



# International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

## Segmentation Of Pectoral Muscle From Digital Mammograms With Breadth -First Search Algorithm Towards Breast Cancer Detection

Pratibha T Joshi<sup>1</sup>, Gurpreet Singh Saini<sup>2</sup>, Shivaji D Pawar<sup>3</sup>, Kamal Sharma<sup>4</sup>, Shankar Patil<sup>5</sup>

<sup>1</sup>Research Scholar, Lovely Professional University, Punjab, and Department of Electronics and Telecomm SIES GST, Nerul, Navi Mumbai, India, Email: [Pratibhaj@sies.edu.in](mailto:Pratibhaj@sies.edu.in)

<sup>2</sup>School of Electronics and Electrical Engineering, Lovely Professional University, Punjab, India, Email: [gurpreet.16889@lpu.co.in](mailto:gurpreet.16889@lpu.co.in)

<sup>3</sup>School of AI and Future Technologies, Universal AI University Karjat, Mumbai, India, Email: [shivajidpawar20@gmail.com](mailto:shivajidpawar20@gmail.com)

<sup>4</sup>Professor, Ambala college of Engineering and Applied Research India, Email: [Kamalsharma111@gmail.com](mailto:Kamalsharma111@gmail.com)

<sup>5</sup>Professor, Smt. Indira Gandhi College of Engineering, Navi Mumbai, India, Email: [smpatil2k@gmail.com](mailto:smpatil2k@gmail.com)

### Abstract

Digital mammography, a powerful tool for early breast cancer detection, can face challenges when the pectoral muscle is present in the breast area, making accurate classification of breast density and diagnosis difficult. This article introduces a robust methodology that effectively segments the pectoral muscle from digital mammographic images by applying the breadth-first search (BFS) algorithm with a heuristic approach. The proposed methodology includes preprocessing digital mammograms using the SAGCWD algorithm to enhance image quality. The BFS algorithm then detects the entire pectoral muscle as a single connected component and removes artifacts and tags from the background region. Finally, the BFS algorithm is used with and without a heuristic technique to eliminate the pectoral muscle from digital mammographic images. The proposed algorithm is subjected to rigorous testing on 2500 images from the DDSM dataset, 322 images from the MIAS dataset, and 194 images from the In-Breast dataset. Expert Radiologists help subjectively calculate the Segmentation Accuracy of the proposed algorithm while quantitative metrics like Jaccard Index and Dice Coefficient are useful in calculating the same objectively. The overall segmentation accuracy subjectively is reported to be an impressive 85.40%, with mean values of the Jaccard index and the Dice similarity coefficient at 0.96 and 0.97, respectively. These metrics highlight the high level of agreement between the algorithm's output and the ground truth, confirming its effective segmentation performance. The successful integration of the proposed algorithm into the preprocessing stage of digital mammograms underscores its potential to enhance the accuracy of pectoral muscle removal, a crucial step in breast cancer classification.

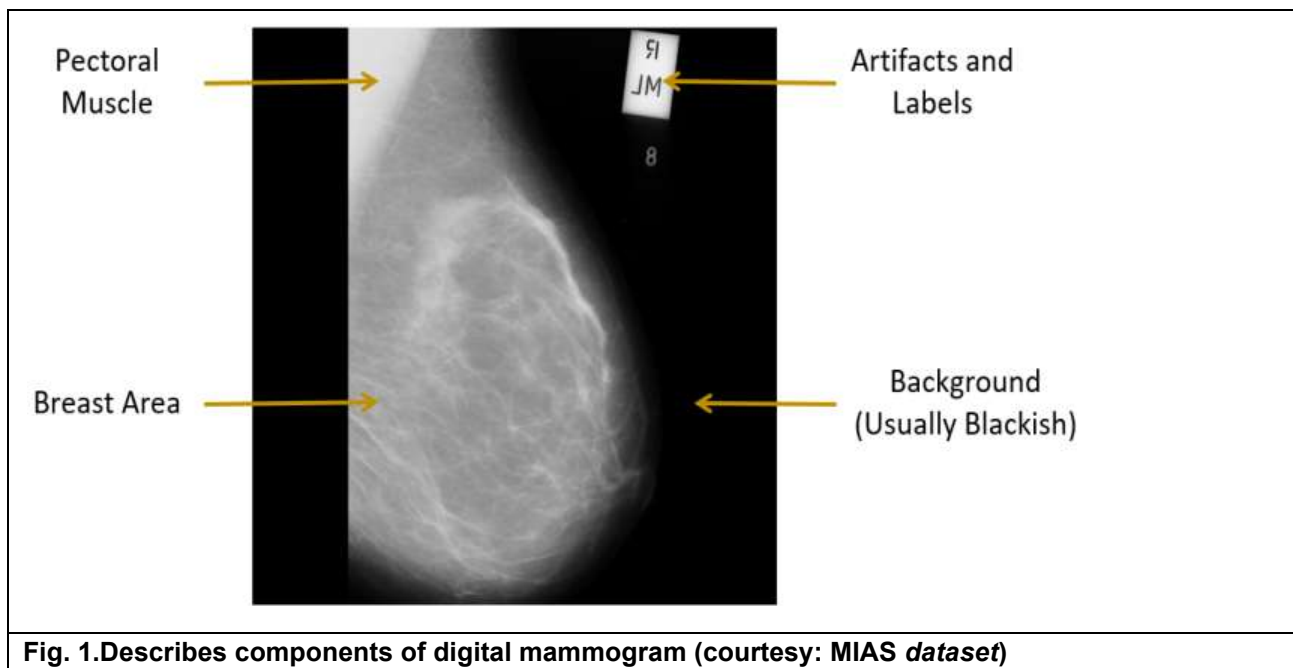
*Keywords: Breast cancer; Digital mammogram; Pre-processing techniques; Breadth First Search; Pectoral muscle removal*

This is an open access article under CC BY 4.0, allowing unrestricted use with proper attribution, a license link, and indication of any changes made.

## 1. Introduction

Breast cancer has been a lethal disease for the population of women worldwide for the last many decades. Early-stage discovery and care are necessary factors for saving the lives of breast cancer patients. [1,2] Detecting breast cancer at an earlier stage allows for more efficient and less aggressive treatment options, ultimately raising the probability of survival. It not only improves treatment results but can also decrease the financial burden and emotional trauma associated with progressive stages of breast cancer. Mammography is amongst the well-regarded means for screening breast cancer because of its effectiveness, widespread availability, and relatively low cost compared to other imaging modalities. [3] Breast cancer detection involves a multi-step process that includes preprocessing, region of interest (ROI) segmentation, feature extraction, and classification. [4] Preprocessing is an integral part of any cancer detection. Preprocessing involves image enhancement, removing noise, artifacts, and pectoral muscle from the image. The pectoral muscle is part of the

chest to which the breast is attached to the body. So, the removal of pectoral muscle becomes very important to reduce the false positive rates of breast cancer diagnosis. [4,5] As indicated in Figure 1, the mammographic image contains different integral parts like artifacts, labels, and pectoral muscle.



**Fig. 1.Describes components of digital mammogram (courtesy: MIAS dataset)**

Microcalcification, masses, architectural alterations, and mammographic breast density are widely accepted biomarkers to predict risk for breast cancer. The mammography tool is identified as effective tool to perform preliminary diagnosis of breast cancer. The complexity in breast cancer detection increases due to Imbalanced brightness, noise, and the fact that fibro-glandular tissues and the pectoral muscle appear similar.[6] For CAD-based effective breast cancer detection, careful preprocessing is an essential step that consists of removing artifacts, labels, noise, and the removal of the pectoral muscle.

The article aims to improve image interpretability by reducing noise, enhancing contrast, removing artifacts, detecting breast borders, and eliminating pectoral muscles. The proposed research article consists of six sections. Section 2 describes an extensive survey of the existing state-of-the-art literature; Section 3 exhibits the dataset details that has been used. Section 4 provides the proposed methodology, while Section 5 explains the experimental results. Finally, Section 6 concludes the work.

## 2. Literature Survey

Advanced image processing techniques, such as thresholding, edge detection, region growing, machine learning, and deep learning-based approaches, are essential for improving the detection and diagnosis of breast cancer. [6,7] These methods accurately identify and delineate the pectoral muscle periphery, enhancing breast tissue visualization and examination. With precise segmentation of the pectoral muscle, radiologists can confidently identify suspicious areas that may indicate cancer or other abnormalities. [8,9] This section highlights some state-of-the-art techniques used to segment pectoral muscle in digital mammograms. Under intensity-based technique Region Growing, Thresholding, and Watersheds are the most popular algorithms used for pectoral muscle removal. Region Growing begins from seed points and iteratively adds neighbouring pixels that meet similarity criteria. It works on the basic idea that pixels with the same intensities belong to the same region. Saeed et al. [6] introduced the Hybridization Bounding Box and Region Growing (HBBRG) algorithm, aiming to accurately segment the pectoral muscle in mammographic images. By combining the strengths of the Bounding Box and Region Growing techniques, the algorithm gives a high level of accuracy. Other technique: Thresholding functions as a fundamental technique for image segmentation by dividing an image into regions based on pixel intensity values. Otsu's method is an outstanding example of local

thresholding, aiming to find optimal thresholds for individual areas. [7,8]. Soukaina et al. [9] implemented a method for breast lump segmentation and extracting the pectoral muscle from mammographic images using a combination of techniques, including Otsu's thresholding, image classification-based k-means, hidden Markov models, and region growing. They used the MIAS database with an accuracy of 91.92%, and the error rate of 8.07%, suggesting that this method performs well in segmenting tumours and eliminating the pectoral muscle. Another intensity-based method, the watershed algorithm is a technique used to divide images into subdivisions based on intensity or gradient information. Camillus et al. [10] proposed a comprehensive segmentation framework integrating edge detection, gradient computation, watershed transformation, and morphological reconstruction. Their method was evaluated on 84 mammograms from the MIAS database, achieving a mean false positive rate of 0.85% and a false negative rate of 4.88%, though it required extensive preprocessing and parameter tuning to enhance performance. Anuradha et al. [11] further progressed watershed-based segmentation by integrating a region merging strategy to address the issue of over-segmentation, conveying accuracy levels ranging from 90% to 95% depending on threshold selection. These studies demonstrate that while the watershed algorithm is a powerful tool for pectoral muscle segmentation, its sensitivity to noise and tendency toward over-segmentation necessitate its use within multi-stage frameworks or in combination with complementary techniques to achieve robust and accurate results. For the segmentation of the pectoral muscle other technique used is a short-term Fourier transform. Wavelets provide spatial frequency information from images. Wavelet decomposition is a method that dissects an image into various scales and orientations, revealing both coarse and fine details. Moghbel et al. [12] proposed a comprehensive process for extracting the breast region from mammogram images to remove the pectoral muscle. This process involves multilevel 2D wavelet decomposition, breaking the image into various frequency bands. The detailed sub-bands resulting from wavelet decomposition are then processed using Minimum Mean Squared Error (MMSE). Once the sub-bands have been processed, inverse wavelet decomposition is applied to reconstruct an enhanced image. The approach shows promise, with an accuracy rate of 92.22% in removing the pectoral muscle. All these methods suffer from several limitations. Intensity based approach rely on grey level differences between pectoral muscle and breast tissue which makes them highly sensitive to noise and intensity variations mostly in dense mammograms. Thresholding method highly depends on seed selection and often results into over or under segmentation. On the other hand, wavelet transform-based methods, even though effective in multiresolution analysis, face challenges such as the difficulty of choosing appropriate wavelet functions and decomposition levels, high computational complexity at higher resolutions, and sensitivity to noise that affects wavelet coefficients. [13] Deep learning-based methods bids significant advantages over traditional based methods. These techniques are crucial for cancer detection as they can analyze complex patterns and extract meaningful features from large datasets. [14] The popularity of deep learning techniques is due to their early detection, high sensitivity, and specificity, as well as their automated tumour segmentation and feature extraction. [15] Guo et al. adopted a two-stage framework for pectoral muscle segmentation, combining U-Net CNN and GAN for refinement. While acquiring a high Dice similarity coefficient of 97.03%, their approach introduces computational complexity due to integrating multiple deep learning models. [16] Despite this limitation, their method significantly enhances accuracy, showcasing robust and precise segmentations. Soleimaniet al's [17] method, offers a promising approach to breast-pectoral boundary segmentation, leveraging the strengths of both deep learning and graph theory-based techniques. Its high accuracy and robust performance across multiple datasets emphasize its potential utility in clinical settings for mammographic image analysis and computer-aided diagnosis tasks. The proposed algorithm is evaluated on three different datasets: MIAS, CBIS-DDSM, and INbreast, with an average accuracy (ACC) of 99.64%. Yu et al. [18] Suggested PeMNet model to segment breast pectoral muscle from mammographic images. PeMNet uses Global Channel Attention Module (GCAM) to improve segmentation accuracy and attains excellent segmentation accuracy and outperforms other methods on the image dataset. Overall, Deep learning's application in automatic pectoral muscle removal has the potential to transform mammogram preprocessing, leading to more accurate and efficient breast cancer detection and analysis. Careful consideration of data quality, model architecture, and training optimization is crucial for achieving optimal results in this context.

## **2.1 Research Gap**

Pectoral muscle segmentation from digital mammograms is challenging, and no technique has been found to be optimal for clinical implementation so far. Researchers face several critical issues in this field, which are listed below.

1. The muscle's diversity and texture make it challenging to develop techniques that can effectively remove it from the images. Additionally, the variety of pectoral muscle shapes and sizes across different images complicates the selection of initial seeds for segmentation algorithms. [23,24]
2. Another concern is the need for a standardized dataset and universally accepted ground truth for validation [21,22]. Without a common platform for comparison, it becomes difficult for researchers to assess and benchmark the effectiveness of their segmentation approaches.
3. There have been limited studies on graph search methods for segmenting the pectoral muscle in mammograms. [27]
4. However, deep learning techniques show promising potential and encourage segmentation task alternatives. Although the computational complexity of these models and their need for larger datasets can pose practical difficulties. [30, 35]

Overcoming the above-mentioned challenges could significantly advance the accuracy and reliability of pectoral muscle segmentation. This could lead to improved diagnosis and treatment in mammography.

### **3. Resource Dataset**

The proposed algorithm is implemented using three publicly available datasets: MIAS, DDSM, and In Breast. The Digital Database for Screening Mammography (DDSM) contains a broad range of cases including normal, benign, and cancer, providing a diverse set for analysis. This database includes multiple views for each patient case, including MLO, LMO, left CC, and right CC. The MIAS dataset comprises 322 images, while the INBreast dataset uses 194 images. [35, 36, 37] Expert radiologists manually annotated all images in the datasets to ensure high-quality ground truth. Their expertise ensures the accuracy and reliability of the annotated contours, which will serve as a benchmark for evaluating the effectiveness of the proposed algorithm.

### **4. Proposed Methodology**

This section illustrates the proposed methodology for separating pectoral muscle. Fig.2 shows the proposed methodology using the Breadth First Search (BFS) algorithm to segment pectoral muscle. It is divided into image preprocessing, BFS search, and heuristic approach segmented pectoral muscle image output.

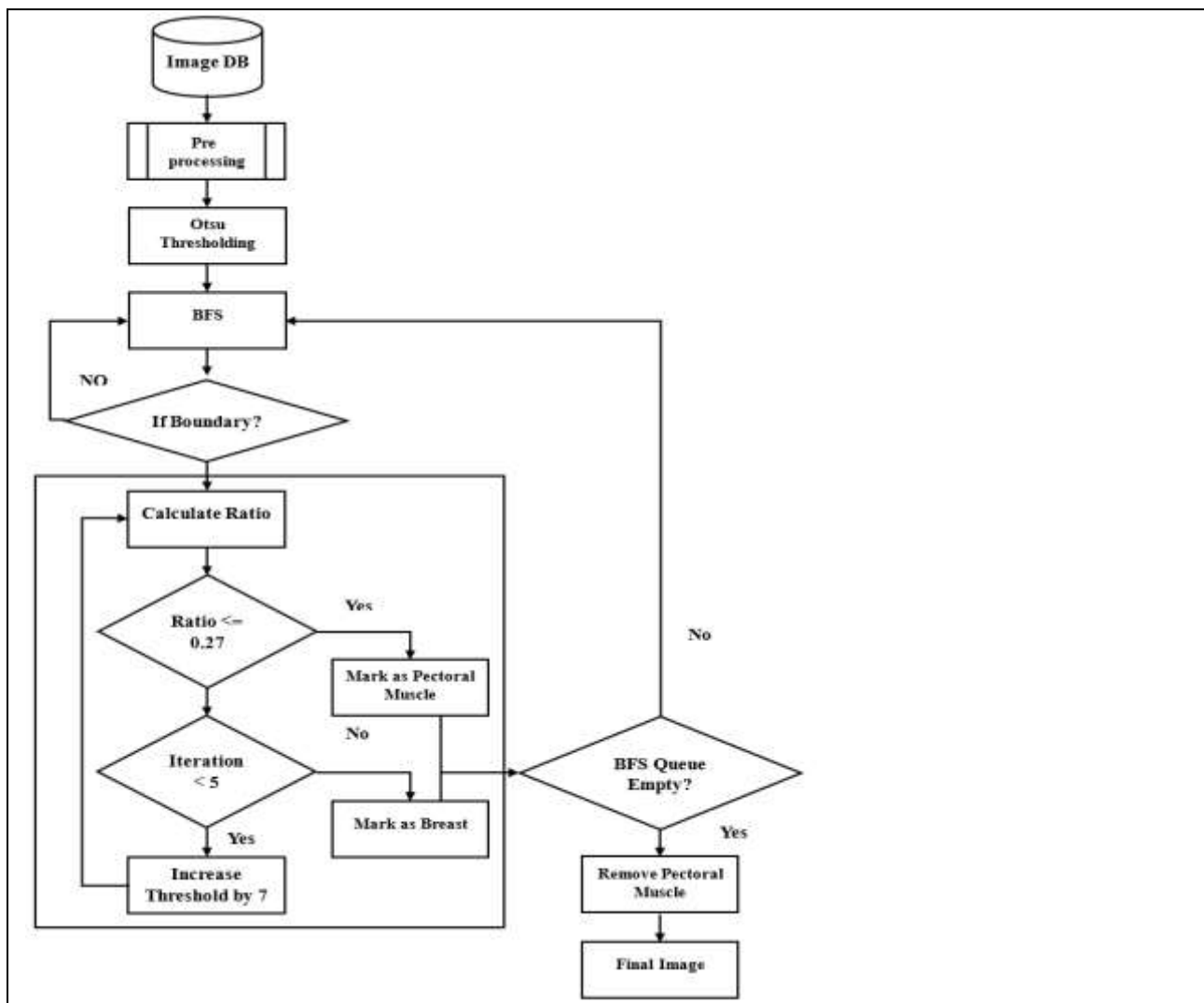


Fig. 2. Proposed BFS Flowchart for pectoral muscle segmentation

### 4.1 Preprocessing

Digital mammography, along with advanced image processing techniques, plays a significant role in early detection and diagnosis of breast cancer, which can lead to improved patient outcomes and survival rates. [29, 34] As part of the pre-processing steps, the image undergoes normalization, followed by flipping. Since some breast images in the dataset are left-oriented, and some are right-oriented, we make all images left-oriented to ensure uniformity. This involves positioning the pectoral muscle in the upper-left corner of the image. Then, histogram-equalization of the image is performed, followed by cropping based on calculations from left and right boundaries. This step equalizes the image's histogram and removes any accompanying labels on the image.

### 4.2 Contrast Enhancement

Various contrast enhancement techniques are commonly used in medical image processing with digital mammography to improve the visibility of structures and reduce noise. Techniques such as wavelet transform, histogram equalization, and adaptive histogram equalization (AHE) are frequently utilized. [19] However, the unsharp masking (USM) technique can introduce artifacts and make the output image look unrealistic, which makes it unsuitable for medical image enhancement. On the other hand, Adaptive Contrast Enhancement (ACE) can enhance high-frequency components, but its implementation requires extensive computational resources and pre-processing time. An alternative approach is to use Adaptive Gamma Correction with Weighting

Distribution (AGCWD) to implement contrast enhancement. This algorithm combines gamma correction and histogram equalization to adjust the histogram and improve image visibility.

$$T(l) = l \max \left( \frac{l}{l_{max}} \right)^\gamma \dots\dots\dots (1)$$

Eq-1 gives gamma correction. where

$l(\max)$  = maximum intensity of the applied image.

$\gamma$ = adaptive parameter

$T(l)$ = image after gamma correction

$$T(l) = F(l). l_{max} \dots\dots\dots (2)$$

Where

$F(l)$  = Cumulative Distribution function (C.D.F.) obtained based on probability Density function (P.D.F.)  $f$  is given by

$$f(l) = \frac{n_l}{n_{tot}} \dots\dots\dots (3)$$

$n_l$ = number of pixels with intensity value  $l$

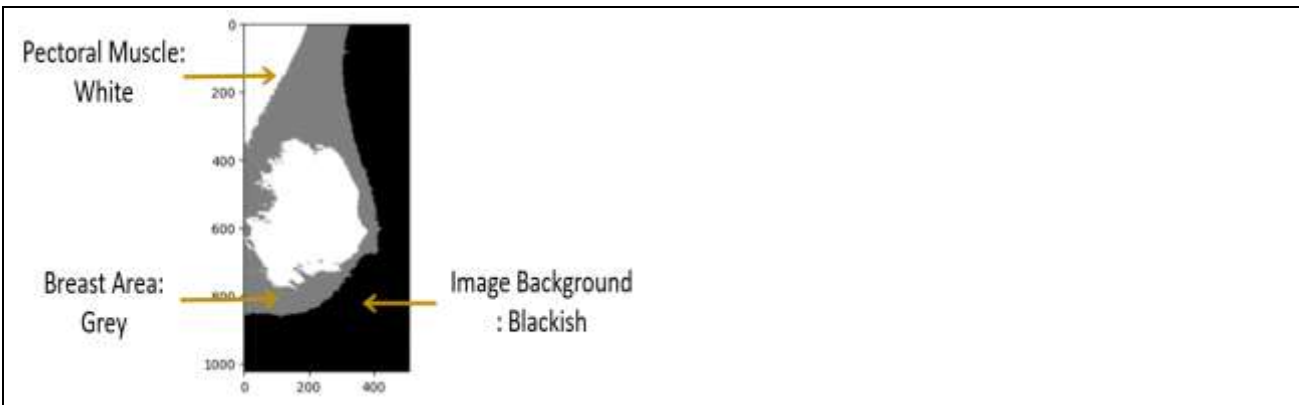
$n_{tot}$  = Totalpixels in the image

C.D.F. can be calculated as given in Equation4

$$(f)l = \sum_{k=0}^l f(k) \dots\dots\dots (4)$$

**4.3 Breast border detection**

The next step is to identify the border of the enhanced image's breast. To achieve this, two threshold values are obtained using local thresholding with Otsu's 3-class classification. Otsu's method is a thresholding technique that automatically determines the optimal intensity threshold to separate pixels into multiple classes, maximizing the inter-class variance to segment objects from the background in an image effectively. [14] It calculates the total number of pixels and the mean intensity of the entire image. It then iterates through all possible values and determines three possible values. Probabilities are calculated based on the histogram of pixel intensity values and the total number of pixels. By summing up the histogram values within the specified intensity ranges and dividing them by the total number of pixels, three probability values are obtained, which divide the image into black, gray, and white parts. If the probability value is less than the threshold, it will be treated as the black part of the image, which represents the background. If it is greater than the threshold, it is the white part of the image, which indicates the pectoral muscle. The probability value between these two threshold values represents the breast part, which is the grey part of the image. Figure 6 displays the different parts of the breast after thresholding.



**Fig. 3.Parts of Breast**

### 4.4 Initial Seed Selection

The images obtained in the previous stage are used to remove the pectoral muscle. The pectoral muscle is located at the top left of the picture, but its size and shape can vary, and in some cases, it may be tiny or absent. [23] To address this, the image is divided into four parts for initial seed selection, with the pectoral muscle in part one. Then, a dynamic seed selection is implemented for the proposed algorithm. The seed pixel is dynamically selected from the first part of the image, which provides optimal results, as shown in Figure 6.

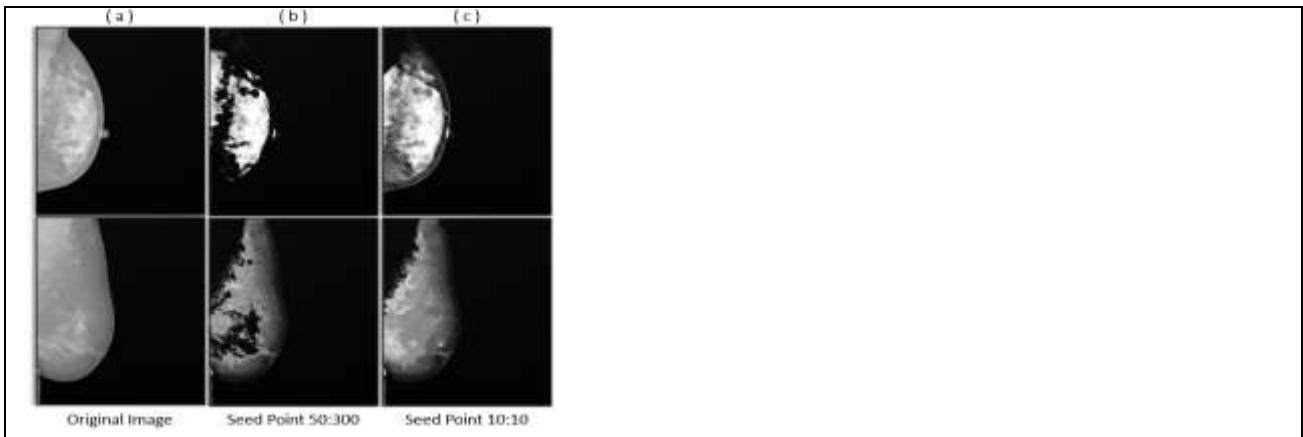


Fig. 4.Improvement in over-segmentation using BFS algorithm with heuristic approach.

Figure 4-a gives the original image. With an initial seed point of (50,300), the output is over-segmented (4-b). And using a heuristic approach to change the seed point to (10,10), as shown in Fig (4-c), the over-segmentation is largely reduced, and the pectoral muscle is successfully removed.

### 4.5 Pectoral muscle removal using BFS

After the seed point selection, the Breadth First Search algorithm is applied to each image from different datasets. Connected components in the image are identified using the Breadth First Search algorithm. It starts at the root node (initial seed) and identifies all the neighbouring nodes at the current depth level before moving on to nodes at the next depth level. BFS implements deque (double encoded queue) used as the main queue [39]. It keeps track of visited pixels. After the selection of the initial seed point, for each pixel, it checks the neighbouring pixels in four directions to check whether they are in image boundaries and the same class as a current pixel [40]. If so, it adds the next neighbouring pixel to the queue for further exploration and makes it visited. The connected component analysis is done to identify the pectoral muscle.



Fig. 5. Pectoral muscle detection using Breadth First Search Algorithm

As shown in the above diagram. BFS algorithm will start from Node A. It will check the adjacent Nodes in all four directions for a given node. Here, Node A has only two connected nodes, Node B and Node F [41]. Breadth-first search algorithm generates a queue of neighbouring pixels, it iterates through the queue and examines the neighbouring pixels & creates an array of visited pixels to keep a record of visited nodes. The algorithm continues until the BFS queue is empty.

**Pixel Checking:** For each pixel, it checks the neighbouring pixel in four directions to check if they are in image boundaries and the same class as a current pixel. If the class changes, then it is considered as the boundary limit. Once the Boundary limit is reached or identified, it skips the nodes in that row and moves on to the next row. e.g. In the above image, when it goes to Node M from Node L, there is a change of boundary, so from Node L, it will go straightaway to Node P. Return Visited array: Returns the visited array for the total visited pixels, and these nodes are ignored for the next iteration even if they are in the neighbouring pixel queue for any node.

#### 4.6 BFS with the heuristic approach

For the selection of the seed point and at the boundary node of the pectoral muscle and breast part of the image, the proposed algorithm uses a heuristic approach. For seed selection, the heuristic approach is already discussed in section 4.4. And at the periphery of the pectoral muscle and breast wall also it applies the heuristic approach. After the selection of the initial seed point, that point will be considered as node 1, then horizontally search for the next node will be there. Accordingly, it will calculate the ratio of pectoral muscle pixels and the total count of breast pixels, including pectoral muscle. If that ratio  $< 0.27$ , it's part of the pectoral muscle; otherwise, it checks a condition involving an optional threshold and increments by 7 to check whether it is less than 0.27. If it is not less than 0.27, it will again increase the threshold check threshold; if less than 0.27, it is a pectoral muscle, and greater than 0.27, it is part of the breast area. It will check the condition for five iterations to check whether the selected node is part of the pectoral muscle or the breast area. If the selected node is part of the breast area, it will skip the node get shifted to the next row and will follow the BFS queue. Thus, for every pixel row at the boundary point, the algorithm will apply this heuristic approach to check the node as part of a muscle or breast area. Once all the nodes from the BFS queue are checked, the entire pectoral muscle will separate from the breast area. Thus, using BFS and a heuristic approach, the pectoral muscle is removed successfully from the breast area

$$\text{Ratio} = \text{count of } x = \frac{\text{count of pectoral pixels}}{\text{count of breast pixel including pectoral muscle}} < 0.27$$

### 5. Results

Proposed segmentation algorithm is applied over different datasets with and without Heuristic approach. For assessments all sample images are classified into five different categories for each dataset. Subjective and objective evaluation is done. Results for Subjective evaluation is evaluated on the basis:

- Correctly segmented—If the ratio of segmented pectoral muscle to actual pectoral muscle in the original image is 95-100%, it is classified as correctly segmented.
- Acceptable accuracy - If the ratio of segmented pectoral muscle to actual pectoral muscle in the original image is 90-94%, it is classified as Acceptable accuracy.
- Under Segmented - If the ratio of segmented pectoral muscle to actual pectoral muscle in the original image is less than 90%, it is classified as Under-Segmented.
- Over-Segmented - if the proposed algorithm removes pectoral muscle and also some part of the breast then it is classified as Over-Segmented
- Lastly, for some images algorithm will not work and there is no pectoral muscle segmentation at all, or images are discarded

Segmentation Accuracy	Criteria for subjective assessment	Dataset Referred		
		% matching with radiologist ground truth	MIAS	DDSM
Images considered		322	2500	194
Images segmented Correctly.	95-100	119	995	46
Images In acceptable range of segmentation	90-95	112	1052	34
Images under-segmented	< 90	29	249	34
Images Over- segmented	images with breast part removed	48	147	56
Images not segmented /Discard	unable to segment	14	57	24
Segmentation Accurac		87.65	85.375	83.20
Overall segmentation Accuracy		85.40%		

### 5.1 Subjective Assessment of Results

This section discusses about the subjective and objective assessments results. To check the accuracy, the obtained output of the proposed algorithms was compared with the ground truth marked by an expert radiologist. This is a crucial validation step, indicating that the algorithm's output corresponds well with the expert's annotations or markings. The analysis is performed on datasets such as MIAS, DDSM, and INBreast.

Subjective assessment of results

$$\% \text{ segmentation Accuracy} = \frac{\text{Total pre-processed Images}}{\text{Total Input images}} * 100$$

### 5.2 Objective assessment of results

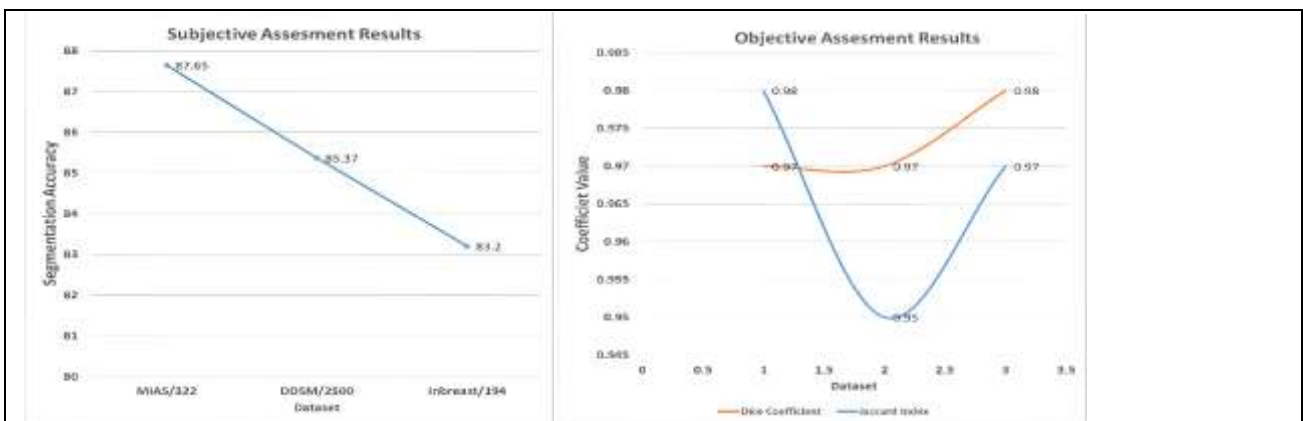
To validate the segmentation accuracy, various indices are calculated. E.g. Jaccard Coefficient and Dice similarity coefficient.

$$\text{Jaccard Index} = \frac{\text{Intersection area}}{\text{Union area}} = \frac{A \cap B}{A \cup B}$$

$$\text{Dice similarity Coefficient} = \frac{2 * \text{Intersection area}}{\text{Segmented area} + \text{Ground Truth area}} = \frac{2 * IA \cap BI}{|AI| + |BI|}$$

where A and B are the SET of ground truth image and sample image respectively. Both the coefficients range from 0 to 1. 1 indicates perfect overlap, while 0 means no overlap. Table 1 gives the experimental result

Dataset/no of images	Accuracy	Jaccard Index	Dice similarity coefficient
MIAS/322	87.65	0.980	0.970
DDSM/2500	85.375	0.950	0.970
InBreast/194	83.20	0.970	0.980



**Fig. 6. Subjective and objective assessment of the proposed algorithm**

Table 2 provides Segmentation accuracy, Jaccard index, and dice similarity coefficient for sample images to elaborate different segmentation levels. Table 3,4 and 5 gives the image outputs when the algorithm is implemented for all the three datasets.

original and segmented output for some images for different datasets are as follows:

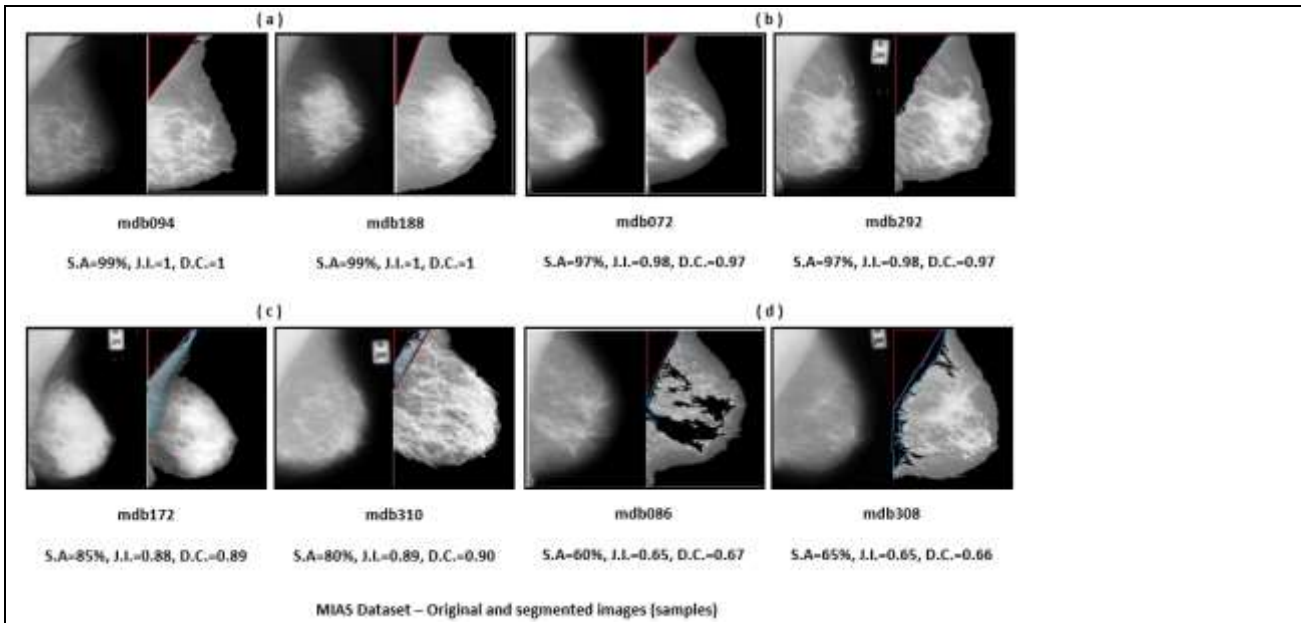


Fig. 7. MIAS Dataset original and segmented images

Subjective and objective assessment of the proposed algorithm Image from MIAS dataset shows (a) Correct segmentation (b) Acceptable segmentation (c) Under segmentation (d) Over segmentation.

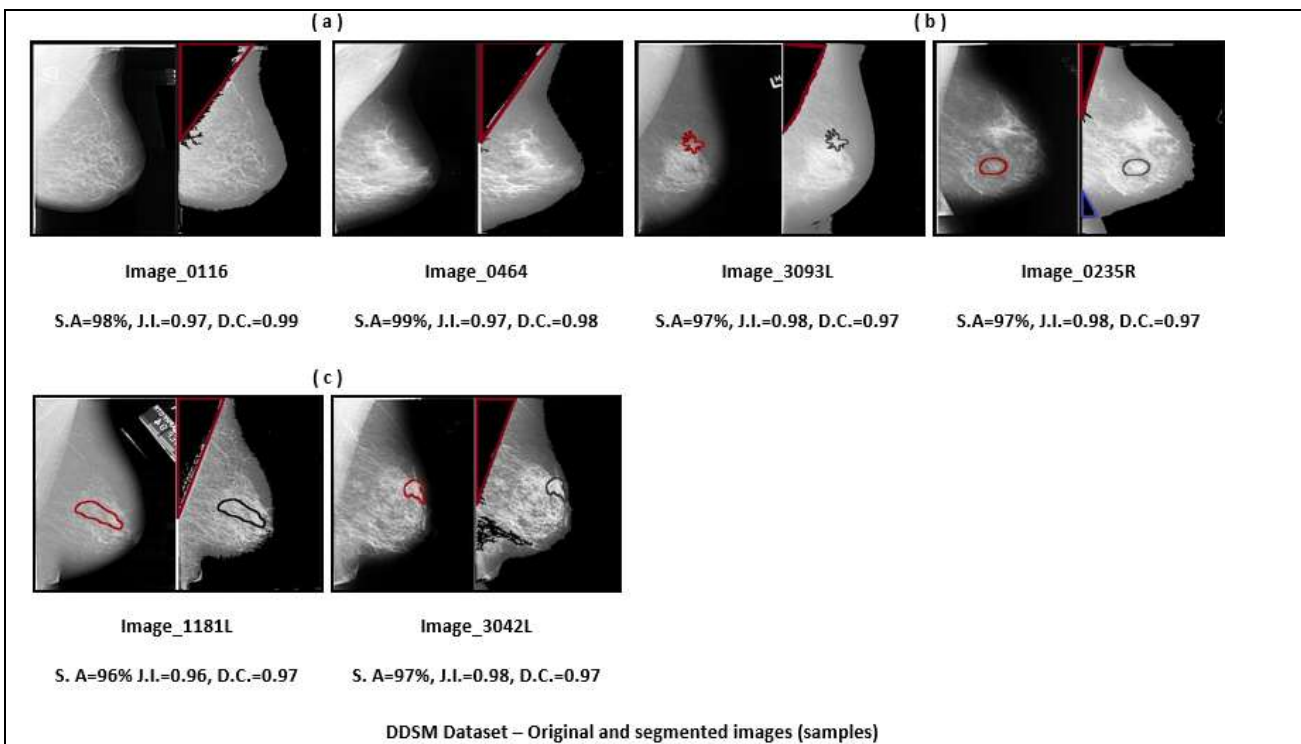
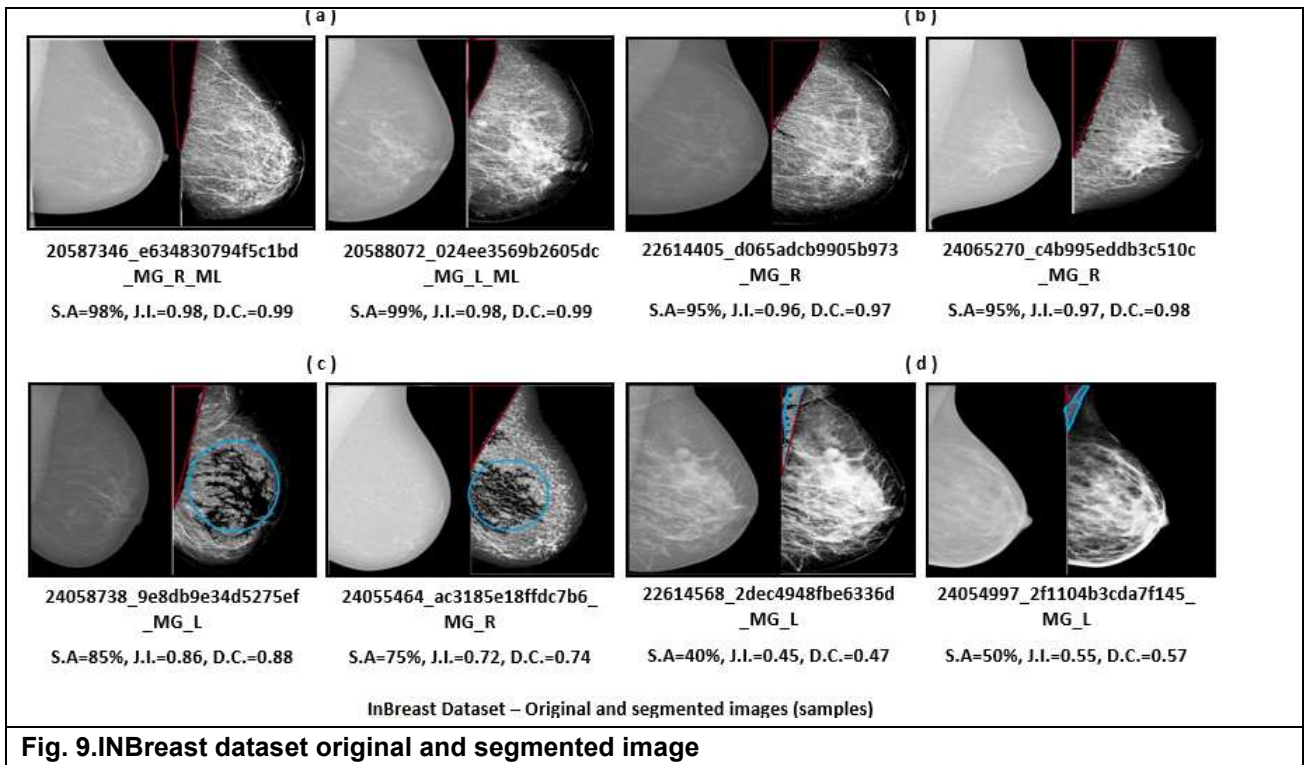


Fig. 8. DDSM dataset original and segmented images

Subjective and objective assessment of the proposed algorithm (a) Normal breast (b) Benign breast (c) cancer Breast.



**Fig. 9.** INBreast dataset original and segmented image

Fig.9 shows Subjective and objective assessment of the proposed algorithm Image from InBreast dataset shows (a) Correct segmentation (b) Acceptable segmentation (c) Under segmentation (d) Over segmentation. The figure (7,8,9) shows the images from different datasets. For each of the image types we have segmentation types as Correct, Acceptable, Over and Under segmentation like previous datasets.

**5.3 Results of pectoral muscle segmentation with and without Heuristic Approach**

As discussed in section 4.4 with initial seed selection results are obtained with and without Heuristic approach with all three datasets. Table (3) demonstrates the outcome of this experiment. The results show an approximate 3% increase in accuracy with the heuristic approach.

Table 3. Impact of Heuristic Approach						
Dataset	Segmented Accuracy		Jaccard Index		Dice Similarity Coefficient	
	without Heuristic	with Heuristic	without Heuristic	with Heuristic	without Heuristic	with Heuristic
MIAS	84.32%	87.65%	0.97	0.98	0.956	0.97
DDSM	81.98%	85.35%	0.93	0.95	0.96	0.97
InBreast	81.02%	83.20%	0.95	0.96	0.966	0.98

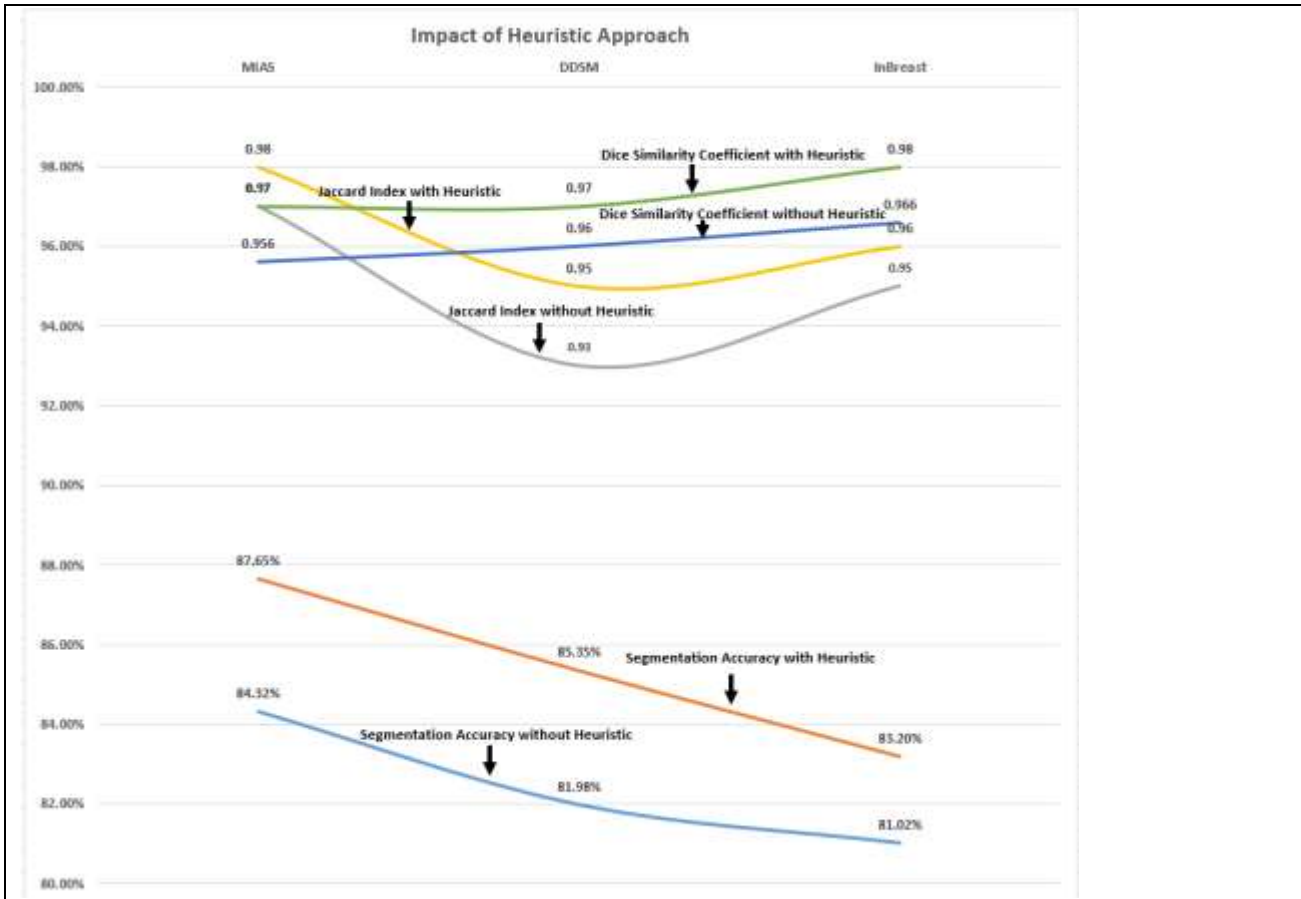


Fig. 10. Outcome with and without Heuristic Approach

## 6. Discussion

Accurate identification and segmentation of the pectoral muscle are critical for improving the precision of breast cancer screening and analysis. Small presence of pectoral muscle in the mammograms decreases the accuracy of breast cancer prediction [28,42]. As the position of the pectoral muscle is at the leftmost corner of the image, dynamic seed selection of the initial seed selection is taken for experimentation, and BFS with a heuristic approach for identification and segmentation of the pectoral muscle is applied. Otsu’s multiclass thresholding is used to classify the image into three different classes, namely, pectoral muscle, breast part, and background of the image. These different classes are used to detect pectoral muscle from the entire image. From the results, we observed that for some images, the algorithm presents over-segmentation. In those cases, the muscle segmentation was done correctly, but since the brightness of pixels near the chest boundary is the same as that of the muscle, the algorithm extended the boundary of the muscle beyond the pectoral muscle (into the breast area). It has been observed that, especially for the InBreast dataset, there were many cases of either over or under-segmentation. The table gives the exact count of these cases. Variation in the intensities of pixels and the shape and size of pectoral muscle plays crucial roles in the success rate of the BFS algorithm. Veins in the breast appear of the same brightness as pectoral muscle in the mammograms. So, the boundary of the pectoral muscle gets extended into the breast area as these veins are connected to the pectoral muscle. The same was observed with MIAS and DDSM datasets also, but it is more prominent in the InBreast dataset. Thus, the accuracy of the segmentation depends on the image quality and standardization.

### 6.1 Comparative study

As researchers use different image databases and evaluation methodologies, direct comparison is difficult, but logical comparison of the proposed methodology with present techniques is bestowed. Comparative results are split into two categories.

### Subjective assessment - Comparison

Medical image analysis research commonly involves qualitative evaluation of pectoral muscle separation methods involving professional specialist radiologists to authenticate accuracy of segmentation. Table 4 gives the subjective analysis. It gives state-of-art accuracy assessment. Mustra et al. [25] methodology combines boundary detection and wavelet decomposition techniques to mark and segment the pectoral muscle while minimizing noise and granulation. The accuracy of pectoral segmentation was 85%. The method proposed by Ergin involves a two-step process: pre-segmentation of the breast border followed by the pectoral muscle separation in mammography images. But this logic needs some post-processing operations, then. Only the segmentation accuracy obtained was 94.40%. Hazarika et al. [26] proposed region growing algo to separate pectoral muscle. A gradient-based smoothing method is employed to improve the segmented pectoral muscle boundary rather than morphological-based filtering. The proposed method provides promising results in the segmentation of pectoral muscle, but further optimization and fine-tuning are required to improve accuracy. Pawar et al. [27] recommended a mechanism to dynamically select seed and combine it with and without the heuristic methodology of the DFS (Depth First Search) algorithm to mark the pectoral muscle. Otsu multi-class thresholding was applied for the initial seed selection for the DFS application. The algorithm was found robust and gave a segmentation accuracy of 86.18% with the DDSM database. The VGG-SVM system is compared with VGG-DLA by researcher Saliha Zahoor et al. [28] Used MEWOA (Modified Entropy controlled Whale Optimization Algorithm) for feature selection of highly correlated features and classification. The author used fine-tuned pre-trained models Nasnet mobile and MobilenetV2 for feature extraction and a serial-based threshold approach for feature fusion. Augmented data for improved model performance achieves high accuracies for INbreast (99.7%), MIAS (99.8%), and CBIS-DDSM (93.8%) datasets with minimized processing time. Yu et al. [29] recommended PeMNet model to separate pectoral muscle from mammogram images. PeMNet implements the Global Channel Attention Module (GCAM) and the Deeplabv3+ architecture with minimal parameter overheads to enhance segmentation accuracy. PeMNet decisively excelled in state-of-the-art models with remarkable segmentation results: Global Pixel Accuracy: 99.48%. Since it's a deep learning model, the algorithm is robust and can handle any image change

Year	Author	Technique	Dataset/ Image Count	Accuracy
2009	Mustra et al.	Wavelet Decomposition	MIAS/322	85.00%
2018	Hazarika et al.	Region Growing	MIAS/150	86.67%
2021	Pawar et al.	Depth first Search with heuristic approach	DDSM/2675	86.18%
2022	Zahoor et al.	Deep learning using transfer learning	MIAS	99.80%
			INBREAST	99.70%
			DDSM	93.80%
2025	Yu et al.	Deep Learning -PeMNet	OPTIMAM +INBreast /882	99.48%
Proposed algorithm		Breadth First Search algorithm	MIAS/322	87.65%
			DDSM/2500	85.38%
			Inbreast/194	83.20%

### Objective Comparative Study

Researchers evaluated the performance of their techniques using statistical tools like the Jaccard index and the Dice coefficient. Shen et al. [30] proposed a method that combines genetic algorithms and morphological selection to segment mammograms. Rampunet al. [31] presented a method to automatically separate breast pectoral muscle in mediolateral oblique (MLO) mammograms by deploying a Convolutional Neural Network (CNN). The proposed method is robust and capable of reducing false positives/negatives resulting in a more accurate segmentation. Zebari et al. [32] addressed challenges associated with noise and low-quality mammogram images in the context of image segmentation. Researchers reduced the impact of these limitations

by improving mammograms prior to segmentation, but with poor or missing borders' segmentation accuracy was affected.

Pawar et al. [27] proposed a mechanism to dynamically select seed and apply with and without heuristic method of the DFS (Depth First Search) algorithm to recognize the pectoral muscle. Otsu multi-class thresholding was applied for the initial seed selection for the DFS application. The algorithm was found robust and gave a segmentation accuracy of 86.18%. With the DDSM database. In some cases, the segmentation accuracy fails because of over segmentation or when there is overlapping of breast tissue and pectoral muscle. The algorithm also fails if the grey level intensity is less than 100. Deb et al. applied modified U-Net architecture using MIAS and INBreast dataset for pectoral muscle removal. Tables 7 and 8 give a state-of-the-art methods summary. Deb et al. [20] implemented a segmentation approach built on U-Net with densely connected convolution layers. While Angelone et al. proposed semiautomatic method with U-Net for elimination of pectoral muscle. [34][35]

**Table 5. Objective Assessment – A Comparison**

Year	Author	Method	Dataset/ Image count	Jaccard Index	Dice coefficient
2018	Shen et al.	Thresholding using Genetic algo.	MIAS	0.912	0.949
			DDSM	0.944	0.971
			INBreast	0.846	0.891
2019	Rampun et al.	Deep Learning	MIAS/322	0.946	0.975
			INBreast	0.926	0.956
2020	Zebari et al.	Thresholding and Machine Learning	MIAS/322	0.981	0.990
			INBreast /200	0.986	0.987
2021	Pawar et al.	Depth first Search + heuristic approach	DDSM/2675	0.932	0.955
2022	Yu et al.	PeMNet	OPTIMAM +INBreast /882	0.963	0.933
2023	Deb et al.	Modified U-Net	Mini -MIAS, INBreast	-	0.96
2024	Angelone et al.	U-Net	1199 patches from 240 FFDM	0.94	0.97
This Paper's method		Breadth First Search + Heuristic approach	MIAS/322	0.980	0.970
			DDSM/2500	0.950	0.960
			INBreast/194	0.970	0.980

### 6.2 Limitations and Futures cope of the study

- In some cases, portion of the inner breast is separated as a pectoral muscle, causing over-segmentation. This is observed when intensity at the skin boundary and muscle is the same. As the algorithm differentiates breast images using three intensity levels, if the breast part at the boundary is of the same intensity, it will consider it the muscle part and not the breast part and continue segmenting until the intensity change occurs.
- At the boundary level of muscle and body, the node, we considered a heuristic approach, considering five iterations. It might be possible that after five iterations, a particular node is a node from pectoral muscle and not breast portion. As per the BFS algorithm, after five iterations, the next node will be considered from the next row.
- **Limitation** The accurate removal of the pectoral muscle from images is crucial for improving the reliability and accuracy of the segmentation process. The proposed algorithm has shown encouraging results, but under and over-segment results can be improved by refining the algorithm. Algorithm performance can be enhanced by analysing cases in the categories of under and no-segmentation. Understand the patterns and specific scenarios where the algorithm struggles. Fine-tuning of the algorithm will further improve the results obtained.

### 7. Conclusion

To detect the Breast cancer, Identifying and segmenting the pectoral muscle in mammographic images constitute the fundamental stages. These two steps greatly influence the improvement in clinical outcomes and research insights. The prime objective of the research is to first clean the image by removing the artifacts, tags along with the pectoral muscle by applying BFS algorithm. With an innovative heuristic approach, the BFS algorithm with the heuristic approach produced 85.40% segmentation accuracy. The proposed algorithm demonstrates robust performance across a diverse set of mammographic images with variety of textures, sizes, and shapes. Moreover, the proposed segmentation method is effective and comparative to previous methods.

## Acknowledgement

The authors gratefully acknowledge Dr. Kiran Sonje Specialist Radiologist, Thane for their valuable and timely help in marking ground truth on some difficult cases of mammogram and providing subjective evaluation of proposed research work.

## References

1. Luo, C., Wang, L., Zhang, Y., Lu, M., Lu, B., Cai, J., ... & Dai, M. (2022). Advances in breast cancer screening modalities and status of global screening programs. *Chronic Diseases and Translational Medicine*, 8(02), 112-123.
2. Maqsood, S., Damaševičius, R., & Maskeliūnas, R. (2022). TTCNN: A breast cancer detection and classification towards computer-aided diagnosis using digital mammography in early stages. *Applied Sciences*, 12(7), 3273.
3. Pawar, S. D., Sharma, K. K., Sapate, S. G., Yadav, G. Y., Alroobaea, R., Alzahrani, S. M., & Hedabou, M. (2022). Multichannel DenseNet architecture for classification of mammographic breast density for breast cancer detection. *Frontiers in Public Health*, 10, 885212.
4. Rezaei, Z. (2021). A review on image-based approaches for breast cancer detection, segmentation, and classification. *Expert Systems with Applications*, 182, 115204.
5. Zebari, D. A., Zeebaree, D. Q., Abdulazeez, A. M., Haron, H., & Hamed, H. N. A. (2020). Improved threshold based and trainable fully automated segmentation for breast cancer boundary and pectoral muscle in mammogram images. *Ieee Access*, 8, 203097-203116.
6. Saeed, E. M. H., & Saleh, H. A. (2020, April). Pectoral muscles removal in mammogram image by hybrid bounding box and region growing algorithm. In *2020 International Conference on Computer Science and Software Engineering (CSASE)* (pp. 146-151). IEEE.
7. Joshi, P. T., Saini, G. S., & Pawar, S. D. (2023, December). Comparative Analysis of Different Preprocessing Techniques for Digital Mammograms Towards Breast Cancer Detection. In *2023 6th International Conference on Advances in Science and Technology (ICAST)* (pp. 234-239). IEEE.
8. Fadhil, S. S., & Dawood, F. A. A. (2021). Automatic pectoral muscles detection and removal in mammogram images. *Iraqi Journal of Science*, 676-688.
9. El Idrissi El Kaitouni, S., Abbad, A., & Tairi, H. (2018). A breast tumors segmentation and elimination of pectoral muscle based on hidden Markov and region growing. *Multimedia Tools and Applications*, 77(23), 31347-31362.
10. Camilus, K. S., Govindan, V. K., & Sathidevi, P. S. (2011). Pectoral muscle identification in mammograms. *Journal of applied clinical medical physics*, 12(3), 215-230.
11. Anuradha, P. V., Jose, B. R., & Mathew, J. (2015). Improved segmentation of suspicious regions of masses in mammograms by watershed transform. *Procedia Computer Science*, 46, 1483-1490.
12. Moghbel, M., Ooi, C. Y., Ismail, N., Hau, Y. W., & Memari, N. (2020). A review of breast boundary and pectoral muscle segmentation methods in computer-aided detection/diagnosis of breast mammography. *Artificial Intelligence Review*, 53(3), 1873-1918.
13. Zebari, D. A., Haron, H., Zeebaree, S. R., & Zeebaree, D. Q. (2019, April). Retracted: Enhance the Mammogram Images for Both Segmentation and Feature Extraction Using Wavelet Transform. In *2019 International conference on advanced science and engineering (ICOASE)* (pp. 100-105). IEEE

14. Pawar, S. D., Sharma, K. K., & Sapate, S. G. (2021). Advances in machine learning and deep learning approaches for mammographic breast density measurement for breast cancer risk prediction: an overview. *Design of Intelligent Applications Using Machine Learning and Deep Learning Techniques*, 125-143.
15. Joshi, P. T., Saini, G. S., & Pawar, S. D. (2025, June). Mass and Microcalcification Detection in Mammograms: A Deep Learning Perspective on Breast Cancer Classification. In *2025 International Conference on Emerging Trends in Industry 4.0 Technologies (ICETI4T)* (pp. 1-8). IEEE.
16. Guo, Y., Zhao, W., Li, S., Zhang, Y., & Lu, Y. (2020). Automatic segmentation of the pectoral muscle based on boundary identification and shape prediction. *Physics in Medicine & Biology*, 65(4), 045016.
17. Soleimani, H., & Michailovich, O. V. (2020). On segmentation of pectoral muscle in digital mammograms by means of deep learning. *IEEE Access*, 8, 204173-204182.
18. Yu, X., Wang, S. H., Górriz, J. M., Jiang, X. W., Guttery, D. S., & Zhang, Y. D. (2022). PeMNet for pectoral muscle segmentation. *Biology*, 11(1), 134.
19. Huang, S. C., Cheng, F. C., & Chiu, Y. S. (2012). Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE transactions on image processing*, 22(3), 1032-1041.
20. Deb, S. D., & Jha, R. K. (2024). Segmentation of pectoral muscle from mammograms using U-Net having densely connected convolutional layers. *Multimedia Tools and Applications*, 83(8), 24505-24526.
21. Pawar, S., Joshi, P., Sharma, K., & Sapate, S. (2023, June). Review on pre-processing algorithms for breast density classification using digital mammograms. In *AIP Conference Proceedings* (Vol. 2717, No. 1). AIP Publishing
22. 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.
23. Michael, E., Ma, H., Li, H., Kulwa, F., & Li, J. (2021). Breast cancer segmentation methods: current status and future potentials. *BioMed research international*, 2021(1), 9962109.
24. Rana, S., & Singh, L. (2023, December). A Survey on Segmentation Techniques in Fractal Classification Systems. In *2023 Third International Conference on Smart Technologies, Communication and Robotics (STCR)* (Vol. 1, pp. 1-6). IEEE.
25. Mustra, M., Bozek, J., & Grgic, M. Breast border extraction and pectoral muscle detection using wavelet decomposition. In *IEEE EUROCON 2009* : 1426-1433
26. Hazarika, M., & Mahanta, L. B. (2017). A novel region growing based method to remove pectoral muscle from MLO mammogram images. In *Advances in Electronics, Communication and Computing: ETAEEER-2016* (pp. 307-316). Singapore: Springer Singapore.
27. 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.
28. Zahoor, S., Shoaib, U., & Lali, I. U. (2022). Breast cancer mammograms classification using deep neural network and entropy-controlled whale optimization algorithm. *Diagnostics*, 12(2), 557.
29. Yu X, Wang SH, Górriz JM, Jiang XW, Guttery DS, Zhang YD. (2022) PeMNet for pectoral muscle segmentation. *Biology* ;11(1):134.
30. Shen, R., Yan, K., Xiao, F., Chang, J., Jiang, C., & Zhou, K. (2018). Automatic pectoral muscle region segmentation in mammograms using genetic algorithm and morphological selection. *Journal of digital imaging*, 31(5), 680-691.
31. Rampun, A., Morrow, P. J., Scotney, B. W., & Winder, J. Fully automated breast boundary and pectoral muscle segmentation in mammograms. *Artificial intelligence in medicine*: 79, 28-41.
32. Zebari, D. A., Zeebaree, D. Q., Abdulazeez, A. M., Haron, H., & Hamed, H. N. A. (2020). Improved threshold based and trainable fully automated segmentation for breast cancer boundary and pectoral muscle in mammogram images. *Ieee Access*, 8, 203097-203116.
33. Pawar, S. D., Joshi, P. T., Savalkar, V. A., Sharma, K. K., & Sapate, S. G. (2022, August). Past, present and future of automated mammographic density measurement for breast cancer risk prediction. In *Journal of Physics: Conference Series* (Vol. 2327, No. 1, p. 012076). IOP Publishing.

34. Angelone, F., Ponsiglione, A. M., Grassi, R., Amato, F., & Sansone, M. (2025). U-net based approach for pectoralis muscle segmentation in digital mammography. *Computer Methods and Programs in Biomedicine Update*, 100210.
35. Joshi, P. T., Saini, G. S., & Pawara, S. D. (2025). Application Of Densenet Architecture And Its Variants Towards Breast Cancer Detection: A Multi-View Analysis. *International Journal of Environmental Sciences*, 11(9s), 835-854.
36. [http://www.eng.usf.edu/cvprg/Mammography/DDSMM/thumbnails/cancers/cancer\\_13/overview.html](http://www.eng.usf.edu/cvprg/Mammography/DDSMM/thumbnails/cancers/cancer_13/overview.html)
37. <https://www.kaggle.com/datasets/kmader/mias-mammography>
38. Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS. Inbreast: toward a full-field digital mammographic database. *Academic radiology*. 2012 Feb 1;19(2):236-48.
39. Ramasubbu, L. (2025). Deep Learning and Optimization Approaches in an Efficient Cascaded Deep Capsule Cell Attention Network Model for Breast Cancer Molecular Subtypes Prediction Using Mammogram Images: A Review. *Multidisciplinary Journal of Research in Engineering and Technology*, 12(2), 149–155.
40. Saeedzada, X. (2025). AI-Based Deep Capsule Attention Network for Breast Cancer Molecular Subtype Prediction Using Mammograms. *ITSI Transactions on Electrical and Electronics Engineering*, 14(2), 97–109.
41. V, V. K., Nayak, A. N., & Arunkumar, T. R. (2026). Sensitivity-Based Breast Cancer Characterization on CBIS-DDSM Mammography Dataset Using Fine-Tuned EfficientNet-B2 Model. *International Journal on Advanced Electrical and Computer Engineering*, 15(1S), 365–371.
42. Khobragade, S., Sambhale, I., Mahajan, A., & Mondal, D. (2025). Breast Cancer Detection System Using Machine Learning. *International Journal on Advanced Computer Theory and Engineering*, 14(1), 623–626.