Bulletin of Electrical Engineering and Informatics
Vol. 11, No. 1, February 2022, pp. 158~167
ISSN: 2302-9285, DOI: 10.11591/eei.v11i1.3458

Breast cancer segmentation using K-means clustering and optimized region-growing technique

Srwa Hasan Abdulla[1], Ali Makki Sagheer[2,3], Hadi Veisi[4]
[1]Department of Basic Science, University of Sulaimani, Sulaymaniyah, Iraq
[2]Department of Computer Science and Information Technology, Al-Qalam University College, Kirkuk, Iraq
[3]College of Computer Science and Information Technology, University of Anbar, Ramadi, Iraq
[4]Department of Computer Engineering, University of Tehran, Enghelab Square ،16 Azar St, Iran

Article Info

Article history:

Received Jul 28, 2021
Revised Sep 8, 2021
Accepted Dec 16, 2021

Keywords:

Image segmentation
K-means clustering
Mammography
Micro-calcification
Region-growing

ABSTRACT

Breast cancer is one of the major causes of death among women, and early detection may decrease the aggressiveness of the disease. The goal of this paper is to create an automated system that can classify digital mammogram images into benign and malignant. This paper presents a new detection technique of micro-calcifications in mammogram images. An automated technique for identifying breast microcalcifications (MCs) proposed utilizing two-level segmentation processes, first crop the breast area from the image using k-means clustering, then, an optimized region growing (ORG) approach has been used, where multi-seed points and thresholds are generated optimally depending on the color values of the image pixels. Then the texture features are extracted based on Haralick definitions of texture analysis. In addition, three features (cross-correlation coefficient, pearson correlation, and average area of segmented spots) are obtained from the segmented image. Support vector machine (SVM) classifier evaluate the efficiency of the system utilizing the images from the digital database for screening mammography (DDSM) dataset. The results were obtained by utilizing 355 images for training and 85 images for testing. The proposed system's sensitivity reached up to 97.05%, the specificity obtained is 98.52%, and accuracy is 98.2%.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Srwa Hasan Abdulla
Department of Basic Science, University of Sulaimani
Kirkuk Road, Sulaimani, Kurdistan Region, Iraq
Email: sirwa.abdulla@univsul.edu.iq

1. INTRODUCTION

One of the leading causes of female deaths worldwide is breast cancer. It has caused more deaths than any other illness, such as malaria or tuberculosis. The cancer research agencies of the world health organization (WHO) (i.e. the international agency for cancer research (IARC) and the American cancer society) announce that 17.1 million new cancer cases were registered worldwide in 2018 [1]. It is anticipated that an estimated 276,480 new cases of invasive breast cancer will be diagnosed in women in the United States by 2020, along with 48,530 new cases of non-invasive breast cancer [2].

In developed and emerging countries, citizens shift their lifestyle from traditional to modern, increasing the incidence of breast cancers among women mainly between the ages of 35-55. It is possible to monitor the incidence of breast cancers by detecting breast cancers in their early stages [3]. Self and clinical breast checks, magnetic resonance imaging (MRI), ultrasound, and mammography [4] are screening techniques used for breast cancer screening. The image produced through mammography is called a

mammogram, which consists of the background, breast region, fat tissue, breast masses, and microcalcifications with high intensities [5]. As the demand for mammography processing is growing, radiologists can produce errors that overlook important clues due to fatigue [6].

Microcalcifications (MCs) are deposits of calcium in a mammogram that appear as tiny bright spots. Due to MCs and Masses are looking close to the context in the mammogram, the identification and classification are difficult. Since image processing methods play a significant role in the earlier diagnosis of MCs. Researchers have developed many strategies to determine the precise position of MCS and Masses [6].

This paper introduces a useful pixel-based technique for region growth. Due to the nature of the calcifications which can be sporadic, multi-points have been utilized to detect calcifications in more than one region of the breast, which cannot be determined by using the standard region growth algorithm, on the other hand, need an initial seed point which increases computing cost and execution time.

The structure of this paper is as shown in: literature review presented in section 2. Section 3 provides a detailed overview of the proposed methodology for data collection, pre-processing, denoising, segmentation, extraction of features, and classification. The performance analysis that demonstrates the segmentation and classification efficiency of the suggested system is given in section 4. The conclusion of this work is obtained in section 5.

2. LITERATURE REVIEW

According to Rouhi *et al*. [7], two approaches for breast mass segmentation were proposed using the growing region. The texture features were collected and utilized a neural classifier in order to differentiate malignant and benign mammograms. 96.87%, 95.94%, and 96.47% are the acquired rates of sensitivity, accuracy, and precision respectively. Sambandam and Jayaraman [8] suggested a self-adaptive segmentation technique for multilevel thresholding based on dragonfly optimization, where optimum thresholds are created utilizing a swarm optimization approach.

Alam *et al*. [9] proposed method that uses the first wavelet-based algorithm to boost the region of interest, followed by morphological operations and interpolation methods of image segmentation. Then sub-regions were obtained by splitting the initial image and bicubic interpolation was used to obtain the intensity level of the local history. Finally, the difference image is obtained by subtracting an interpolated image from the original image, and area-ranking technique is used to cluster microcalcifications. Liu and Zeng [10] suggested an optimized region-growing methodology called multiple concentric layers (MCL) in order to increase accuracy and achieved a sensitivity of 82.4% when tested on a collection of 164 mammograms.

Anitha and Peter [11] suggested a method that are eliminated the noise and objects using Weiner filtering, followed by background separation by morphological operations, global thresholding, kernel-based level set, and fuzzy clustering to segment the mass after the region of interest is defined. Ibeni *et al*. [12] provides a full Bayesian method to evaluating the predictive distribution of all classes using three classifiers: Naive bayes (NB), bayesian networks (BN), and tree augmented Naive Bayes (TAN) with three datasets: breast cancer, breast cancer Wisconsin, and breast tissue dataset. After that, the prediction accuracy of Bayesian methods is compared to three common machine-learning algorithms: K-nearest neighbor (K-NN), support vector machine (SVM), and decision tree (DT). The findings indicated that the Bayesian networks (BN) algorithm had the best performance, with an accuracy of 97.281%.

According to Touil *et al*. [13] suggested a new conditional region growth (CRG) method for determining correct MC bounds beginning from a set of seed points. Regional maxima detection and superpixel analysis are used to find the beginning seed points. In terms of contrast and shape variation, the region growing step is governed by a set of criteria adapted to MC detection. These criteria are separated into two groups and are generated from prior knowledge to characterize MCs. The first one is about the size of the search neighborhood. The second one is about analyzing gradient information and shape evolution during the growing process.

3. PROPOSED METHOD

The methodology proposed involves multiple steps including image processing techniques. The first step is the image acquisition from the dataset collected in the digital database for screening mammography (DDSM) where regular and irregular mammograms are collected. The optical mammograms are then pre-processed by Gaussian filters for noise reduction. The images are further processed using the proposed multi-points (seeds) region growing method to extract the region of interest (ROI), which targets breast MCs. For feature extraction, the ROIs are then processed where a collection of texture features are extracted using Haralick texture characteristics. Then the extracted textures are further fed into the SVM classifier. The diagrammatic illustration of the suggested procedure is shown in Figure 1.

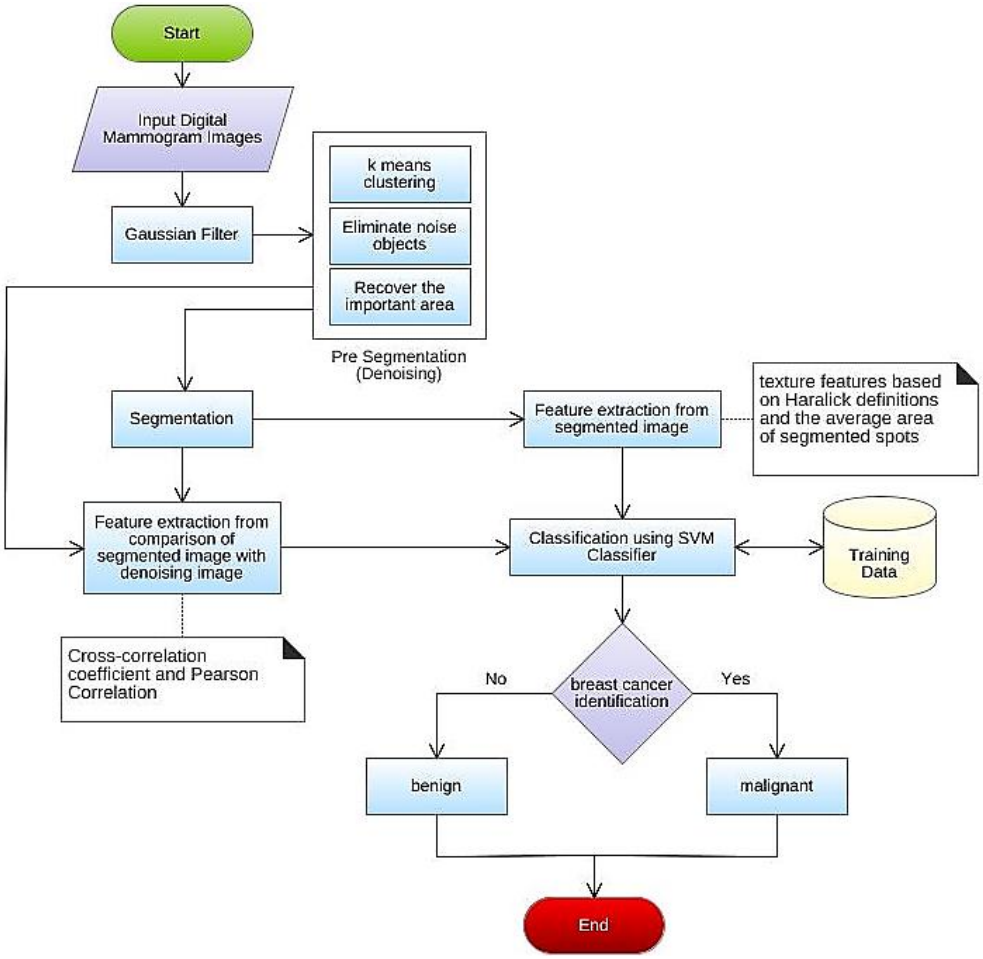


Figure 1. The workflow of the proposed algorithm

3.1. Data acquisition

The curated breast imaging subset of DDSM (CBIS-DDSM) dataset are available at [14], which are the digital images used in the proposed technique. In this work, a series of 440 digital mammograms of breast microcalcification containing benign and malignant severity was used in cranio-caudal (CC) views for training and testing purposes. This collection comprises 329 benign cases, 111 malignancies cases in both left and right breast.

3.2. Pre-processing

In any image processing technique, pre-processing is regarded as the fundamental step. The ultimate objective of this technique is to improve the image quality and the image characteristics that are necessary for further processing. Mammogram images are difficult to view compared with other medical images, so pre-processing is important [15]. The proposed method utilizes a Gaussian filter to preprocess the digital mammograms where the noises are removed and the images softened out. Figure 2(a) is the original image and Figure 2(b) shows the effect of applying a Gaussian filter to the image.

3.3. Pre-segmentation (denoising)

Usually, medical images contain some symbols, words, or letters that show the type or some of the medical-physical characteristics of the image. This is generally considered image noise and may affect classification accuracy. To overcome these problems the denoising is used. This process goes through three stages (K-means clustering, eliminate noise objects, recover the important area):

3.3.1. K-means clustering

K-means clustering is a method of grouping or partitioning a pattern into multiple clusters such that similar patterns are allocated to the same cluster. Clustering is used in many forms of analysis to blot out the

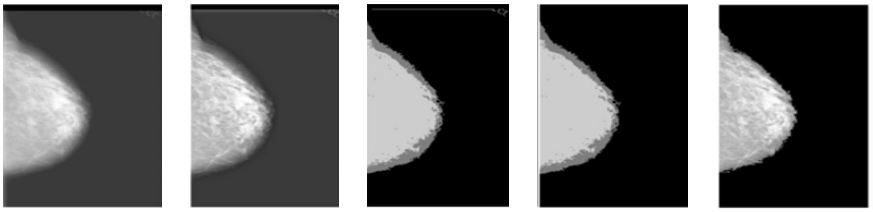
field of image segmentation. The K-means clustering is an unsupervised algorithm, and it is one of the most widely used techniques [16]. At this stage, the image is divided into a group of converged areas in color intensity, and as a result, the background of the image will be isolated from its components more clearly. Also, the noise in the images will be isolated from the rest of the image components, which facilitates the process of cutting them later. Figure 2(c) illustrates the process of applying K-means clustering.

3.3.2. Eliminate noise objects

Morphological operation is rearranging the order of pixel values. It operates on structuring elements and input images, structuring elements are attributes that probe features of interest. Erosion is an essential operation used here, during erosion, the rock bottom value is chosen by comparing all the pixel values in the region of input image [17]. After dividing the image in the previous step into specific areas of converging intensity in color and isolating them from the background, the extra foreign objects that cause noise in the image are cut off. Thus, we have an image that contains only the breast area without any noise as shown in Figure 2(d), this area will be used to restore the breast area from the original image, which we call the breast area mask.

3.3.3. Recover the important area

It is the last step in this stage, where the mask that was produced in the previous step is relied on and applied to the original image to restore the equivalent area of the mask, and neglect the components of the image and consider it as background for the area resulting from the retrieval process. Figure 2(e) demonstrates the area of the resected breast.

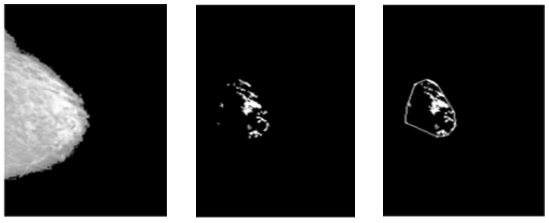


(a) (b) (c) (d) (e)

Figure 2. The pre-segmentation steps; (a) original image, (b) Gaussian filter, (c) applying k-means, (d) erosion filter, and (e) breast area retrieval

3.4. Segmentation

The method of segmentation is distinguishing the benign and malignant area by splitting the digital mammograms through non-overlapping segments from the background portions [2]. The region-based strategies find a seed point and growing regions until a criterion of homogeneity is reached [18]. This paper presents an effective variant of the region growing pixel-based technique that produces optimal seeds and thresholds. Due to the nature of the calcifications, which may be sporadic, multi-points (seeds) have been used to determine the calcifications if they are in more than one area of the breast, which cannot be determined in the case of using the traditional region growth algorithm. Figure 3(a) shows the breast area retrieval, Figure 3(b) is the segmented area, and Figure 3(c) shows the ROI.



(a) (b) (c)

Figure 3. The segmentation process, (a) breast area retrieval, (b) segmented area, and (c) ROI

The following presents the main steps for generating optimal seeds using multi-points region-growing segmentation:

a. Input digital mammogram image.
b. Let mammogram values is a list of mammogram pixel values (It will contain all the pixel values in the image without duplicating).
c. Sort mammogram Values in descending order.
d. Determine the segmentation threshold, which will determine the approved values from the list and which will be adopted in the segmentation process. This is done by dividing the sorted Mammogram Values into ten sections and adopting the first section of it (higher values), which will represent the list of segmentation values or seed points.
e. For each pixel in the Mammogram image, if the pixel value belongs to the Segmentation Values, then the pixel will be dimming, otherwise, the pixel will be discarded.

The algorithm for the multi-points region-growing method is presented in Algorithm 1:

Algorithm 1: Optimized Region Growing

```
Input: Digital Mammogram Image (M).
Output: Segmented Image (S).
Begin
  W = Digital Mammogram Image width
  H = Digital Mammogram Image height
  MammValues<>: list of Mammogram pixel values
  SegmentationValues<>: list of segmentation threshold values
  Aggregation (M)
  For i ← 1 to W do
    For j ← 1 to H do
      If (! MammValues.Contains(M(i,j))
        MammValues.Add(M(i,j))
      End
    End
  End
  Sort (MammValues)
  Determine thresholds (MammValues)
  For i ← 1 to (MammValues.Count / 10) do
    SegmentationValues.Add(MammValues[i])
  End
  Segmentation (M, SegmentationValues)
  For i ← 1 to W do
    For j ← 1 to H do
      For t ← 1 to SegmentationValues.Count do
        If (GetPixelValue(M(i,j)) == SegmentationValues[t])
          S.SetPixel(i, j, Color.White)
        Else
          S.SetPixel(i, j, Color.Black)
        End
      End
    End
  End
  Return S
End
```

3.5. Features extraction

Features of the image show the current attributes and characteristics. The extracted features utilized for classification should also be identifiable, effective, and autonomous [19], [20]. In the first step of features extraction, statistical textural analysis-features (cross-correlation coefficient and pearson correlation) information from the comparison of the original image with the segmented image intensities extracted.

– Cross-correlation coefficient: It is a measure of similarity of two series as a function of the displacement of one relative to the other. The cross-correlation coefficients are more robust to changes of illumination than the mean square error (MSE) [1].

$$\text{Cross-correlation coefficient}=\frac{\sum_k(x_k-\bar{x})(y_k-\bar{y})}{\sqrt{\sum_k(x_k-\bar{x})^2}} \quad (1)$$

– Pearson correlation coefficient: evaluates if there is statistical support for a linear relationship, represented by a population correlation coefficient, between the same pairs of variables in the population. A parametric calculation is the pearson correlation [21].

$$\text{Pearson correlation coefficient}=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}} \quad (2)$$

Then, the average area of segmented spots was obtained. In this extraction process, the average area of segmented spots is calculated which represents the average of the infected areas. Suppose that B={B1, B2, ..., BN} is the set of segmented blobs, where N is the number of the blob segmented from the mammograms.

$$\text{Average Area}=\frac{\sum_{i=1}^{N}\text{Area(Bi)}}{N} \quad (3)$$

In the second step of features extraction, the proposed technique utilizes a collection of texture features based on Haralick's texture analysis concepts [22]. Where twenty-six texture features (angular second moment, contrast, correlation, variance, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, first information measure, second information measure, and invariance was achieved for each of these statistics by averaging them over the four directional co-occurrence matrices) are defined by the proposed methodology.

Thus, for training the SVM classifier, a collection of twenty-nine features was extracted (three features from the first step and twenty-six features from the second step). In this paragraph, a sample is shown of the obtained results. It displays a comparison of some extracted features of the segmented images. Figures 4(a), (b), (c), and (d) shows the distance between features values of malignant and benign samples (angular second moment, variance, first information measure, and cross-correlation coefficient) respectively for a group of 30 segmented samples, and illustrated the extracted features that can be used to distinguish and observed the values of the malignant tissue differed from the values of the benign tissue.

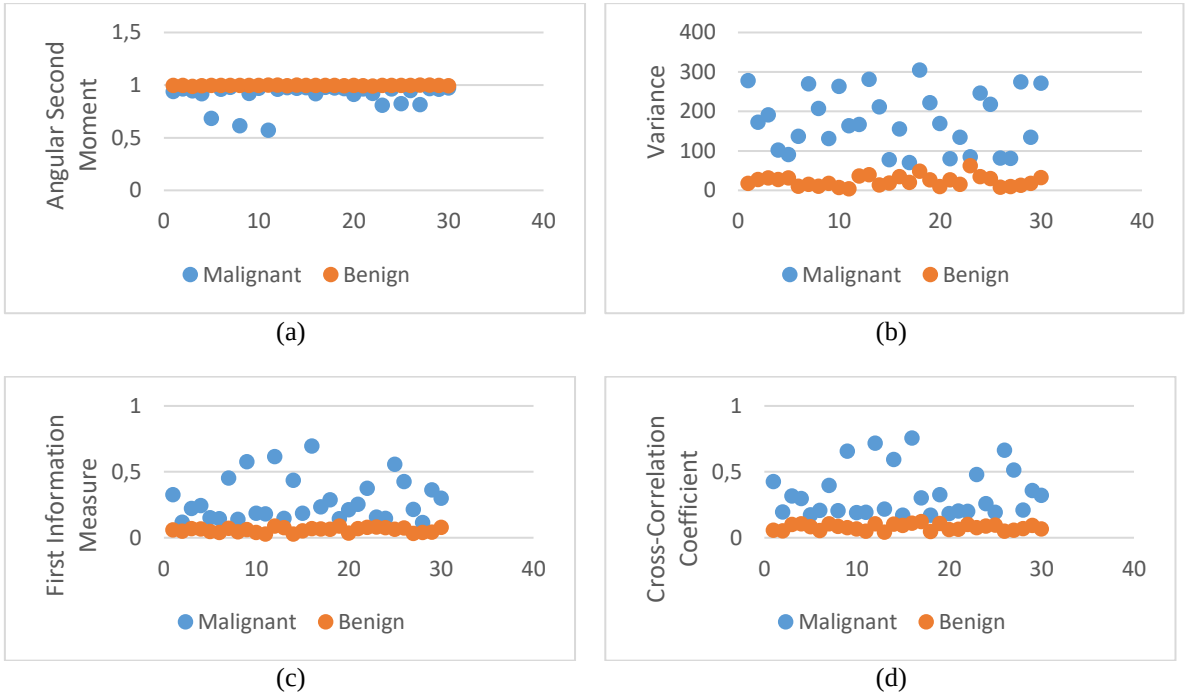


Figure 4. Distance between features values of malignant and benign samples; (a) distance between (angular second moment) values of malignant and benign samples, (b) distance between (variance) feature values of malignant and benign samples, (c) distance between (first information measure) feature values of malignant and benign samples, and (d) distance between (cross-correlation coefficient) feature values of malignant and benign samples

3.6. Support vector machine classifier

SVM is one of the best classifiers and most used as a classification algorithm. The working principle support vector machine is mainly based on marginal calculations [23], [24]. In this paper, the Gaussian kernel

function is used for transformation. Nonlinear samples are converted into a high-dimensional feature space by using the kernel function, where it may be possible to isolate nonlinear samples or data to make the classification convenient [25].

4. RESULTS AND DISCUSSION

For the calculation of accuracy, sensitivity, and specificity, the confusion matrix distinguishes the terms true positive (TP), true negative (TN), false positive (FP), and false negative (FN) from the predicted and ground truth result. Accuracy, sensitivity, specificity are calculated utilizing as shown in (4)-(5), measure the efficiency of the proposed method outlined in the paper. 440 images are used to test the classification's efficiency by using the (ORG) in the segmentation stage, where 329 benign cases and 111 malignant cases in CC view are used. The images were selected randomly from the CBIS-DDSM dataset are divided in advance into training and testing. For training, 355 images are used and 85 images are utilized to assess the proposed method. The quality of classification can be determined as [26]:

– Sensitivity: is a test that decides the chances of outcomes that are correctly identified when the cancer is present.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (4)$$

– Specificity: is a test that determines the probability of the outcomes that are true negative which are correctly identified.

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (5)$$

– Accuracy: is a test that determines the probability that how many samples are correctly identified.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (6)$$

In the classification process, these evaluations are expressed in terms of various parameters [26], the TP, TN, FP, and FN.

– TP: positive instance classified as positive.
– TN: negative instance classified as negative.
– FP: negative instance classified as positive.
– FN: positive instance classified as negative.

The proposed method was validated using SVM for desired results. Concretely, the proposed method achieved significant accuracy in classifying the CBIS-DDSM dataset. It achieved good results with the proposed ORG and SVM by showing 97.05% sensitivity, 98.52% specificity, and 98.2% accuracy in identifying both the benign and malignant samples. The ROC curve area of the proposed method was validated, as shown in Figure 5.

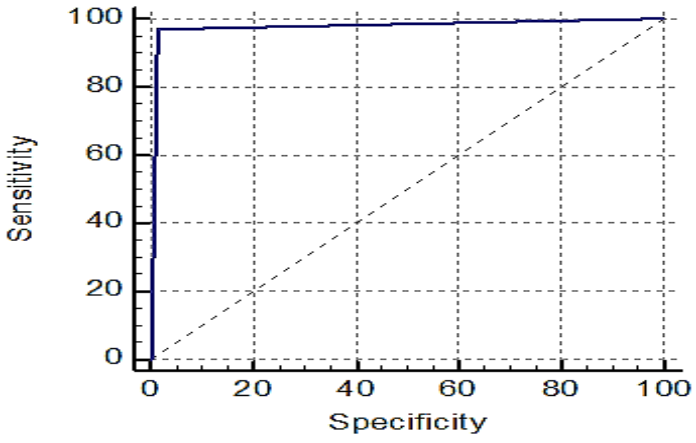


Figure 5. ROC curve of the classification results

The proposed system was implemented by using Visual Studio.Net framework 2017-C# developer, which was implemented on the Windows 10-64 bit OS, Core i7 processor, and 8GB RAM. The proposed system processes the images of the left and right breast in Cranio-Caudal (CC) views obtained from the DDSM dataset. The proposed system first utilized Gaussian filtering to pre-process the mammogram images, then applying the denoising step that consists of isolating the image background from its components using K-means clustering, eliminating noise objects, and breast area recovery. Then the Optimized Region Growing used to segment ROI includes the MCs. The work deal with the extraction of segmented area features to detect and distinguish medical digital mammogram image of benign and malignant. The experimental results of the proposed algorithm are contrasted with prior research as shown in Table 1. The obtained results, relative to the (DDSM) database, have shown that the proposed method is more reliable than other reported literature approaches and lead to the conclusion that makes it possible for clinical experts to decide and diagnose.

Table 1. Comparison of the existing technique with the proposed system

References	Segmentation	Features methods	Classifier	Dataset	Sensitivity	Specificity	Accuracy
Rouhi *et al.* [7]	Region growing optimized using GA adaptive threshold method	GLCM, contour-related, and morphological features	Random forest, NB, SVM, and KNN	MIAS and DDSM	-	-	96.47%
Setiawan *et al*. [27]	Cropping	GLCM laws texture features	FFNN	MIAS	-	-	93.90%
Kashyap *et al.* [28]	Thresholding technique and Fuzzy C-means	Shape features and Tamura	SVM using (RBFK)	Mini-MIAS	-	-	96.92%
Patel and Sinha [29]	Region growing	Shape and texture	Multilayer perceptron neural network	DDSM	-	-	95.6%
Shen *et al.* [30]	-	pixel-level annotations	CNN	DDSM	86.7%	96.1%	-
Varela *et al*. [31]	Adaptive threshold method	Gray level and contour-related features	Backpropagation neural network	Images from hospitals in Santiago de Compostela's health district, Spain.	88 %	-	-
Xie *et. al*. [32]	level set model	gray-level features and textural features	ELM and SVM	Mini MIAS+DDSM	-	-	96.02%
Punitha *et al*. [3]	Dragonfly region growing optimization	GLCM and GLRLM texture features	FFNN using backpropagation	DDSM	98.1%	97.8%	98%
Proposed system	Multi-seed points optimized region growing	Cross-correlation coefficient, pearson correlation, the average area of segmented spots and texture features based on Haralick definitions	SVM	DDSM	97.05 %	98.52 %	98.2 %

5. CONCLUSION

In this work, we presented a new optimized region growing method where multi-points (seeds) have been used to detect and segment microcalcifications (MCs) of mammographic images accurately. To detect the benign and malignant, a collection of twenty-nine features was extracted for training the SVM classifier. The precise detection of the proposed multi-points (seeds) region growing method with the SVM reached an accuracy of 98.2%. The proposed method has proven to provide a nearly accurate diagnosis of benign and malignant microcalcifications and can be used as a reference for radiologists and as a second opinion in crucial cases. Further focus should be placed on the methods of feature collection and extraction in future work. In addition, evolutionary algorithms can generate the optimized threshold generated in the growing region technology.

REFERENCES

[1] S. J. S. Gardezi, A. Elazab, B. Lei, and T. Wang, "Breast Cancer Detection and Diagnosis Using Mammographic Data: Systematic Review," *Journal of Medical Internet Research*, vol. 21, no. 7, p. e14464, Jul. 2019, doi: 10.2196/14464.

[2] "U.S. Breast Cancer Statistics," Breastcancer.org. https://www.breastcancer.org/symptoms/understand_bc/statistics (accessed Dec. 07, 2021).
[3] S. Punitha, A. Amuthan, and K. S. Joseph, "Benign and malignant breast cancer segmentation using optimized region growing technique," *Future Computing and Informatics Journal*, vol. 3, no. 2, pp. 348–358, Dec. 2018, doi: 10.1016/j.fcij.2018.10.005.
[4] J. Dheeba and S. T. Selvi, "A Swarm Optimized Neural Network System for Classification of Microcalcification in Mammograms," *Journal of Medical Systems*, vol. 36, no. 5, pp. 3051–3061, Oct. 2012, doi: 10.1007/s10916-011-9781-3.
[5] M. M. Pawar and S. N. Talbar, "Genetic Fuzzy System (GFS) based wavelet co-occurrence feature selection in mammogram classification for breast cancer diagnosis," *Perspectives in Science*, vol. 8, pp. 247–250, Sep. 2016, doi: 10.1016/j.pisc.2016.04.042.
[6] W. Qian, F. Mao, X. Sun, Y. Zhang, D. Song, and R. A. Clarke, "An improved method of region grouping for microcalcification detection in digital mammograms," *Computerized Medical Imaging and Graphics*, vol. 26, no. 6, pp. 361–368, Dec. 2002, doi: 10.1016/S0895-6111(02)00045-9.
[7] R. Rouhi, M. Jafari, S. Kasaei, and P. Keshavarzian, "Benign and malignant breast tumors classification based on region growing and CNN segmentation," *Expert Systems with Applications*, vol. 42, no. 3, pp. 990–1002, Feb. 2015, doi: 10.1016/j.eswa.2014.09.020.
[8] R. K. Sambandam and S. Jayaraman, "Self-adaptive dragonfly based optimal thresholding for multilevel segmentation of digital images," *Journal of King Saud University - Computer and Information Sciences*, vol. 30, no. 4, pp. 449–461, Oct. 2018, doi: 10.1016/j.jksuci.2016.11.002.
[9] N. Alam, A. Oliver, E. R. E. Denton, and R. Zwiggelaar, "Automatic Segmentation of Microcalcification Clusters," *Annual Conference on Medical Image Understanding and Analysis, MIUA 2018: Medical Image Understanding and Analysis*, Springer, Cham, 2018, pp. 251-261, doi: 10.1007/978-3-319-95921-4_24.
[10] X. Liu and Z. Zeng, "A new automatic mass detection method for breast cancer with false positive reduction," *Neurocomputing*, vol. 152, pp. 388–402, Mar. 2015, doi: 10.1016/j.neucom.2014.10.040.
[11] J. Anitha and J. D. Peter, "Mass segmentation in mammograms using a kernel-based fuzzy level set method," *International Journal of Biomedical Engineering and Technology*, vol. 19, no. 2, p. 133, 2015, doi: 10.1504/IJBET.2015.072933.
[12] W. N. L. W. H. Ibeni, M. Z. M. Salikon, A. Mustapha, S. A. Daud, and M. N. M. Salleh, "Comparative analysis on bayesian classification for breast cancer problem," *Bulletin of Electrical Engineering and Informatics*, vol. 8, no. 4, Dec. 2019, doi: 10.11591/eei.v8i4.1628.
[13] A. Touil, K. Kalti, P.-H. Conze, B. Solaiman, and M. A. Mahjoub, "A new conditional region growing approach for microcalcification delineation in mammograms," *Medical & Biological Engineering & Computing*, vol. 59, no. 9, pp. 1795–1814, Sep. 2021, doi: 10.1007/s11517-021-02379-x.
[14] Kirk Smith, "Cancer Imaging Archive Wiki," *Cancer Imaging Archive*. https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM (accessed Dec. 07, 2021).
[15] R. Ramani, N. S. Vanitha, and S. Valarmathy, "The Pre-Processing Techniques for Breast Cancer Detection in Mammography Images," *International Journal of Image, Graphics and Signal Processing*, vol. 5, no. 5, pp. 47–54, Apr. 2013, doi: 10.5815/ijigsp.2013.05.06.
[16] N. Dhanachandra, K. Manglem, and Y. J. Chanu, "Image Segmentation Using K -means Clustering Algorithm and Subtractive Clustering Algorithm," *Procedia Computer Science*, vol. 54, pp. 764–771, 2015, doi: 10.1016/j.procs.2015.06.090.
[17] H.-C. Lu, E.-W. Loh, and S.-C. Huang, "The Classification of Mammogram Using Convolutional Neural Network with Specific Image Preprocessing for Breast Cancer Detection," in *2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD)*, May 2019, pp. 9–12, doi: 10.1109/ICAIBD.2019.8837000.
[18] J. Dabass, S. Arora, R. Vig, and M. Hanmandlu, "Segmentation Techniques for Breast Cancer Imaging Modalities-A Review," in *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Jan. 2019, pp. 658–663, doi: 10.1109/CONFLUENCE.2019.8776937.
[19] G. Kumar and P. K. Bhatia, "A Detailed Review of Feature Extraction in Image Processing Systems," in *2014 Fourth International Conference on Advanced Computing & Communication Technologies*, Feb. 2014, pp. 5–12, doi: 10.1109/ACCT.2014.74.
[20] H. B. Mitchell, "Image Similarity Measures," in *Image Fusion*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 167–185.
[21] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson Correlation Coefficient," 2009, pp. 1–4.
[22] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973, doi: 10.1109/TSMC.1973.4309314.
[23] Z. Wang, G. Yu, Y. Kang, Y. Zhao, and Q. Qu, "Breast tumor detection in digital mammography based on extreme learning machine," *Neurocomputing*, vol. 128, pp. 175–184, Mar. 2014, doi: 10.1016/j.neucom.2013.05.053.
[24] A. Yadav, I. Jamir, R. R. Jain, and M. Sohani, "Comparative Study of Machine Learning Algorithms for Breast Cancer Prediction - A Review," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, pp. 979–985, Apr. 2019, doi: 10.32628/CSEIT1952278.
[25] J. Yao, J. Chen, and C. Chow, "Breast Tumor Analysis in Dynamic Contrast Enhanced MRI Using Texture Features and Wavelet Transform," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 1, pp. 94–100, Feb. 2009, doi: 10.1109/JSTSP.2008.2011110.
[26] C. E. Metz, "Basic principles of ROC analysis," *Seminars in Nuclear Medicine*, vol. 8, no. 4, pp. 283–298, Oct. 1978, doi: 10.1016/S0001-2998(78)80014-2.
[27] A. S. Setiawan, Elysia, J. Wesley, and Y. Purnama, "Mammogram Classification using Law's Texture Energy Measure and Neural Networks," *Procedia Computer Science*, vol. 59, pp. 92–97, 2015, doi: 10.1016/j.procs.2015.07.341.
[28] K. L. Kashyap, M. K. Bajpai, and P. Khanna, "Breast cancer detection in digital mammograms," in *2015 IEEE International Conference on Imaging Systems and Techniques (IST)*, Sep. 2015, pp. 1–6, doi: 10.1109/IST.2015.7294523.
[29] B. C. Patel and G. R. Sinha, "Mammography Feature Analysis and Mass Detection in Breast Cancer Images," in *2014 International Conference on Electronic Systems, Signal Processing and Computing Technologies*, Jan. 2014, pp. 474–478, doi: 10.1109/ICESC.2014.89.
[30] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep Learning to Improve Breast Cancer Detection on Screening Mammography," *Scientific Reports*, vol. 9, no. 1, p. 12495, Dec. 2019, doi: 10.1038/s41598-019-48995-4.
[31] C. Varela, P. G. Tahoces, A. J. Méndez, M. Souto, and J. J. Vidal, "Computerized detection of breast masses in digitized mammograms," *Computers in Biology and Medicine*, vol. 37, no. 2, pp. 214–226, Feb. 2007, doi: 10.1016/j.compbiomed.2005.12.006.

[32] W. Xie, Y. Li, and Y. Ma, "Breast mass classification in digital mammography based on extreme learning machine," *Neurocomputing*, vol. 173, pp. 930–941, Jan. 2016, doi: 10.1016/j.neucom.2015.08.048.

BIOGRAPHIES OF AUTHORS

Srwa Hasan Abdulla received the B.Sc. in Computer and Statistics Science from University of Sulaimani, Sulaimani Iraq in 2002, the M.Sc. in Information Technology from Jawaharlal Nehro Techonolgical University (JNTUH), Hyderabad, India in 2011, working as Lecturer at University of Sulamani from 2012–now; she is a Ph.D. student in Computer Science. She can be contacted at email: sirwa.abdulla@univsul.edu.iq.

Ali Makki Sagheer was born in Iraq-Basrah 1979. He got on B.Sc. of Information System in Computer Science Department at the University of Technology (2001)-Iraq, M.Sc. in Data Security from the University of Technology (2004)-Iraq and Ph.D. in Computer Science from the University of Technology (2007)-Iraq. He is interesting in cryptology, information security, cyber security number theory, multimedia compression, image processing, coding systems and artificial intelligence. He published more than eighty papers in different scientific journals and conferences thirty-one of them are published in Scopus. Finally, he obtained Professor scientific degree in Cryptography and Information Security since 18 Jul 2015. He can be contacted at email: ali.m.sagheer@gmail.com.

Hadi Veisi is an Associate Professor of Computer Science at the Faculty of New Sciences and Technologies, University of Tehran, Tehran. He is also the Head of the Digital Signal Processing Lab. He received his Ph.D. degree in Computer Engineering, Sharif University of Technology, Tehran-Iran, 2011. His current researches are digital signal processing, speech processing (speech recognition, speech enhancement, speech synthesis, and speaker recognition), natural language processing, artificial neural network, and deep learning. He can be contacted at email: h.veisi@ut.ac.ir.