



Automatic Pectoral Muscle Removal and Microcalcification Localization in Digital Mammograms

Kevin Alejandro Hernández Gómez¹, Julian D. Echeverry-Correa², Álvaro Ángel Orozco Gutiérrez¹

¹Automatics Research Group, Faculty of Engineering, Universidad Tecnológica de Pereira (UTP), Risaralda, Colombia

²Data Analysis and Computational Sociology Research Group, Faculty of Engineering, Universidad Tecnológica de Pereira (UTP), Risaralda, Colombia

Objectives: Breast cancer is the most common cancer diagnosed in women, and microcalcification (MCC) clusters act as an early indicator. Thus, the detection of MCCs plays an important role in diagnosing breast cancer. **Methods:** This paper presents a methodology for mammogram preprocessing and MCC detection. The preprocessing method employs automatic artefact deletion and pectoral muscle removal based on region-growing segmentation and polynomial contour fitting. The MCC detection method uses a convolutional neural network for region-of-interest (ROI) classification, along with morphological operations and wavelet reconstruction to reduce false positives (FPs). **Results:** The methodology was evaluated using the mini-MIAS and UTP datasets in terms of segmentation accuracy in the preprocessing phase, as well as sensitivity and the mean FP rate per image in the MCC detection phase. With the mini-MIAS dataset, the proposed methods achieved accuracy scores of 99% for breast segmentation and 95% for pectoral segmentation, a sensitivity score of 82% for MCC detection, and an FP rate per image of 3.27. With the UTP dataset, the methods achieved accuracy scores of 97% for breast segmentation and 91% for pectoral segmentation, a sensitivity score of 78% for MCC detection, and an FP rate per image of 0.74. **Conclusions:** The proposed preprocessing method outperformed the state-of-the-art methods for breast segmentation and achieved relatively good results for pectoral muscle removal. Furthermore, the MCC detection module achieved the highest test accuracy in identifying potential ROIs with MCCs compared to other methods.

Keywords: Machine Learning, Breast Neoplasms, Classification, Calcinosi, Diagnosis

Submitted: March 25, 2021

Revised: 1st, May 30, 2021; 2nd, July 17, 2021

Accepted: July 19, 2021

Corresponding Author

Kevin Alejandro Hernández Gómez

Automatics Research Group, Faculty of Engineering, Universidad Tecnológica de Pereira (UTP), Carrera 27 #10-02, Pereira, Risaralda 660003, Colombia. Tel: +57-315-5642658, E-mail: Kevin_loco@utp.edu.co (<https://orcid.org/0000-0003-4506-9004>)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2021 The Korean Society of Medical Informatics

1. Introduction

Breast cancer is considered to be one of the most common global health problems [1], and more than 700,000 women die of breast cancer annually [2]. Mammography is the most widely used method for screening asymptomatic women aiming at the early detection of breast cancer, and has been demonstrated to reduce the cancer-related mortality rate by 30%–70% [3]. Mammograms are examined by specialist radiologists, who may miss a significant proportion of abnormalities or make incorrect diagnoses due to the fatigue caused by examining a large number of patients daily and inherent difficulties associated with reading mammograms.

Studies have shown that approximately 10%–30% of cancers are missed on examinations [4].

Microcalcifications (MCCs) can be found in mammograms and act as an early indicator of up to 50% of all non-palpable breast cancers. They are also present in about 93% of all cases of ductal carcinoma in situ. Thus, the detection of MCC clusters is crucial in the diagnosis of breast cancer [5]. Various computer-assisted detection (CAD) methods have been proposed to improve the accuracy of the detection of potential MCC clusters. The majority of these methods consist of several stages; the initial stages are focused on preprocessing and finding individual MCC candidates, while subsequent stages aim at grouping candidates into clusters and removing false-positive (FP) clusters [6].

Mammogram preprocessing remains a challenging task due to the noise and artefacts present in images, the complex shapes of the pectoral muscle contour, the high density of some tissues in the breast, the appearance of auxiliary folds, and pectoral muscle superposition with fibroglandular tissue [7]. Some methods perform a stage of contrast enhancement using the gamma correction technique or contrast-limited adaptive histogram equalization [8]. The main problem with these contrast enhancement methods is the fact that they increase the contrast of both the breast and the background, leading to an increase in the intensity of scan-induced artefacts that can be difficult to address using threshold-based approaches. Furthermore, many pectoral muscle segmentation techniques fail to achieve satisfactory segmentation when labels appear too close to the muscle contour [9]. Some methods also assume that the contour of the pectoral muscle is triangular, whereas in the vast majority of actual clinical cases, the curvature of the pectoral muscle contour is not linear [7].

Preprocessing techniques have been proposed to increase the contrast between MCCs and other high-intensity regions [10]. The segmentation of individual MCCs has been achieved using gray-level-based methods with empirically defined parameters such as region growing [11] and gray-level thresholding in the preprocessed region-of-interest (ROI) [12], along with other sophisticated parameter-free segmentation techniques such as morphological operations [13], watershed algorithms [14], and Bayesian pixel classification [15]. Many methods are focused on feature extraction and selection based on spatial gray-level dependence, gray-level run-length, and surrounding-region dependence [16]. Other methods employ features such as shape, texture wavelet, and multi-wavelet-based features [17].

1. Related Work

Extensive research has focused on artefact removal and breast region segmentation. For example, Qayyum and Basit [18] proposed using Otsu thresholding to extract the breast region and morphological operations to remove artefacts. An evaluation based on visual inspection showed that the segmentation was accurate for 99.37% of the images. Slavkovic-Ilic et al. [19] proposed using a K-means clustering algorithm for segmenting the breast tissue, with enhanced image contrast using the adaptive gamma correction with weighting distribution (AGCWD) method to remove artefacts; in their analysis of segmentation accuracy, their proposed method provided good and acceptable segmentation in 97.51% of cases. Yoon et al. [20] also proposed using a K-means algorithm for segmenting the breast region. In an evaluation based on a visual inspection of the segmentation accuracy, their proposed method provided good and acceptable segmentation in 93.16% of images.

Others works centered on pectoral muscle segmentation. For example, Camilus et al. [21] proposed a graph-cut approach for removing the pectoral muscle. The initial pectoral muscle region was approximated using thresholding and graph-cut segmentation methods with the final refined boundary computed by a Bezier curve-fitting approach. Their proposed method achieved an FP rate of 0.64% and a false-negative (FN) rate of 5.58%. A similar approach based on the graph-cut method and Bezier curve estimation was also proposed by Abdellatif et al. [22], whose method achieved an FP rate of 1.2% and an FN rate of 20.4%. Shinde and Rao [23] proposed an approach for removing the pectoral muscle from a presegmented breast region using a support vector machine (SVM). The initial pectoral muscle region was then segmented by three different segmentation methods, which included K-means clustering, region growing, and Otsu thresholding techniques. Next, the statistical and texture features from the initial pectoral muscle region identified by each of the segmentation methods were extracted using a gray-level co-occurrence matrix and used in an SVM classifier to determine the best segment corresponding to the pectoral muscle region. Their proposed method provided acceptable segmentation in 93.7% of cases using a visual inspection of the accuracy of the segmentation, making it comparable (but not more accurate) than other proposed methods.

Several studies have focused on the detection and classification of MCCs using machine learning techniques. For example, El-Naqa et al. [24] employed an SVM to detect MCC clusters, achieving a sensitivity score of 94% and one FP case

per image. Liu et al. [25] proposed an MCC cluster detection method based on probabilistic fuzzy C-means clustering and weighted SVM, reporting a sensitivity score of 92% and an FP rate of 2.3 per image. Wang et al. [26] used a convolutional neural network (CNN)-based approach to locate MCC clusters and reported 90% sensitivity with an FP rate of 0.69 per image. Wang and Yang [27] proposed a context-sensitive deep neural network and reported 85% sensitivity with an FP rate of 0.4 per image. Many approaches reported in the literature do not automatically localize MCC clusters, instead relying on manual classification of selected regions as MCC clusters or normal regions [10]. In contrast, one of the main contributions of this work is the presentation of a beginning-to-end methodology, where, from a raw mammographic image, the ROI and existing MCCs are automatically identified.

This paper presents a method for mammogram preprocessing and MCC detection. The preprocessing method employs an automatic artefact deletion and pectoral muscle removal algorithm based on the background estimation, homogeneity-based region-growing segmentation, and contour fitting. The MCC detection method employs a CNN-based approach for ROI classification, along with morphological operations and wavelet reconstruction for reducing the number of FP cases. The proposed methods were evaluated using the publicly available the mini-Mammographic Image Analysis Society (mini-MIAS) dataset and a private dataset (UTP dataset) in terms of segmentation accuracy in the preprocessing phase, as well as sensitivity and the mean FP rate per image in the MCC detection phase.

II. Methods

1. Datasets

The proposed methodology for MCC detection was tested on two datasets; first, the publicly available mini-MIAS dataset contains 322 mammograms, 28 of which correspond to

suspicious MCC cases as confirmed by biopsies (gold standard), with a size of 1024×1024 pixels. Second, the UTP dataset is part of the EJEALS dataset collected in 2014 by the Automatic Research Group at the Universidad Tecnológica de Pereira. This dataset contains 510 mammograms, 49 of which correspond to confirmed MCC cases; both datasets have craniocaudal (CC) and mediolateral oblique (MLO) views, with a size of 3560×4640 pixels.

2. Mammogram Preprocessing

Mammograms should be preprocessed before CAD algorithms can be applied to them for the tasks of classification and detection. The preprocessing steps include noise removal, radiopaque artefact suppression, and pectoral muscle removal. The pectoral muscle constitutes a predominant density region in the majority of MLO views of mammograms, which can affect the results of image processing, in terms of both the accuracy and the speed of analysis due to the significantly reduced area of the image to be examined.

1) Noise removal and radiopaque artefact suppression

Let $I \in \mathbb{N}^{h \times w}$ be an original image, where h and w are the height and width of the image. Each image is converted to grayscale (Figure 1A). Next, we use a median filter with a 3×3 neighborhood window. Subsequently, the Otsu thresholding method is applied to obtain a binary image (B) (Figure 1B), with the threshold value set to 0.1. Next, the frame is removed using a binary mask (F) built from the image original boundaries (h, w), with a margin shift experimentally set to $\lceil w/14 \rceil$ because the relationship between the image size and frame size is about 1/14 (Figure 1C). The Hadamard product (\otimes) between B and the binary frame F is further applied. The connected components are identified as white regions. The largest region represents the breast; the rest of the region (B_f) is discarded (Figure 1D). $I \otimes B_f$ is then applied to extract the breast tissue region from the original image (Figure 1E). Finally, the image is reduced to a rectangle that

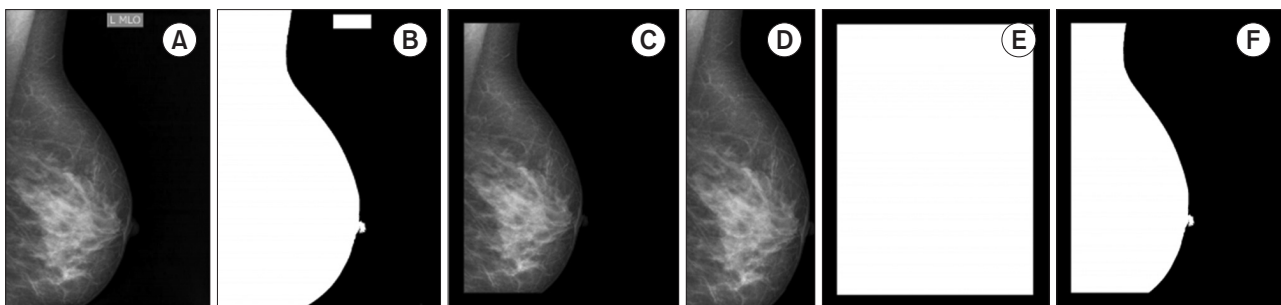


Figure 1. Artefact removal steps. (A) Original image (I). (B) Binary image (B). (C) Binary frame (F). (D) After applying $B \otimes F$ and retaining the largest object (B_f). (E) After applying $I \otimes B_f$. (F) Preprocessed image (J).

circumscribes the breast (the bounding box of B_p), resulting in the preprocessed image J , as illustrated in Figure 1F.

2) Pectoral muscle removal

A four-step method is applied to remove the pectoral region. In the first step, the pectoral muscle is localized by dividing the preprocessed image J (Figure 2A) into two images, J_L and J_R , with non-zero pixels being counted. If the number of non-zero pixels is bigger in the J_L portion, then the image is left-oriented; otherwise, the image is right-oriented. Next, all images are flipped to left (J_F) (Figure 2B). In the second step, the background is cleared; the idea behind this step is to create a background copy to simulate the original background and subtract it, such that all structures different from the pectoral muscle are removed. To do so, a background estimation (BG) is implemented starting from the superposition of the vertical background (BG_v) and horizontal background (BG_h) images. Each of these images is obtained by blurring the $MPIV = \max(J_F)$ maximum pixel intensity value in the respective directions (Figure 2C). After that, the background of the image is removed using the expression $J_F - BG$, resulting in image P (Figure 2D).

In the third step, segmentation of the pectoral muscle is performed using the region-growing algorithm with two parameters, namely the seed and threshold. The seed for all images is set to (10, 10) pixel coordinates. This ensures that the segmented region grows while covering the entire pectoral muscle. The optimal threshold T is searched iteratively considering the fact that the pectoral muscle is a homogeneous region in terms of having similar pixel values. This method for calculating the threshold is one of the main contributions of this work. The uniform intensity value index (UIV) is calculated as [12]

$$UIV = 100 \left(1 - \frac{\sigma_p}{\mu_p} \right)$$

where μ_p and σ_p denote the mean and standard deviation of image P without considering the zero pixel values, respec-

tively. This guaranteed that the UIV is calculated for the pectoral region. The threshold searching procedure starts with initializing $T = \mu_p$ and updates this value by subtracting 0.01 in each iteration, always getting a different region as a result of segmentation. The region-growing algorithm is run with each new T , and the difference between the new (UIV_{new}) and the old (UIV_{old}) is computed in each iteration. The process finishes when the image difference in uniformity remains than 2% ($|UIV_{old} - UIV_{new}| < 2\%$). The decrement threshold value (0.01), minimum uniformity value (70%), and maximum difference in uniformity value (2%) were chosen experimentally. The output of the algorithm is pectoral muscle segmentation (B_p). All residual holes that may have been left from segmentation are filled in by applying the morphological operation of dilatation with a square structuring element of size 5×5 over B_p (Figure 3A).

In the final step, pectoral muscle curve fitting is performed by obtaining the perimeter of B_p and deleting the vertical and horizontal components using the erode operation with an equal line structuring element of length 10 and orientations of 90° and 0° , respectively. This procedure removes the limit lines of the considered image and preserves only the contour that separates the pectoral muscle from the rest of the breast. Next, an approximation of the pectoral muscle perimeter is

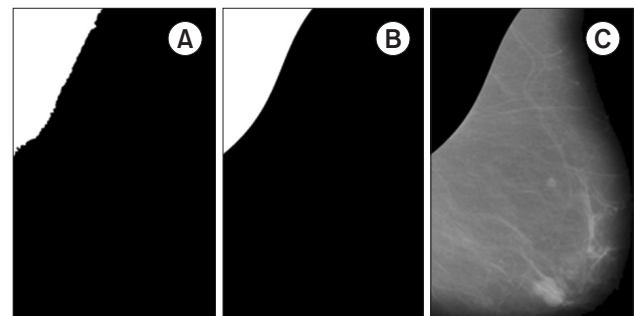


Figure 3. Steps 3 and 4 for pectoral muscle suppression. (A) Region-growing segmentation after perimeter fitting (B_p). (B) Pectoral muscle segmentation (B_{F_p}). (C) After applying $J_F \circ B_p$, suppression of the pectoral muscle (I_p).

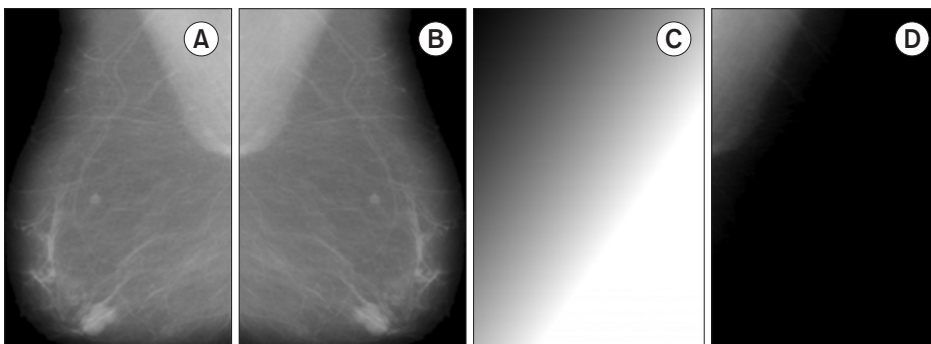


Figure 2. Steps 1 and 2 for pectoral muscle removal. (A) Artefact removal from the image (J). (B) Image flipping (J_p). (C) Background estimation (BG); in this case, $MPIV = 251$. (D) After applying ($J_F - BG$) to remove structures different from the pectoral muscle (P).

constructed using a third-order polynomial model over the edge of the considered image (Figure 3B). Finally, the new fitting \mathbf{BF}_p mask is applied over \mathbf{J}_F to suppress the pectoral muscle and obtain the preprocessed mammogram image (\mathbf{I}_p) (Figure 3C).

The perimeter of the pectoral muscle (Figure 4) was approximated using a third-order polynomial model over the edge of the image \mathbf{B}_p . The cubic polynomial function was chosen considering the shape of the pectoral muscle.

3. MCC Localization

The detection of MCCs remains a major challenge due to the large variety of their widths, contrasts, lengths, and intersections, in addition to backgrounds with similar structures to that of MCCs and similar noise and signal levels. Furthermore, linear structures in mammograms such as ducts, blood vessels, and Cooper's ligaments often produce a characteristically textured appearance that may contribute to a high level of FPs.

1) Potential ROI classification

The proposed method for identifying MCCs in mammograms consists of two steps. The first step is the classification of ROIs with the aim of identifying MCCs, while the second step is MCC contrast enhancement for localization. In the first step, a CNN model was trained on a dataset manually constructed using the mini-MIAS and UTP datasets. The dataset included two classes, with "non-MCC" including normal ROIs (2,360 images) and "MCC" corresponding to MCC ROIs (1,932 images). Out of total 4,292 ROI images, 3,500 images (80%) were used for training the model and 792 images (20%) were used for its evaluation. The ROI window size was set to 101×101 pixels; this size is suitable for identifying MCCs without affecting their geometric distortion effects caused by pixelation. The CNN architecture

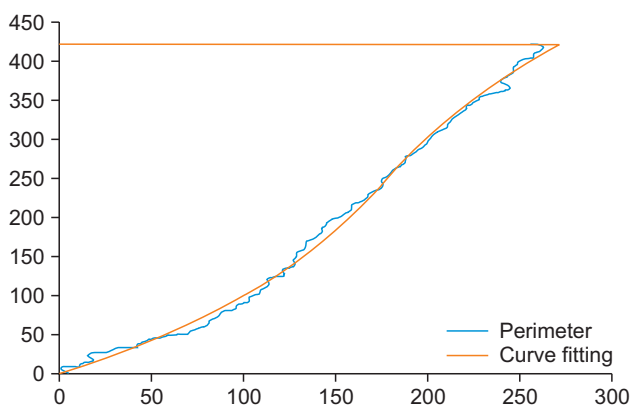


Figure 4. Third-order polynomial pectoral contour fitting.

comprised an input layer of size 101×101 ; seven hidden layers of sizes 3×3 , with 8, 16, 32, 64, 128, 256, and 512 filters, respectively, each having a max-pooling layer of size 2×2 for extracting large-scale features; an activation layer with the rectified linear unit function for fast training; and an output layer with the softmax function for outputting the probabilistic class interpretation. The model was trained using the stochastic gradient descent algorithm with the momentum optimizer, an initial learning rate of 0.01, 10 epochs, and 25 iterations. The architecture parameters were selected experimentally. The model has the objective of classifying whether a specific ROI in a mammogram contains MCCs or not, in order to obtain a set of possible MCC candidates. To do so, the mammogram is divided into regions of size 101×101 pixels and the system evaluates each region automatically in the CNN classifier.

2) MCC enhancement and FP reduction

The second step is applied after obtaining ROIs containing MCCs. The basic idea is to enhance the contrast of MCCs to improve the accuracy of segmentation. To achieve this, each ROI is processed in three stages. The background is removed in the first stage, irrelevant structures are removed in the second stage, and MCCs are identified in the third stage. First, a morphological operation is used to remove the background. Next, wavelet reconstruction from approximation coefficients is performed to enhance the appearance of MCCs. Finally, segmentation is implemented using the binarization method. For background removal, a morphological transformation opening \circ is applied to the original ROI (\mathbf{I}_{ROI}) using a structuring element S to determine the information corresponding to the background of the considered image. This information is then subtracted from the original image: $\mathbf{I}_{ROI} - (\mathbf{I}_{ROI} \circ S)$, where S denotes a disk with a radius r (Figure 5B). The reconstruction of \mathbf{I}_{RBG} from the approximation wavelet coefficients W_ϕ is performed in the second stage to increase the contrast of MCCs by increasing the intensity of the regions of their location [28]. Irrelevant structures such as tissues are filtered out using the following transformation:

$$\begin{aligned} W_\phi &= \Sigma \mathbf{I}_{RBG} \phi \\ \mathbf{I}'_{RBG} &= \Sigma W_\phi \phi \end{aligned}$$

The wavelet (ϕ) biorthogonal (bior3.5) is used in the decomposition at level 4, and the reconstruction is performed at level zero to maintain the resolution of the original image (Figure 5C). The binarization of \mathbf{I}'_{RBG} is performed in the third stage using the Otsu thresholding technique, given the number of objects present in \mathbf{I}_{ROI} .

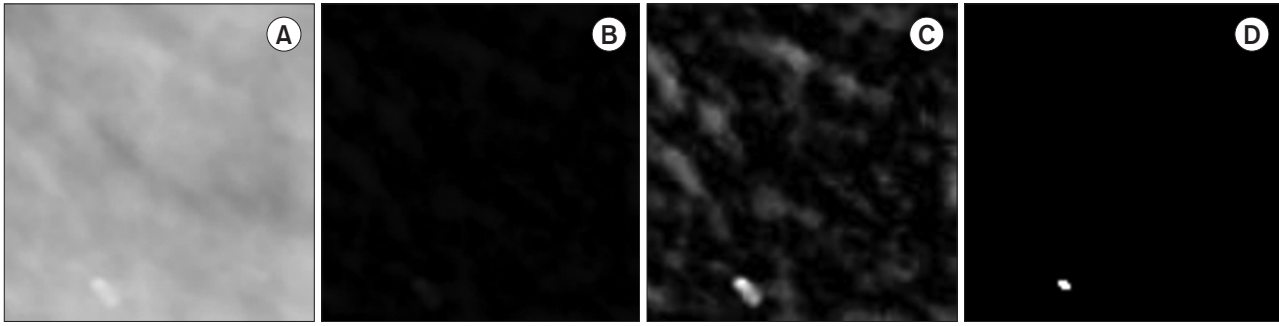


Figure 5. Microcalcification enhancement. (A) Original ROI I_{ROI} . (B) After subtracting the background (I_{RBG}). (C) After wavelet enhancement I'_{RBG} . (D) After binarizing I''_{RBG} .

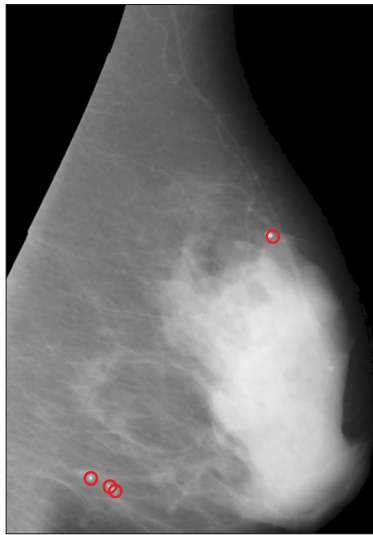


Figure 6. Microcalcification detection result.

The assumption that the diameter of MCCs varies from 0.1 mm to 1 mm [29] was used to reduce the number of FPs. This means that objects linear in shape and greater than 1 mm are discarded based on the length of their major axis. Furthermore, MCCs tend to appear as tiny circular bright spots; thus, an intensity-based feature is used to distinguish true MCCs. In summary, MCC candidates satisfying one or more of the following conditions are removed [24]:

$$\begin{aligned} I_{obj} &> I_{max} \\ A_{obj} &> A_{max} \\ A_{obj} &< A_{min} \\ I_{obj} &< I_{min} \end{aligned}$$

where I_{max} , A_{max} , A_{min} , and I_{min} denote the maximum length of the major axis, maximum area, minimum area, and mean intensity of the selected objects, respectively. The value of I_{max} is defined as $\lceil 1/res \rceil$, where res denotes the resolution of the image in millimeters; A_{max} is defined as $\lceil \pi(I_{max}/2)^2 \rceil$;

Table 1. Results of the proposed noise removal and radiopaque artefact suppression method compared to those of other methods reported in the literature (unit: %)

Study	Dataset	
	UTP	Mini-MIAS
Proposed method	97.25	99.69
Qayyum and Basit [18]	-	99.37
Slavkovic-Ilic et al. [19]	-	97.51
Yoon et al. [20]	-	93.16

and the values of $A_{min} = 3$ and $I_{min} = 190$ were selected experimentally. The centroids of all objects are subtracted from the resulting binary image to obtain the coordinates of MCCs in mammograms (Figure 6).

III. Results

1. Mammogram Preprocessing

The proposed algorithm for noise removal and radiopaque artefact suppression was applied to the CC and MLO views of mini-MIAS and UTP datasets, using a total of 832 mammograms. Segmentation accuracy was obtained by visual inspection; it was categorized as “successful” if the mammogram did not contain artefacts after the algorithm was applied, and “unacceptable” if artefacts remained. The success rate was 99.69% and 97.25% with the mini-MIAS and UTP datasets, respectively (Table 1). The unacceptable cases included mammograms where artefacts and the breast overlapped.

The proposed algorithm for pectoral muscle removal was applied to MLO mammogram views. Segmented accuracy was obtained by visual inspection; it was categorized as “successful” if the pectoral muscle was considered to be well-segmented, and “unacceptable” otherwise. Furthermore, the mean rates of FPs and FNs were calculated for each dataset.

Table 2. Results of the proposed pectoral muscle removal method compared to those of other methods reported in the literature

Study	Dataset ^a	Accuracy (%)	Other performance measures (%)
Our method	Mini-MIAS (282)	91.92	7.01 (FP) – 11.34 (FN)
	UTP (159)	95.12	2.70 (FP) – 14.95 (FN)
Shinde and Rao [23]	Mini-MIAS	93.70	Not mentioned
Abdellatif et al. [22]	Mini-MIAS (80)	Not mentioned	1.20 (FP) – 20.4 (FN)
Qayyum and Basit [18]	Mini-MIAS	93.00	Not mentioned
Slavkovic-Ilic et al. [19]	Mini-MIAS	87.57	Not mentioned
Camilus et al. [21]	Mini-MIAS (84)	Not mentioned	0.64 (FP) – 5.58 (FN)

FP: false positive, FN: false negative.

^aThe numbers in parenthesis indicate the number of subset images.

Table 3. Confusion matrix results for region of interest classification

		Predicted		Total
		Non-MCC	MCC	
Actual	Non-MCC	590 (98)	20 (2)	610
	MCC	13 (7)	169 (93)	182
Total		603	189	792

Values are presented as number (%).

MCC: microcalcification.

The success rate was 91.92% and 95.12% for the mini-MIAS and UTP datasets, respectively. The unacceptable cases, as well as FPs and FNs, were due to an insignificant difference between the pixel intensity values of the breast and pectoral muscle since the algorithm considers pectoral region homogeneity and requires a well-defined pectoral boundary in terms of pixel intensity values (Table 2).

2. MCC Detection

The accuracy score for the training set was 96.26%, while for the validation set (792 images), it was 95.83% (Table 3). The classification results demonstrate that the CNN model was effective in rejecting ROIs that did not contain MCCs, with 20 FP images. This is due to the capability of convolutional layers to extract patterns, such edges and linear structures, thus allowing discrimination of MCC from non-MCC images.

Table 4 lists the results achieved by the proposed MCC detection method. A total of 400 mammograms were used in the evaluation. The results indicate that the proposed MCC enhancement step based on morphological operations and wavelet reconstruction in conjunction with the FP reduction step based on geometrical MCC assumptions achieved relatively good performance (with a sensitivity of 80% and an FP rate of 2 per image) on different datasets. This is remarkable

Table 4. Sensitivity and FP rate per image for microcalcification enhancement and localization

Study	Dataset ^a	Sensitivity (%)	FP per image (%)
Proposed method	Mini-MIAS (260)	78	0.28
	UTP (140)	82	2.33
Wang and Yang [27]	Private (292)	85	0.13
Wang et al. [26]	Private (292)	90	0.24
Liu et al. [25]	Private (205)	92	1.12
El-Neqa et al. [24]	Private (76)	94	1.31

FP: false-positive.

^aThe numbers in parenthesis indicate the number of subset images.

because MCC-based detectors are more likely to be susceptible to FPs caused by MCC-like noise and image patterns such as tubular structures. Nonetheless, our method was able to handle this issue using two different datasets.

IV. Discussion

This paper presents a method for MCC detection from raw mammograms. The proposed method includes two modules, one for raw mammogram preprocessing and another for MCC detection. In the first module, noise and artifacts are removed and segmentation of the breast region is achieved. In the segmentation and subsequent removal of the pectoral muscle, an iterative algorithm is implemented to search for the optimal threshold to be used in the region growing technique. This search is based on the homogeneity of the anatomical patterns of the image (pixel intensity) in the region to be segmented. Third-order polynomial adjustment is used to correct the pectoral contour. The second module conducts a search of MCC candidates in the entire image using a CNN trained to recognize MCCs. Irrelevant structures in the ROI

background are removed using morphological operations, and the MCC intensity is enhanced using the wavelet reconstruction method. Finally, the number of FP cases is reduced based on geometrical MCC assumptions. The performance of the proposed methodology was evaluated on the mini-MIAS and UTP datasets. The results demonstrated that the proposed preprocessing module outperformed state-of-the-art methods for breast segmentation and achieved relatively good results for pectoral muscle removal. We believe that this is due to the employed strategy of performing an iterative search based on the homogeneity of the ROI with the aim of finding optimal parameters for the segmentation of each pectoral muscle. Furthermore, the filtering and removal of unwanted regions from mammograms in conjunction with the CNN's pattern-recognition and abstraction properties allowed the MCC detection module to achieve the highest test accuracy in identifying potential ROIs with MCCs compared to other state-of-the-art methods.

This work makes two main contributions. The first contribution lies in the fact that the method we present is a beginning-to-end methodology; that is, we start from a raw image and deliver a mammogram in which the microcalcifications are detected and localized.

Our second contribution is way in which we segment the pectoral muscle. Specifically, the threshold with which the region growing algorithm operates is adjusted iteratively based on the fact that the pectoral region has a uniform intensity of the image pixels.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

This work was developed under the framework of the research project "Prototipo de un Sistema de recuperacion de información por contenido orientado a la clasificación de grupos de microcalcificaciones en mamografías - protocam" funded by Vice-Rectoría for Research of Universidad Tecnológica de Pereira.

ORCID

Kevin Alejandro Hernández Gómez (<https://orcid.org/0000-0003-4506-9004>)
 Julian D. Echeverry-Correa (<https://orcid.org/0000-0002-0956-6161>)
 Álvaro Ángel Orozco Gutiérrez (<https://orcid.org/0000-0002-1167-1446>)

References

1. Kwan ML, Kushi LH, Weltzien E, Maring B, Kutner SE, Fulton RS, et al. Epidemiology of breast cancer subtypes in two prospective cohort studies of breast cancer survivors. *Breast Cancer Res* 2009;11(3):R31.
2. Raba D, Oliver A, Marti J, Peracaula M, Espunya J. Breast segmentation with pectoral muscle suppression on digital mammograms. In: Marques JS, Perez de la Blanca N, Pina P, editors. *Iberian conference on pattern recognition and image analysis*. Springer, Berlin, Heidelberg, Germany: Springer; 2005. p. 471-8.
3. Linguraru MG, Marias K, English R, Brady M. A biologically inspired algorithm for microcalcification cluster detection. *Med Image Anal* 2006;10(6):850-62.
4. Elshinawy MY, Abdelmageed WW, Badawy AH, Chouikha MF. Pre-CAD system for normal mammogram detection using local binary pattern features. *Proceedings of 2010 IEEE 23rd international Symposium on Computer-Based Medical Systems (CBMS)*; 2010 Oct 12-15; Bentley, Australia. p. 352-7.
5. Karale VA, Ebenezer JP, Chakraborty J, Singh T, Sadhu A, Khandelwal N, et al. A screening CAD tool for the detection of microcalcification clusters in mammograms. *J Digit Imaging* 2019;32(5):728-45.
6. Giger ML, Karssemeijer N, Schnabel JA. Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer. *Annu Rev Biomed Eng* 2013;15:327-57.
7. Moghbel M, Ooi CY, Ismail N, Hau YW, Memari N. A review of breast boundary and pectoral muscle segmentation methods in computer-aided detection/diagnosis of breast mammography. *Artif Intell Rev* 2020;53(3):1873-918.
8. Ancy CA, Nair LS. Tumour classification in graph-cut segmented mammograms using GLCM features-fed SVM. In: Bhateja V, Coello Coello C, Satapathy S, Pattnaik P, editors. *Intelligent engineering informatics*. Singapore: Springer; 2018. p. 197-208.
9. Rahimeto S, Debelee TG, Yohannes D, Schwenker F. Automatic pectoral muscle removal in mammograms. *Evol Syst* 2019;12:519-26.
10. Karale VA, Singh T, Sadhu A, Khandelwal N, Mukhopadhyay S. Reduction of false positives in the screening CAD tool for microcalcification detection. *Sadhana* 2020;45:44.
11. Chan HP, Sahiner B, Lam KL, Petrick N, Helvie MA, Goodsitt MM, et al. Computerized analysis of mam-

- mographic microcalcifications in morphological and texture feature spaces. *Med Phys* 1998;25(10):2007-19.
12. Leichter I, Lederman R, Buchbinder S, Bamberger P, Novak B, Fields S. Optimizing parameters for computer-aided diagnosis of microcalcifications at mammography. *Acad Radiol* 2000;7(6):406-12.
 13. Tsujii O, Freedman MT, Mun SK. Classification of microcalcifications in digital mammograms using trend-oriented radial basis function neural network. *Pattern Recognit* 1999;32(5):891-903.
 14. Betal D, Roberts N, Whitehouse GH. Segmentation and numerical analysis of microcalcifications on mammograms using mathematical morphology. *Br J Radiol* 1997;70(837):903-17.
 15. Paquerault S, Yarusso LM, Papaioannou J, Jiang Y, Nishikawa RM. Radial gradient-based segmentation of mammographic microcalcifications: observer evaluation and effect on CAD performance. *Med Phys* 2004;31(9):2648-57.
 16. Kim JK, Park HW. Statistical textural features for detection of microcalcifications in digitized mammograms. *IEEE Trans Med Imaging* 1999;18(3):231-8.
 17. Soltanian-Zadeh H, Rafiee-Rad F. Comparison of multiwavelet, wavelet, Haralick, and shape features for microcalcification classification in mammograms. *Pattern Recognit* 2004;37(10):1973-86.
 18. Qayyum A, Basit A. Automatic breast segmentation and cancer detection via SVM in mammograms. *Proceedings of 2016 International Conference on Emerging Technologies (ICET)*; 2016 Oct 18-19; Islamabad, Pakistan. p. 1-6.
 19. Slavkovic-Ilic M, Gavrovska A, Milivojevic M, Reljin I, Reljin B. Breast region segmentation and pectoral muscle removal in mammograms. *Telfor J* 2016;8(1):50-5.
 20. Yoon WB, Oh JE, Chae EY, Kim HH, Lee SY, Kim KG. Automatic detection of pectoral muscle region for computer-aided diagnosis using MIAS mammograms. *Biomed Res Int* 2016;2016:5967580.
 21. Camilus KS, Govindan VK, Sathidevi PS. Computer-aided identification of the pectoral muscle in digitized mammograms. *J Digit Imaging* 2010;23(5):562-80.
 22. Abdellatif H, Taha TE, Zahran OF, Al-Nauimy W, Abd El-Samie FE. K2. Automatic pectoral muscle boundary detection in mammograms using eigenvectors segmentation. *Proceedings of 2012 29th National Radio Science Conference (NRSC)*; 2012 Apr 10-12; Cairo, Egypt. p. 633-40.
 23. Shinde V, Rao BT. Novel approach to segment the pectoral muscle in the mammograms. In: Mallick P, Balas V, Bhoi A, Zobaa A, editors. *Cognitive informatics and soft computing*. Singapore: Springer; 2019. p. 227-37.
 24. El-Naqa I, Yang Y, Wernick MN, Galatsanos NP, Nishikawa RM. A support vector machine approach for detection of microcalcifications. *IEEE Trans Med Imaging* 2002;21(12):1552-63.
 25. Liu X, Mei M, Liu J, Hu W. Microcalcification detection in full-field digital mammograms with PFCM clustering and weighted SVM-based method. *EURASIP J Adv Signal Process* 2015;2015:73.
 26. Wang J, Nishikawa RM, Yang Y. Global detection approach for clustered microcalcifications in mammograms using a deep learning network. *J Med Imaging (Bellingham)* 2017;4(2):024501.
 27. Wang J, Yang Y. A context-sensitive deep learning approach for microcalcification detection in mammograms. *Pattern Recognit* 2018;78:12-22.
 28. Alam N, Oliver A, Denton ER, Zwigelaar R. Automatic segmentation of microcalcification clusters. In: Nixon M, Mahmoodi S, Zwigelaar R, editors. *Medical image understanding and analysis*. Cham, Switzerland: Springer; 2018. p. 251-61.
 29. McCann J, Stockton D, Godward S. Impact of false-positive mammography on subsequent screening attendance and risk of cancer. *Breast Cancer Res* 2002;4(5):R11.