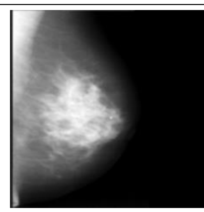
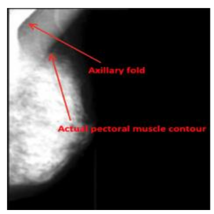
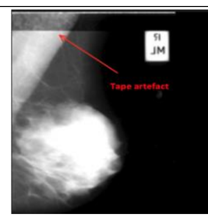
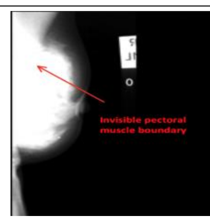
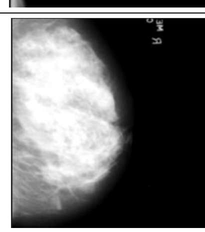
TABLE I. Evaluation Criteria

Evaluation	*Mammogram*	*Criteria*
Acceptable		The segmentation algorithm removes the pectoral muscle region from the mammogram completely.
Unacceptable		The algorithm does not remove the entire pectoral muscle region. Part of it has not been removed (red boundary)
Unacceptable		The algorithm removes the entire pectoral muscle region, and removes some of the breast regions that might contains tumor (red circle).

C. Classification of the MIAS dataset

To better understand the performance of each segmentation algorithm, all 322 mammograms in the MIAS dataset are classified into the five types as outlined in Table II. Different types of mammograms, along with their total numbers, brief descriptions, and examples, are presented in Table II.

TABLE II. CLASSIFICATION OF INPUT IMAGES OF THE MIAS DATASET

Type Code	*Count*	*Description & Image No.*	*Mammogram*
T1	210	Normal Mammogram. mdb187.pgm	
T2	84	Mammograms have an inner boundary (axillary fold) within the pectoral muscle region which results in high false positives when detecting the pectoral muscle boundary [7]. mdb039.pgm	
T3	14	Mammograms that contain tape artefact. mdb002.pgm	
T4	8	Mammograms in which the contour of the pectoral muscle region is invisible. mdb288.pgm	
T5	6	Mammograms in which the pectoral muscle region is not visible. mdb236.pgm	

D. Performance of the Segmentation Algorithms

As mentioned earlier, a professional radiologist was invited to evaluate the mammograms segmented by the Level Set algorithm and two professional radiologists to evaluate the mammograms segmented by the Improved Region Growing algorithm (IRG). The results of each algorithm were also evaluated by the development team. The results are presented in Tables III, IV, V, VI and VII. For each category of mammogram (as described in Table II), the tables show the number of mammograms classified into that category (count), followed by the number of segmented images classified by the evaluator as acceptable or unacceptable respectively. Finally, the acceptance rate for each category is recorded, and an overall acceptance rate is reported at the bottom.

1) Level set: Tables III and IV show the Level Set (LS) results of the radiologist and the team's evaluation respectively.

2) Improved Region Growing (IRG): For the results of the Improved Region Growing (IRG) algorithm, Tables V and VI show the results of the two radiologist's evaluations, and table VII shows the results of the team's evaluation.

TABLE III. EVALUATION OF LS METHOD BY RADIOLOGIST

Type Code	*Count*	*Acceptable*	*Unacceptable*	*Success Rate*
T1	210	113	97	54%
T2	84	6	78	7%
T3	14	5	9	36%
T4	8	2	6	25%
T5	6	3	3	50%
Overall Acceptance Rate				40%

TABLE IV. EVALUATION OF LS METHOD BY THE TEAM

Type Code	*Count*	*Acceptable*	*Unacceptable*	*Success Rate*
T1	210	115	95	55%
T2	84	3	81	4%
T3	14	5	9	36%
T4	8	2	6	25%
T5	6	3	3	50%
Overall Acceptance Rate				39.75%

TABLE V. EVALUATION OF THE IRG METHOD BY RADIOLOGIST 1

Type Code	*Count*	*Acceptable*	*Unacceptable*	*Success Rate*
T1	210	168	42	80%
T2	84	50	34	60%
T3	14	5	9	36%
T4	8	2	6	25%
T5	6	5	1	83%
Overall Acceptance Rate				71.42%

TABLE VI. EVALUATION OF IRG METHOD BY RADIOLOGIST 2

Type Code	*Count*	*Acceptable*	*Unacceptable*	*Success Rate*
T1	210	167	43	80%
T2	84	52	32	62%
T3	14	2	12	14%
T4	8	2	6	25%
T5	6	4	2	67%
Overall Acceptance Rate				70.40%

TABLE VII. EVALUATION OF IRG METHOD BY THE TEAM

Type Code	*Count*	*Acceptable*	*Unacceptable*	*Success Rate*
T1	210	161	49	77%
T2	84	50	34	60%
T3	14	3	11	21%
T4	8	0	8	0%
T5	6	3	3	50%
Overall Acceptance Rate				67.30%

E. Analysis

The IRG method performs better than the LS method in terms of the following aspects:

1) Removal of the pectoral muscle region from normal mammograms: The IRG method can generally remove the pectoral muscle more accurately than the LS method. Following the performance evaluation methodology, this means that the IRG method is more likely to remove the entire pectoral muscle region and is less likely to leave out part of the pectoral muscle. Fig. 9 and Fig. 10 show the performance differences between these two algorithms.

2) Removal of the pectoral muscle region from mammograms that contain axillary fold: The LS method performs poorly because the algorithm cannot properly segment a mammogram containing an axillary fold (Type T2). The axillary fold causes the algorithm to detect a false pectoral muscle contour, which results in incomplete pectoral muscle removal. In the radiologist's and team's evaluation, the algorithm only achieves a 7% and 4% acceptance rate, respectively. On the other hand, the IRG method is less likely than the LS method to detect a false contour. It achieves a higher acceptance rate when segmenting a Type T2 mammogram.

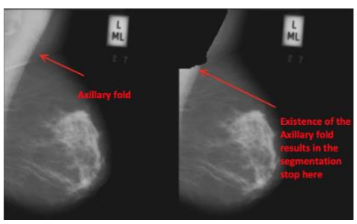


Fig. 9. An example of how the LS method fails to remove the entire pectoral muscle region.

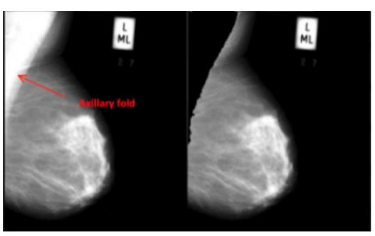


Fig. 10. An example of how the IRG method successfully removes the entire pectoral muscle region.

Fig. 11 compares the acceptable rate between the LS and IRG methods when segmenting the different types of mammograms.

Another observation that can be made is that both algorithms do not work well (acceptance rate less than 50%) with mammograms that contain tape artifacts and those that have invisible pectoral muscle contours (Type T3 and T5). This is because both segmentation algorithms are designed to identify a region based on the boundary found. First, the algorithm does not work well with type T3 mammograms because the boundary that is identified is the artifact boundary. Next, the type T5 mammogram will also not be segmented well by the algorithms as its boundary is not visible.

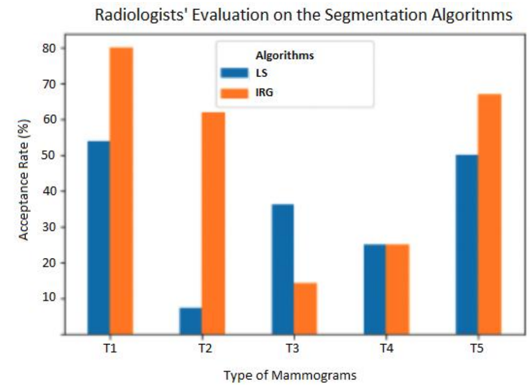


Fig. 11. Group bar chart comparing the result between the LS method and IRG method.

V. CONCLUSION

While the automatic pectoral muscle removal system is a complete system, it contains a few limitations on the preprocessing steps, segmentation algorithms, and web application. The result tabulation shows that none of the algorithms works well with mammograms containing the tape artifact (type T3). This shows that the existing preprocessing steps are not enough to ensure that both segmentation algorithms run well. The IRG method performs significantly better than the LS method. However, both algorithms have limitations. First, both algorithms do not work well with type T2 mammograms. For instance, the IRG method achieves an 80% acceptance rate on the normal mammogram (Radiologists' evaluation for T1), yet it only achieves a 62% and 60% acceptance rate on type T2 mammograms (Radiologists' evaluation for T2). Considering the number of T2 mammograms in the MIAS dataset (84 out of 322), it can be said that this type of mammogram is very common. Therefore, the low acceptance rate to segment this type of mammogram is a big limitation of the algorithm. Second, both algorithms fail when the mammogram contains an invisible pectoral muscle contour (type T4). While the LS method achieves a 25% acceptance rate, the IRG method performs poorly when segmenting the type T4 mammograms. Next, it is noticed that both segmentation algorithms cannot remove the pectoral muscle completely from a mammogram if it contains an axillary fold. Fig. 12 shows a type T2 mammogram in which the pectoral muscle is not removed completely.

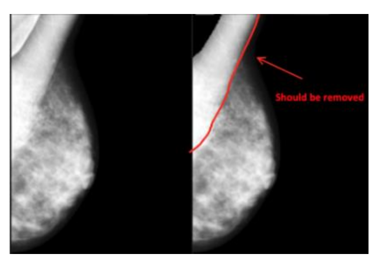


Fig. 12. A mammogram in which the pectoral muscle region is not removed completely.

One possible improvement that could be made to the IRG method to solve this problem is to place a triangle over the mammogram image to estimate the pectoral muscle region. The triangle can be drawn using the difference between the intensity of the pectoral muscle region and the breast region, along with some mathematical formulas [15]. Having estimated the pectoral muscle boundary, the region growing algorithm can be modified to dynamically adjust the threshold value based on where the current pixel is. For example, a higher threshold value can be used when the current pixel lies within the estimated boundary. In other words, a looser criterion is used for grouping similar pixels if the current pixel is within the boundary. This will allow the algorithm to completely avoid the false pectoral boundary and group the pectoral muscle region. Fig. 13 shows how the algorithm could be improved by drawing an additional triangle before the algorithm runs.

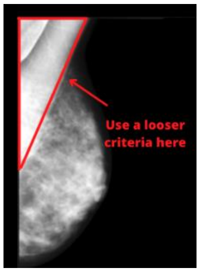


Fig. 13. An example of how a triangle can be placed on top part of the mammogram.

The study has a constraint of not validating the algorithm's performance with other datasets. Additionally, utilizing deep networks can enhance the accuracy of the proposed technique by extracting improved features from segmented images. In the future, it is recommended to automatically crop the tumor region and utilize pre-trained networks.

ACKNOWLEDGMENT

The Initial part of this work was carried out when the first author was affiliated with Monash University Malaysia.

REFERENCES

[1] Mutebi, M., Anderson, B. O., Duggan, C., Adebamowo, C., Agarwal, G., Ali, Z., ... & Eniu, A. (2020). Breast cancer treatment: A phased approach to implementation. Cancer, 126, 2365-2378.

[2] DeSantis, C. E., Bray, F., Ferlay, J., Lortet-Tieulent, J., Anderson, B. O., & Jemal, A. (2015). International variation in female breast cancer incidence and mortality rates. Cancer epidemiology, biomarkers & prevention, 24(10), 1495-1506.

[3] Raghavendra, U., Acharya, U. R., Fujita, H., Gudigar, A., Tan, J. H., & Chokkadi, S. (2016). Application of Gabor wavelet and Locality Sensitive Discriminant Analysis for automated identification of breast cancer using digitized mammogram images. Applied Soft Computing, 46, 151-161.

[4] Rampun, A., Morrow, P. J., Scotney, B. W., & Winder, J. (2017). Fully automated breast boundary and pectoral muscle segmentation in mammograms. Artificial intelligence in medicine, 79, 28-41.

[5] Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., ... & Savage, J. (2015). Mammographic image analysis society (mias) database v1. 21.

[6] Rezaei, Z. (2021). A review on image-based approaches for breast cancer detection, segmentation, and classification. Expert Systems with Applications, 182, 115204.

[7] Rampun, A., López-Linares, K., Morrow, P. J., Scotney, B. W., Wang, H., Ocaña, I. G., ... & Macía, I. (2019). Breast pectoral muscle segmentation in mammograms using a modified holistically-nested edge detection network. Medical image analysis, 57, 1-17.

[8] Mustra, M., & Grgic, M. (2013). Robust automatic breast and pectoral muscle segmentation from scanned mammograms. Signal processing, 93(10), 2817-2827.

[9] Pawar, S. D., Sharma, K. K., Sapate, S. G., & Yadav, G. Y. (2021). Segmentation of pectoral muscle from digital mammograms with depth-first search algorithm towards breast density classification. Biocybernetics and Biomedical Engineering, 41(3), 1224-1241.

[10] Digabel, H.L., C, Iterative Algorithms. Proceedings of the 2nd European Symposium Quantitative Analysis of Microstructures in Material Science, Biology and Medicine, 1978. 66(1): p. 240-256.

[11] Gonzales, R.C. and P. Wintz, Digital image processing. 1987: Addison-Wesley Longman Publishing Co., Inc.

[12] Vikhe, P. S., & Thool, V. R. (2016). Intensity based automatic boundary identification of pectoral muscle in mammograms. Procedia Computer Science, 79, 262-269.

[13] Taifi, K., Ahdid, R., Fakir, M., Elbalaoui, A., Safi, S., & Taifi, N. (2017, May). Automatic breast pectoral muscle segmentation on digital mammograms using morphological watersheds. In 2017 14th International Conference on Computer Graphics, Imaging and Visualization (pp. 126-131). IEEE.

[14] Gómez, K. A. H., Echeverry-Correa, J. D., & Gutiérrez, Á. Á. O. (2021). Automatic pectoral muscle removal and microcalcification localization in digital mammograms. Healthcare Informatics Research, 27(3), 222-230.

[15] Hazarika, M., & Mahanta, L. B. (2018). A novel region growing based method to remove pectoral muscle from MLO mammogram images. In Advances in Electronics, Communication and Computing: ETAEERE-2016 (pp. 307-316). Springer Singapore.

[16] Osher, S., & Sethian, J. A. (1988). Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. Journal of computational physics, 79(1), 12-49.

[17] Li, C., Huang, R., Ding, Z., Gatenby, J. C., Metaxas, D. N., & Gore, J. C. (2011). A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI. IEEE transactions on image processing, 20(7), 2007-2016.

[18] Zhou, S., Wang, J., Zhang, M., Cai, Q., & Gong, Y. (2017). Correntropy-based level set method for medical image segmentation and bias correction. Neurocomputing, 234, 216-229.

[19] Anitha, J., & Peter, J. D. (2015). Mass segmentation in mammograms using a kernel-based fuzzy level set method. International Journal of Biomedical Engineering and Technology, 19(2), 133-153.

[20] Wang, K., Khan, N., Chan, A., Dunne, J., & Highnam, R. (2019, November). Deep learning for breast region and pectoral muscle segmentation in digital mammography. In Pacific-Rim Symposium on Image and Video Technology (pp. 78-91). Cham: Springer International Publishing.

[21] 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, 30(4), 1108-1118.

[22] Kim, Y. J., Yoo, E. Y., & Kim, K. G. (2021). Deep learning based pectoral muscle segmentation on Mammographic Image Analysis Society (MIAS) mammograms. Precision and Future Medicine, 5(2), 77-82.

[23] Elkorany, A. S., & Elsharkawy, Z. F. (2023). Efficient breast cancer mammograms diagnosis using three deep neural networks and term variance. Scientific Reports, 13(1), 2663.

[24] Ferrari, R. J., Rangayyan, R. M., Desautels, J. L., Borges, R. A., & Frere, A. F. (2004). Automatic identification of the pectoral muscle in mammograms. IEEE transactions on medical imaging, 23(2), 232-245.

[25] Wang, Z., Ma, B., & Zhu, Y. (2021). Review of level set in image segmentation. Archives of Computational Methods in Engineering, 28, 2429-2446.

[26] Rezaei, Z., Selamat, A., Taki, A., Rahim, M. S. M., & Abdul Kadir, M. R. (2019). Systematic mapping study on diagnosis of vulnerable plaque. Multimedia Tools and Applications, 78, 21695-21730.