



Performance evaluation of breast lesion detection systems with expert delineations: a comparative investigation on mammographic images

Bikesh K. Singh¹ · Pankaj Jain¹ · Sumit K. Banchhor¹ · Kesari Verma²

Received: 20 August 2018 / Revised: 8 March 2019 / Accepted: 2 April 2019 /

Published online: 16 April 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Performance of computerized diagnostic systems yearning to be approved by medical regulatory bodies must meet the expectations of human experts. Highly accurate lesion segmentation techniques have thus turned out to be an essential part for clinical acceptability of mammography based computer-aided diagnosis systems. The objective of this study is to evaluate the performance of six popular breast tumor detection techniques with manual delineations provided by two experienced radiologists on the mammographic images. In our study, 20 mammographic images from the mini-MIAS database are utilized. For the analysis, input mammographic images are first manually cropped to generate the region of interest (ROI). The ROI images are then pre-processed and segmentation is performed using different techniques, namely: expected maximization, K-means, Fuzzy c-Means (FCM), multilevel thresholding, region growing, and particle swarm optimization. The results were compared against the manual tracings. Among the other five segmentation techniques, FCM achieves the highest Jaccard Index (0.73 ± 0.06) and Dice Similarity Coefficient (0.82 ± 0.08) values. Statistical analysis (t-test, Mann Whitney U test, Wilcoxon test, Chi-Square test, and Kolmogorov–Smirnov test) and graphical analysis (Bland Altman and Regression plots) further prove the stability and reliability of the segmentation methods. Segmentation using FCM demonstrates the most accurate results and can be employed for the detection of breast cancer in the mammographic images. Further, it is concluded that computer-aided lesion detection systems can be used to assist Radiologists in routine clinical practice for the detection of breast tumors in mammographic images.

Keywords Mammogram · Lesion detection · Breast cancer · Performance evaluation · Manual delineations

✉ Bikesh K. Singh
bsingh.bme@nitrr.ac.in

¹ Department of Biomedical Engineering, National Institute of Technology Raipur, Raipur, CG 492001, India

² Department of Computer Applications, National Institute of Technology Raipur, Raipur, CG, India

1 Introduction

Breast cancer is one of the main reasons for death in women especially in Asia and worldwide [26]. Breast cancer is ranked second most dangerous cancer in overall. An estimate of more than 2.6 million new breast cancer cases will be diagnosed in 2019 [4].

Among all cancers, breast cancer is the most perilous diseases in women and can be treated only when detected at its initial stage [52]. Breast cancer is due to abnormal growth of the cell in the genes or body tissue [48]. Mammography is one of the widely used imaging modality [30] and plays an important role in the clinical examination for the diagnosis of breast disease [57]. Regular screening of mammograms for detection of masses i.e. small deposits of calcium in the breast which appear as small bright spot [2] is essential [6, 60]. During diagnosis, the identification of suspicious lesions which may be small or ill-defined are also important [8] and need to be segmented. Therefore, segmentation plays an important role in the diagnosis of breast cancer [19, 20, 49, 54].

Initially, mammogram images are manually diagnosed by the Radiologist and are referred to biopsy for final diagnosis. Due to the large size of the mammogram database, manual tracing of the lesions in all the images is tedious and thus can lead to observation errors [12]. To overcome this limitation, the opinion of more than one experts can be obtained, however, this can further lead to inter-observer variability [10, 37, 43]. To overcome these issues, Computer-aided diagnosis (CAD) system was introduced for segmentation and classification of the lesions which can provide another opinion to the radiologist [7, 20, 21].

Segmentation systems are consistent [19, 32, 36, 53] and thus provide a high level of accuracy. Segmentation is an important initial step of the CAD system [49]. In CAD system, existing techniques such as: expected maximization (EM) [45, 47, 55], K-means (KM) [13, 17, 22, 31, 44], Fuzzy c-Means (FCM) [5, 15, 16, 23, 42, 44, 46], multilevel thresholding (MT) [18, 34, 38, 51], region growing (RG) (single seed and multi-seed) [1, 14, 28, 59], and particle swarm optimization (PSO) [9, 29, 34] are found as rapid, fast, and robust. These techniques are free of tuning parameters, can be easily manipulated with smaller storage space, and are very popular in solving statistical estimation problems [19].

Many studies in the literature had employed these segmentation techniques for the detection of masses in the mammogram images. In 2015, Neto et al. [29] presented a segmentation technique followed by Otsu thresholding and PSO. False Positives Reduction was also performed by graph clustering and area filter. The study achieved a 95% segmentation correctness for 100 images from the DDSM (Digital Database for Screening Mammography) dataset. Along with successful segmentation of masses, the technique was also detecting small false regions which should be minimized. Raja et al. [34], proposed an improved PSO with multilevel thresholding using Otsu's function. With image histogram analysis from a chosen RGB image data set, a threshold from Otsu's is processed with Improved Particle Swarm Optimization (IPSO), PSO [35], and Darwinian Particle Swarm Optimization (DPSO) techniques. IPSO demonstrate result with less computational time as compared to the other two techniques (0.037 min as compared to DPSO and 0.053187 min as compared to PSO). In the same year, Elmoufidi [13], proposed a hybrid segmentation using K-means and Region Growing to detect different breast tissue regions in mammograms. Using MIAS databases, the proposed method obtained a mean precision percentage of 92.87% on 170 mammograms.

In 2016, Nurhasanah et al. [31], proposed a combined method of fractal 2D Fourier analysis and K-means clustering. Fractal analysis was used to identify the density of normal and abnormal tissues and K-means clustering was used to determine the area of lesions by

segmenting the boundary of abnormal tissue. The fractal dimension and intercept value are interpreted with a limit of 1 and 32 respectively for a normal and abnormal mammogram, respectively. K-means showed a better result for computing the area of an abnormal mammogram with less than 1 value for fractal dimension and intercept value greater than 32.

In 2017, Vedanarayanan et al. [55] proposed an efficient technique for segmenting breast cancer region with improved expected maximization and modified snake method. For feature extraction, abnormality, region, and boundary features were used and the backpropagation network was employed to classify the abnormality.

Recently in 2018, Sadad [42] proposed a CAD system by cascading Fuzzy c-Means and region-growing for segmentation and pre-processing, respectively. Decision Tree and KNN classifier were then applied to both MIAS and DDSM databases with hybrid features. For MIAS and DDSM databases, the system observed the accuracy of 98.2% and 95.8%, respectively. In the same year, Melouah [28] performed a study with two region growing methods for mammogram and IRM (Integrated Region Matching) datasets, which are (i) thresholding based region growing [1] and (ii) features similarity based region growing [59]. The methods were applied on 28 mammogram images acquired from mini Mammographic Image Analysis Society (MIAS) dataset and 20 IRM images acquired from RIDER (Reference Image Database to Evaluate Therapy Response) dataset. In this study, both thresholding and region growing methods showed a better result on mammographic images, but thresholding based region growing shows better result on IRM images as compared to features similarity based region growing. It was observed that the SSTT method does not give a good result with the high-intensity regions such as artifacts and labels. Also in the SSFS method, the reference values should be chosen very carefully that too at the center of the mass region.

As mentioned above, many studies were performed on a small dataset. Also, a comparison of the techniques on a similar dataset was missing. Further, none of the previous studies had performed a detailed performance evaluation and statistical analysis of the obtained results. Benchmarking with previous studies was also missing in many studies. All these limitations motivate us to implement and compare the above-mentioned segmentation techniques on a single dataset to identify the best technique for the detection of breast masses in mammograms, which is the novelty of the proposed work. In this study, performance evaluation is demonstrated by the Jaccard Index (JI) and the Dice Similarity Coefficient (DSC). Bland Altman and regression plots indicate the graphical similarity between segmentation and manual (ground truth) results. Statistical tests such as t-test, Mann Whitney, Wilcoxon, Chi-Square, and Kolmogorov–Smirnov (KS) are also performed to prove the stability and reliability of the study [3, 24, 33, 39].

The paper is organized as follows. Section 2 includes a brief summary of all the segmentation methods. Section 3 shows binary results of all the segmentation techniques and manual tracings performed by the two Radiologists. Section 4 presents a comparison of the segmentation results with the manual results using JI and DSC measures. Various statistical tests results are also presented. Finally, the contribution of the work is summarized in the Section 6.

2 Materials and methods

The mini-MIAS database is the widely used publicly available database of mammographic data. This database contains about 322 screening mammography cases, which contains approximately 207, 64, and 51 normal, benign, and malignant cases, respectively [50]. For

this study, 20 images are randomly selected from this mini-MIAS database which includes 10 images of both benign and malignant cases. The proposed methodology is a five-stage process as shown in Fig. 1.

2.1 Generation of ROI

The manually cropped region encloses the tumor area. In the first stage, input mammographic images from the database are manually cropped to obtain the desired ROI of fixed size (100×100). All the manual cropping are performed by Expert 1 who had a clinical experience of more than 20 years.

2.2 Pre-processing

Pre-processing enhances the intensity of the ROI image and further remove the noise. In the second stage, pre-processing is performed using open and close morphological operation using a single MATLAB command. Morphological opening involves erosion followed by the dilation and the inverse i.e. dilation followed by erosion happens in the case of morphological closing. In a binary image, dilation increase the thickness of the objects and erosion shrinks the thickness of the objects. The extent of thickening and shrinking is controlled by a structuring element. In our study, the structural element with a disk shape was chosen. The radius of the structuring element is chosen as four and six for malignant and benign images, respectively. The reason for this is because a higher and a lower radius are causing blurring of the ROI regions. The parameters fixed are found to be suitable for all the images.

2.3 Segmentation

The third stage involves extraction of the desired mass lesions by various segmentation techniques, namely: EM, FCM, KM, MT, PSO, and RG. In the fourth stage, segmentation results are compared with the manual tracings obtained by two experienced radiologists. Finally, in the fifth stage, performance evaluation is performed. The segmentation techniques employed in the current study are discussed briefly in the sub-sections below.

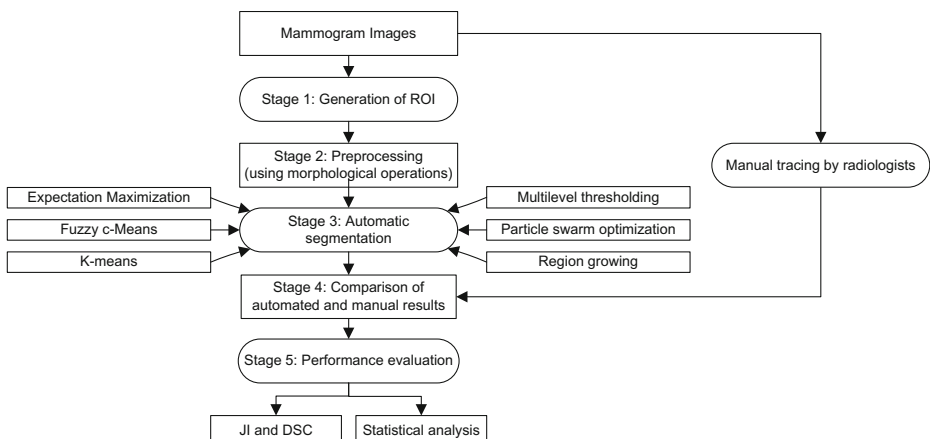


Fig. 1 Process flow diagram

2.3.1 Expectation maximization (EM)

The expectation maximization algorithm is a repetitive process to compute the Maximum Likelihood or probabilities of missing/hidden data from the most likely observed dataset. The EM algorithms basically consist of two processes: E-process and M-process [45, 47, 55]. In our work, a vector of class means, a vector of class variances, and vector of class proportions are used for computing the Expectation and Maximization. The vector of class means, a vector of class variances, and the vector of the class are computed by using the grayscale image and class value. The class value (which is the number of iterations) is morphologically chosen as 10. The maximum likelihood condition (which is termination condition) is the difference of likelihood of Maximization and Expectation. In our study, the maximum likelihood condition is morphologically chosen to be greater than 0.001.

2.3.2 Fuzzy c-means (FCM)

Fuzzy c-means is an unsupervised clustering iterative method in which the dataset is divided into many clusters with a membership function for each cluster. This method continuously updates the center of clusters with associated membership function and moves the center location within the dataset [5, 15, 16, 23, 44, 46]. To computer membership function and updated cluster center, the fuzziness coefficient, number of clusters, and maximum iteration were morphologically chosen as 2, 10, and 50, respectively.

2.3.3 K-means

K-means is an unsupervised clustering algorithm in which k number of clusters is defined with k number of the centroid. Now, points near the centroid are associated to make a group. Again the process of choosing centroid for the group is repeated until the new centroid location is fixed [13, 17, 22, 31, 44]. For computing cluster centers, minimum Euclidean Distance is used for assigning the nearest pixel. The number of clusters is morphologically chosen as 10.

2.3.4 Multilevel thresholding (MT)

Multilevel thresholding is a segmentation technique in which multiple thresholds for the gray image is determined by segmenting the image into many different regions. The segmentation is performed according to the brightness of different mass lesions. From multiple thresholds, one threshold is eliminated each time. A larger desired region is obtained by combining the threshold of a smaller segmented region with its adjacent region. Statistics like mean, standard deviation, and variance of gray images are evaluated to separate objects with multiple thresholds [18, 34, 38, 51]. Histogram entropy is used to compute the threshold. For optimization, the maximum threshold value is morphologically chosen as 10.

2.3.5 Particle swarm optimization (PSO)

In particle swarm optimization, each particle in the population has objective functions which are moving with adaptable velocity. The objective functions are maintained in

search space, so as to optimize the best fitness function objective. The velocity and positions of each particle are updated by two values: p-best (fitness by own) and g-best (fitness by neighbour) [9, 29, 34, 35]. In our study, we have chosen the minimum number of particles in the swarm as 10, dimension of the fitness function is twice the number of particles, swarm size is three times the dimension of the fitness function, lower fitness bound is 0, upper fitness bound is 255, maximum speed is 255; starting speed is 0, inertia constant is 0.73, cognition constant and social constant is 2.05. These values are the same as used by Neto et al. [29]. Apart from these parameters, the number of iteration was morphologically chosen as 10.

2.3.6 Region growing (RG)

The region growing algorithm, start from a manually located seed points applied over suspicious regions of masses in a mammogram. In region growing, initial seed points are located and the region regularly increases by comparing neighbour pixels with the same properties with the pixel of seed point [1, 14, 28, 59]. In our study, a seed (pixel value) is selected in the ROI. The segmented region is evaluated by using density weighting function. The density weighting function depends on the threshold and mean value of four neighbouring pixels. The algorithm will terminate when the distance between the segmented region and the new four neighbouring pixels become higher than a threshold (optimized intensity difference of the region). A value of 10,000 is specially chosen as the maximum threshold.

2.4 Manual segmentation by experts

Manual segmentation is a process of tracing masses in mammogram images by a Radiologist or medical practitioner. Manual delineations by the Radiologist is taken as a gold standard in this study for the evaluation of segmentation methods [25]. In this study, two experienced Radiologists with an experience of 20 years perform manual tracing of the mass lesions in all the mammogram images.

2.5 Statistical analysis

To evaluate the similarity between the different segmented techniques, parametric and nonparametric statistical tests are performed [3, 11, 39, 40]. The t-test (commonly known as Students t-test) is a parametric test used to show the difference between group sample data with respect to mean or average. Mann Whitney, Wilcoxon, Chi-Square and KS tests are nonparametric statistical tests [39]. Here, Mann Whitney test is used to compare how sample data are similar from the observed data, by evaluating the median of the sample data [3]. Wilcoxon test is used to evaluate the similarity between matched or repeated data in a single sample to perform difference rank mean, Chi-Square test is used to calculate dissimilarity between the measured and the expected data which are computed from the null hypothesis condition, and lastly, KS test is used to evaluate the normality test [41]. The significance of all tests depends on the p value (probability of chance) which should be greater than 0.05 to accept the null hypothesis condition. All the statistical tests are performed using MedCalc 16.8.4 software (Ostend Belgium) and a p value less than 0.05 is considered statistically significant [27].

Jaccard Index (also known as Jaccard Similarity Index) and Dice Similarity index (also known as Sorensen Dice Coefficient) is performed to evaluate the similarity between the segmentation and the manual results [56, 58]. Bland Altman and regression plots are also

studied to evaluate the difference and to identify the correlation between the segmentation and the manual results [33, 41]. Descriptions of all the statistical and performance tests are shown in Table 6 in Appendix 1.

3 Results

Table 1 shows the general statistical summary of the manual tracing results performed by both the Radiologists and the segmentation techniques for all the patients. All these statistical analyses are performed by taking mass lesion area as the parameter. Table 2 shows the detected area of mass regions (for both malignant and benign cases) observed by Radiologists and different segmentation techniques.

Figure 2 presents the ROI generation process. Figure 2a shows the original mammographic image, Fig. 2b show the original mammographic image with the manually cropped region (yellow overlay) and Fig. 2c indicate the manually cropped image (i.e. ROI). Figs. 3 and 4 shows the binary and overlay results of detected mass lesions using six different segmented techniques. Similarly, Figs. 5 and 6 shows the binary and overlay results of the detected mass lesions obtained from the manual tracings performed by Radiologist 1 and Radiologist 2, respectively. It can be observed that the results of all segmentation techniques are very similar to the manual results. Therefore, to identify the best segmentation technique, performance analysis is carried out in the next section.

4 Performance evaluation

The performance evaluation of all segmentation techniques with manual tracing results are observed in three parts: (i) comparison using JI and DSC measures; and (ii) performing statistical analysis namely: t-test, Mann Whitney, Wilcoxon, Chi-Square and KS test; and (iii) observing Bland Altman and regression plots.

4.1 Comparison of segmentation results with manual tracing results

The validation of the segmentation results is tested by comparing the observed results with the manual tracing results. The results of the JI and DSC test are shown in Table 3.

Table 1 Statistic summary of the detected area of mass lesions observed by different segmentation techniques and by both the Radiologists

Observation	Mean	Variance	SD	Median
Using different segmentation techniques				
Expectation maximization	2.6556	0.8297	0.9109	2.7756
Fuzzy c-means	2.7103	0.7951	0.8917	2.7264
K-means	2.8725	1.2282	1.1083	2.9582
Multilevel thresholding	2.4344	1.1209	1.0587	2.2226
Region growing	3.5387	2.3044	1.5180	3.3028
Particle swarm optimization	2.3908	0.6070	0.7791	2.2309
By Radiologist				
Radiologist 1	2.7002	0.6854	0.8279	2.6984
Radiologist 2	2.7897	1.0965	1.0471	2.7421

Table 2 Detected area of mass regions observed different segmentation techniques and by both the Radiologists

Observation	For malignant cases		For benign cases	
	Mean	SD	Mean	SD
Using different segmentation techniques				
Expectation maximization	2.478	0.799	2.832	1.020
Fuzzy c-means	2.451	0.700	2.969	1.019
K-means	2.690	0.904	3.054	1.304
Multilevel thresholding	2.357	1.009	2.511	1.155
Region growing	3.540	1.607	3.536	1.509
Particle swarm optimization	2.430	0.874	2.350	0.716
By Radiologist				
Radiologist 1	2.422	0.563	2.977	0.978
Radiologist 2	2.463	0.786	3.116	1.207

4.2 Statistical analysis

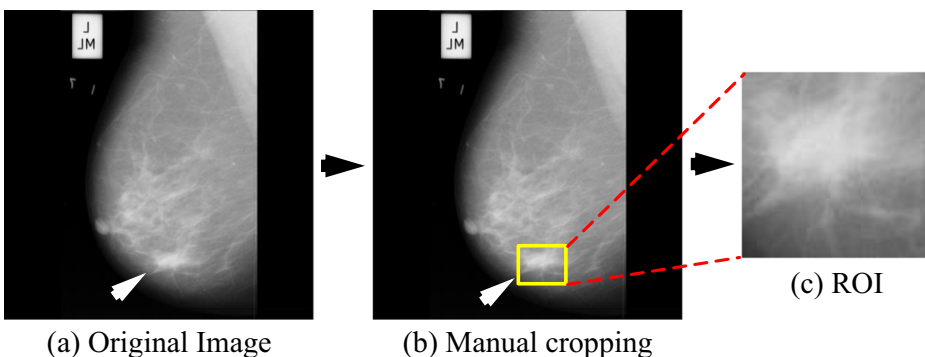
For statistical analysis t-test, Mann Whitney, Wilcoxon, Chi-Square, and KS tests are performed. Table 4 shows a statistical analysis and coefficient correlation of all the six different segmentation techniques and the two manual tracing results performed by both the Radiologists.

4.3 Graphical analysis

In Appendix 2, Figs. 7 and 8 present the Bland Altman plots and Figs. 9 and 10 present the regression plots between different segmentation techniques and the manual tracing results obtained from both the Radiologists.

5 Discussion

In our study, mass lesions in mammogram images are segmented by different segmentation techniques and the results are validated by using the manual tracings as shown in Fig. 1. The results of the study are now discussed in the below subsections.

**Fig. 2** ROI generation process

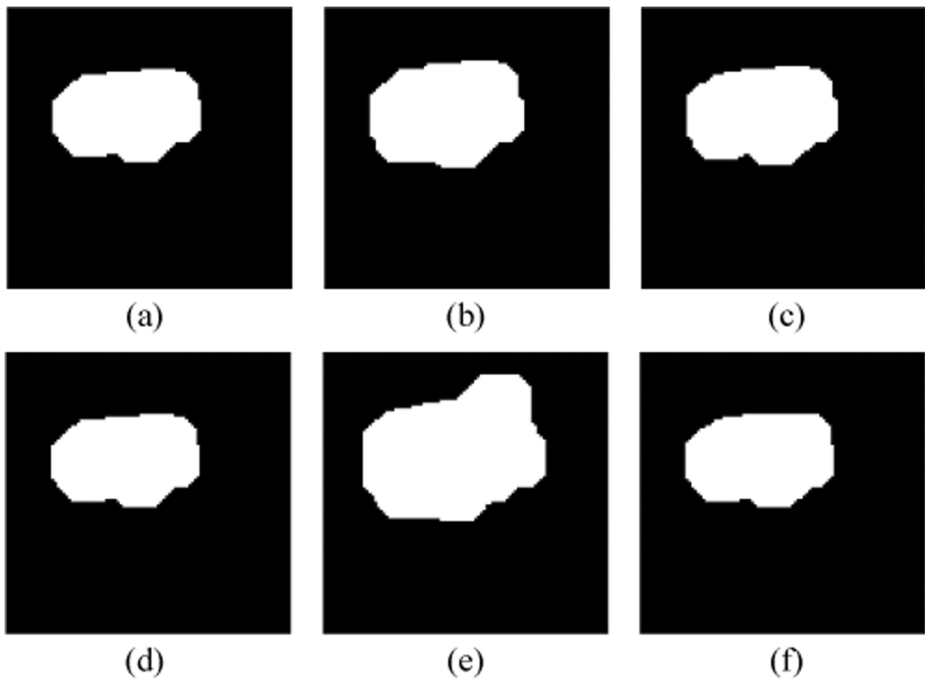


Fig. 3 Binary results of detected mass lesions using six different segmented techniques: **a** Expected Maximization, **b** Fuzzy c-Means, **c** K-means, **d** Multilevel thresholding, **e** Region Growing, and **f** PSO

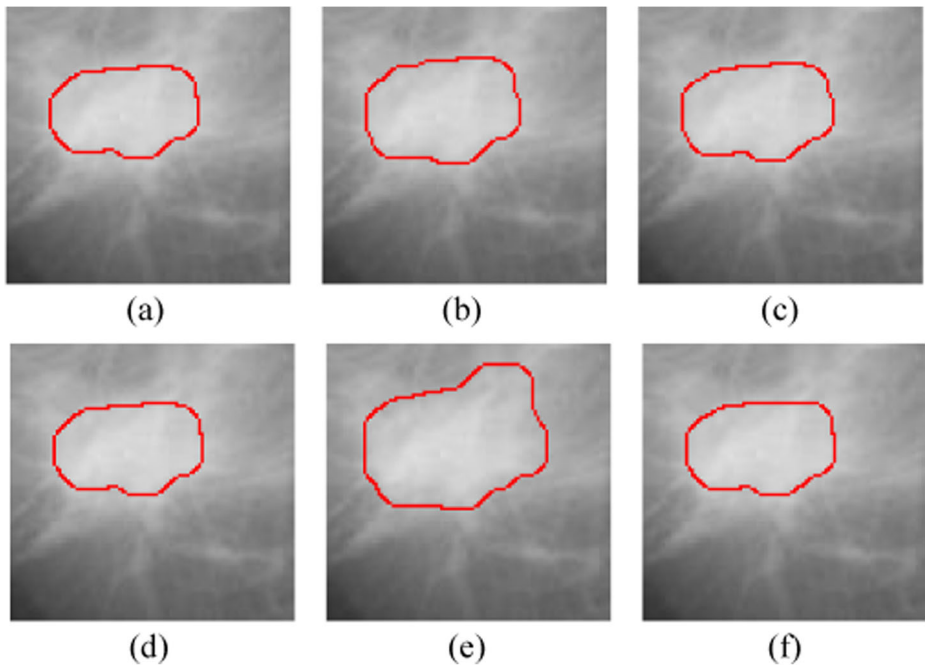


Fig. 4 Overlay results of six different segmented techniques: **a** Expected Maximization, **b** Fuzzy c-Means, **c** K-means, **d** Multilevel thresholding, **e** Region Growing, and **f** PSO. Red contour is the detected mass lesion

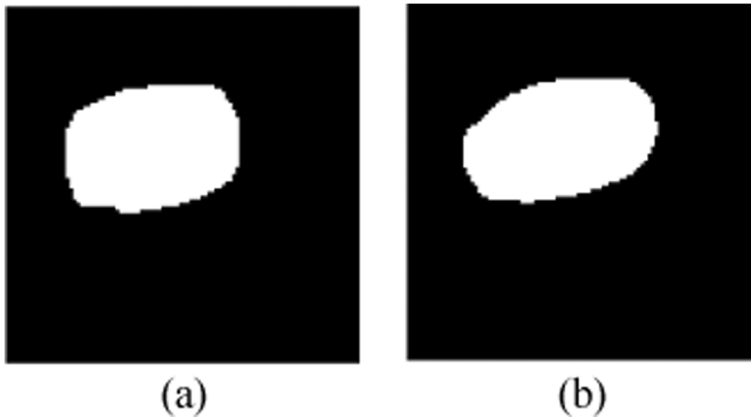


Fig. 5 Binary results of detected mass lesions obtained from the manual tracings performed by **a** Radiologist 1 and **b** Radiologist 2, respectively

5.1 A note on statistical summary

The statistical summary (see Table 2) of breast lesion area was performed to compare segmentation techniques with both the manual tracing results. From Table 2, it was observed that the median of lesion area for FCM (2.7264) and EM (2.7756) are very close to Radiologist 1 (2.6984) and Radiologist 2 (2.7421). This proves a high degree of similarity between segmentation and manual results. In Table 2, the median of the lesion area for region growing (3.3028) shows less similarity with both the Radiologists. This demonstrates poor performance of region growing method in detection of breast masses.

5.2 A note on the comparison of segmentation results with manual tracing results

From Table 3 it can be observed that JI for FCM and EM showed the highest mean value (0.73, 0.72) and lowest SD value (0.06, 0.08), with both the Radiologists, whereas region

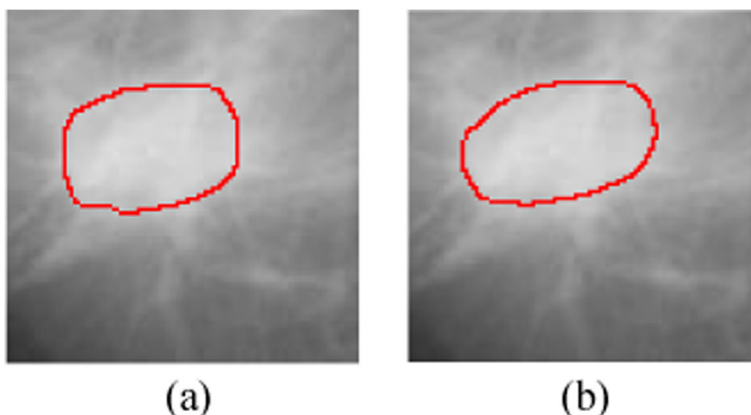


Fig. 6 Overlay results obtained from the manual tracings performed by **a** Radiologist 1 and **b** Radiologist 2, respectively. Red contour is the detected mass lesion

Table 3 Comparison between different segmentation technique and manual tracing results performed by both Radiologists 1 and Radiologists 2

Methods	Jaccard Index				Dice Similarity Coefficient			
	Radiologist 1		Radiologist 2		Radiologist 1		Radiologist 2	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
EM	0.72	0.08	0.72	0.07	0.84	0.05	0.84	0.05
FCM	0.73	0.06	0.73	0.06	0.82	0.08	0.82	0.08
KM	0.70	0.11	0.70	0.11	0.84	0.04	0.84	0.04
MT	0.67	0.12	0.68	0.10	0.80	0.10	0.81	0.08
PSO	0.65	0.13	0.63	0.13	0.78	0.10	0.77	0.10
RG	0.61	0.17	0.60	0.16	0.74	0.15	0.74	0.15

EM Expectation Maximization, FCM Fuzzy c-Means, KM K-Means, MT Multilevel Thresholding, RG Region Growing, PSO Particle Swarm Optimization

growing showed the lowest mean value (0.61) and highest SD (0.17) value for the same. Similarly for DSC, KM, and EM shows the highest mean value (0.84, 0.84) and lowest SD value (0.04, 0.05), with both the Radiologists, whereas region growing shows the lowest mean value (0.61) and highest SD (0.17) value for the same. FCM and EM are clustering algorithms with fewer inhomogeneities variation and have the robustness to noise ambiguity. This feature allows them to retain much information as compared to other techniques especially region growing. Region growing suffers from high inhomogeneities variation as it merges neighbour pixels with the statistical properties.

Higher the value of JI and DSC more is the similarity between the two results. It can be observed that all segmentation techniques have good accuracy and hence, the results are validated. In Table 3, the mean value of JI and DSC for both FCM and EM proved their

Table 4 Statistical test summary between segmentation technique and manual tracing results performed by both Radiologists 1 (R1) and Radiologists 2 (R2)

Combinations	T-test	Mann Whitney-test	Wilcoxon test	Chi Squared-test		KS-test	CC
	p value	p value	p value	CoC	p value	p value	
Segmentation techniques and Radiologist 1 (R1)							
R1 vs. EM	0.700	<0.0001	0.123	0.975	0.236	>0.10	0.83 (p < 0.001)
R1 vs. KM	0.922	<0.0001	0.870	0.975	0.236	>0.10	0.77 (p < 0.001)
R1 vs. FCM	0.288	<0.0001	0.090	0.975	0.236	>0.10	0.86 (p < 0.001)
R1 vs. MT	0.142	1.000	0.017	0.975	0.236	>0.10	0.69 (p < 0.001)
R1 vs. RG	0.025	1.000	0.014	0.975	0.236	>0.10	0.25 (p < 0.286)
R1 vs. PSO	0.124	1.000	0.076	0.975	0.236	>0.10	0.43 (p < 0.059)
Segmentation techniques and Radiologist 2 (R2)							
R2 vs. EM	0.316	1.000	0.090	0.975	0.236	>0.10	0.83 (p < 0.001)
R2 vs. KM	0.399	<0.0001	0.498	0.975	0.236	>0.10	0.83 (p < 0.001)
R2 vs. FCM	0.568	0.448	0.368	0.975	0.236	>0.10	0.92 (p < 0.001)
R2 vs. MT	0.034	<0.0001	0.019	0.975	0.236	>0.10	0.78 (p < 0.001)
R2 vs. RG	0.044	1.000	0.022	0.975	0.236	>0.10	0.31 (p < 0.178)
R2 vs. PSO	0.088	1.000	0.114	0.975	0.236	>0.10	0.44 (p < 0.050)

CoC Contingency Coefficient, CC Correlation Coefficient, EM Expectation Maximization, KM K-Means, FCM Fuzzy c-Means, MT Multilevel Thresholding, RG Region Growing, PSO Particle Swarm Optimization

similarity with both the manual tracing results. Hence, FCM and EM segmentation accuracy are very much acceptable with manual segmentation as compare to the results obtained from other segmentation techniques.

5.3 A note on statistical test

In Table 4, for all the combinations we observed that the KS test with a p value greater than 0.10 and Chi-Square test with a constant p value of 0.236 showed no statistically significant difference. In the t-test for all the combinations, the p value was greater than 0.05 which proved that there is no difference between the two results. Similarly, in the Wilcoxon test, the p value also satisfies the condition which proved that the median of paired observations differences is statistically significant. Region growing technique in both the tests does not satisfy the condition of similarity with both Radiologists. The higher p value of both EM and KM showed their ability to perform accurate mass lesion segmentation as compare with the manual tracing results.

5.4 A note on graphical analysis

Bland Altman plots as presented in Figs. 7 and 8 in the Appendix 2. From the plots, it was observed that variability between EM, FCM, KM, and Radiologist 1 and 2 is lower since all the points are uniformly scattered within the two predictor intervals ($\text{mean} \pm 1.96 \text{ SD}$) excluding few outliers. The low variations proved higher similarity between segmentation and manual tracing results.

Similarly, in the regression plots as presented in Figs. 9 and 10 in the Appendix 2. The regression line, 95% confidence interval, and 95% predictor intervals interpret accurate analysis of the similarity between the two results. For both EM and KM, most of the points lie within the confidence interval which showed a high degree of similarity between segmentation and manual tracing results.

Previous studies had used both FCM and EM for the segmentation of mass lesions in mammogram images. Pavan et al. [42], proposed a segmentation technique for mammogram using FCM. They employed various FCM features on breast tissues and showed performance with 93% consistency with Breast imaging reporting and data system (BIRADS). Also, Vedanarayanan et al. [55] proposed an efficient technique with improved EM for mammograms with showed 100% sensitivity. In the proposed study we have used the same segmentation techniques and the observed results are consistent with the above studies.

5.5 Comparison against current literature and benchmarking

A lot of work has been done so far on breast segmentation. Table 5 shows the comparison between the proposed study and the current literature using ten different attributes, namely: year, number of frames, objective, the technique used, validation, ground truth, number of manual tracers, accuracy, benchmarking, and performance evaluation. In 2014, Dheeba [9], proposed Particle Swarm Optimized Wavelet Neural Network for the detection of breast abnormalities in digital mammograms. A pattern classifier was used for the classification of the suspicious regions. The proposed algorithm was tested on 216 mammograms and a sensitivity and specificity of 94.167% and

Table 5 Survey of other breast cancer segmentation methods in literature and proposed work

Author (Year)	No. of frames	Objective	Technique used	Validation	Ground truth	No. of manual tracers	Accuracy (%)	Benchmarking	Performance evaluation
Dheeba (2014) [9]	216	Detection of breast cancer	PSO, Wavelet, and NN	Yes	Radiologist	More than one	93.67	None	Sensitivity and specificity
Elmoufidi et al. (2015) [13]	170	Density-based breast segmentation	K-means and Seed-based region growing	Yes	Radiologist	1	Mean precision percentage: 92.87%	Yes	None
Neto et al. (2015) [29]	100	Automatic segmentation of masses	PSO and graph clustering	None	None	None	95.2	Yes	None
Raja et al. (2015) [34]	3	Segmentation of cancer infected breast	PSO-based Multilevel thresholding	None	None	None	None	None	None
Nurhasanah et al. (2016) [31]	6	To generate boundaries of normal and abnormal tissue organs	Fractal method and K-means clustering	None	None	None	None	None	None
Vednarayanan et al. (2017) [55]	50	Advanced segmentation to enhance breast cancer detection	Modified EM and modified snake algorithm	None	None	None	Sensitivity: 100%	None	None
Saad et al. (2018) [42]	181	Classification of the tumor from Mammograms	FCM, RG, DT, KNN, SVM, LR, LDA and Ensemble Classifier	Yes	MiniMIas dataset	None	98.2	Yes	Sensitivity, specificity, MCC, and F-score
Melouah (2018) [28]	28	Comparison of automatic seed generation methods	RG	Yes	Radiologist	1	None	None	None
Proposed (2019)	20	Comparison of segmentation methods	EM, K-means, FCM, multilevel thresholding, RG, and PSO	Yes	Radiologist	2	JI: 0.73 ± 0.06, DSC: 0.82 ± 0.08	Yes	JI, DSC, Bland Altman, and Regression Plots

PSO Particle Swarm Optimization, NN Neural Network, EM Estimation Maximization, RG Region Growing, DT Decision Tree, KNN K-nearest Neighbors, SVM Support Vector Machine, LR Logistic Regression, LDA Linear Discriminant Analysis, MCC Matthews's Correlation Coefficient, JI Jaccard Index, DSC Dice Similarity Index

92.105%, respectively was observed. Elmoufidi [13], Neto [29], and Raja [34] had tried to perform segmentation of cancer masses in the infected breast using mammogram images. Elmoufidi [13] validated their results against one radiologist and obtained a mean precision percentage of 92.87%. Neto [29] also observed the accuracy of 95.2% but similar to Raja [34], Nurhasanah [31], and Vedanarayanan [55] did not validate their results. In 2018, Sadad [42] performed classification of the tumor from Mammograms using many classifiers and observed the accuracy of 98.2%. The study also validated their results against the MIAS dataset. Very recently, Melouah [28] performed a comparison of automatic seed generation methods using region growing technique. The observed results were benchmarked against one radiologist. Ours is the first study that had tried to compare six different segmentation methods, namely; EM, K-means, FCM, multilevel thresholding, RG, and PSO on a single database which brings in the novelty. The results are validated by two different radiologists and performance evaluation was also performed by JI, DSC, Bland Altman, and Regression plots.

5.6 Strength, weakness, and extensions

The main objectives of this study were to evaluate the performance of lesion detection in the mammogram using different segmentation technique and its comparison with manual delineations provided by expert Radiologists. For validation JI and DSC, measures were employed. Similarly, extensive statistical and graphical tests are used to validate the findings of the study. However, the study also suffers from some limitations such as low data size and manual ROI generation. In the future, the proposed semi-automated system can be made fully automated by using the automated ROI generation system. The segmentation results can be tested on more real-time mammogram datasets. In future, with a larger database, we will implement and compare the performance of different approaches discussed in this study with that of deep neural network. Inter- and intra-observer variability analysis can also be performed in the future. Developing methods to enhance interaction between CAD and medical experts is also important for its clinical acceptance and belief.

6 Conclusion

This study reported a comparative evaluation of six different segmentation techniques namely: EM, FCM, K-means, multilevel thresholding, RG, and PSO for the detection of mass lesions in the mammographic images. Further, the main aim was to compare the performance of these techniques with manual delineations provided by two experienced Radiologists. The result indicates that FCM and EM as the most accurate techniques which can be implied in the clinical settings. Performance evaluation and statistical analysis further proved the stability and reliability of the study. Results of this study indicate that computer-aided lesion detection systems can assist Radiologists in routine clinical practice for the detection of breast tumors in mammographic images.

Acknowledgements This work was supported by the Chhatisgarh Council of Science & Technology, Raipur, India (grant number 2481/CCOST/MRP/2016). The authors of this article are extremely grateful to them.

Appendix 1

Table 6 Description of all the statistical and performance tests

Test performed	Formula	Variable description
T-test	$t = \frac{\Sigma P / X}{\sqrt{\frac{\Sigma P^2 - (\Sigma P / X)^2}{(X) \times (X-1)}}$	ΣP = Sum of the differences, ΣP^2 = Sum of the squared differences, X = number of data.
Mann Whitney test	$U = a_1 \times a_2 + \frac{a_2(a_1+1)}{2} - \sum_{i=a_1+1}^{a_2} X_i$	a_1 = sample size one data, a_2 = sample size second data, X_i = Rank of sample size.
Wilcoxon test	$W = \sum_{i=1}^{K_r} \text{sgn}(a_{2i} - a_{1i}) \times X_i$	K_r = paired sample data, a_{1i} and a_{2i} = measured data, X_i = rank of sample
Chi-Square test	$\chi^2 = \frac{\sum (O_d - E_d)^2}{E_d}$	O_d = measured observed data, E_d = Expected data.
KS test	$D = \sup_x (P_n(x) - P(x))$	\sup = supremum of the set of distances, $P_n(x)$ = commutative distribution function of hypothesis distribution, $P(x)$ = EDF of observed data.
Jaccard Index	$J = \frac{ M \cap N }{ M \cup N }$	M and N are binary images.
Dice Similarity Coefficient	$DSC = 2 \times \frac{ M \cap N }{ M + N }$	M and N are binary images.

Appendix 2

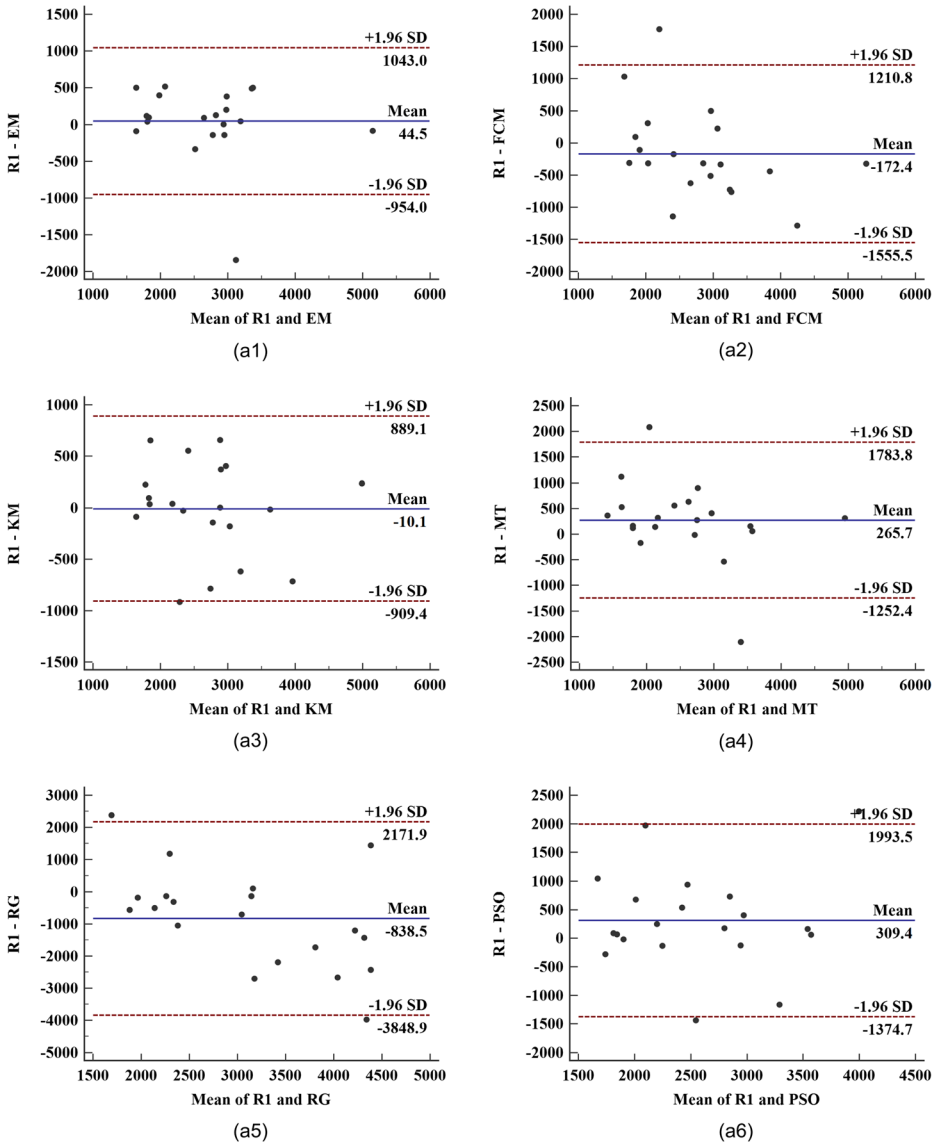


Fig. 7 Bland Altman plots between all the segmentation techniques (expected maximization (EM), Fuzzy c-Means (FCM), K-means (KM), multilevel thresholding (MT), region growing (RG), particle swarm optimization (PSO), and Radiologist 1 (R1))

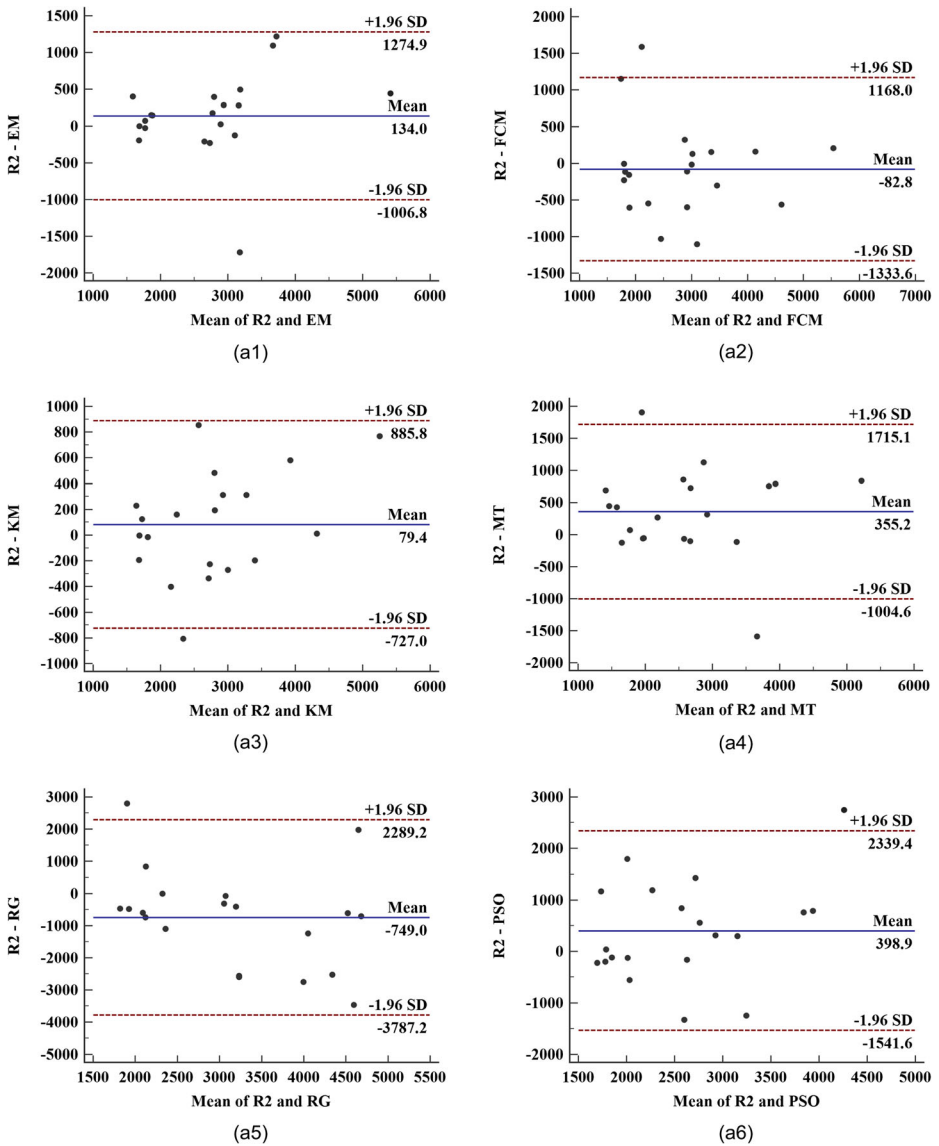


Fig. 8 Bland Altman plots between all the segmentation techniques (expected maximization (EM), Fuzzy c-Means (FCM), K-means (KM), multilevel thresholding (MT), region growing (RG), particle swarm optimization (PSO), and Radiologist 2 (R2))

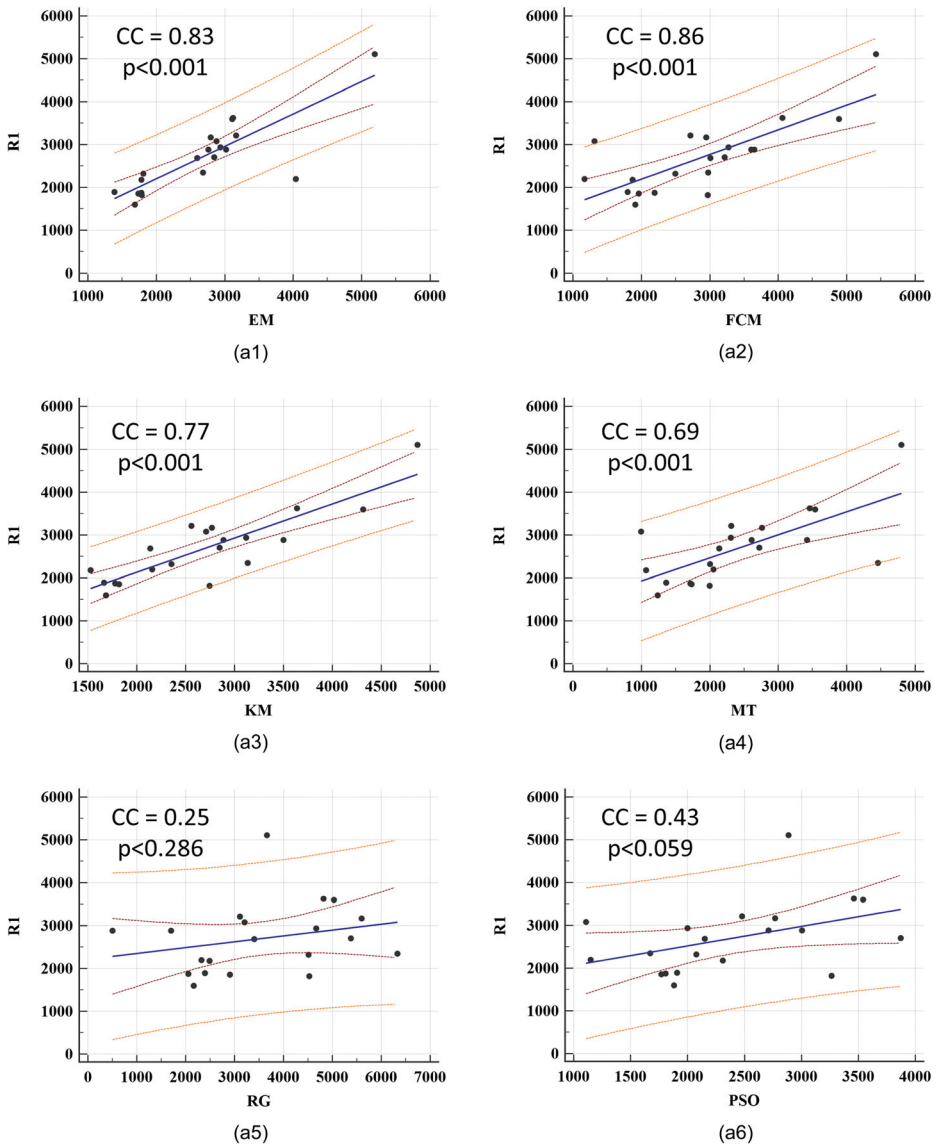


Fig. 9 Regression plots between all the segmentation techniques (expected maximization (EM), Fuzzy c-Means (FCM), K-means (KM), multilevel thresholding (MT), region growing (RG), particle swarm optimization (PSO), and Radiologist 1 (R1))

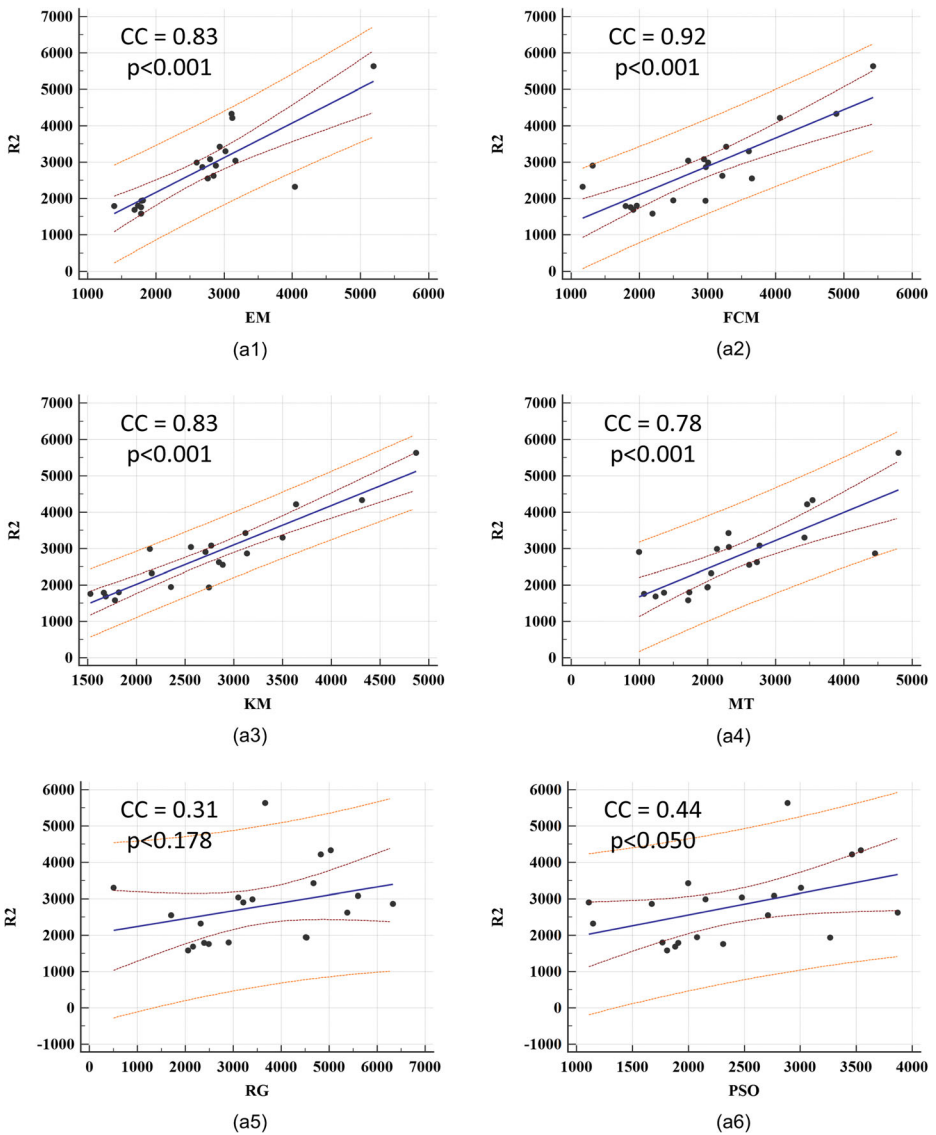


Fig. 10 Regression plots between all the segmentation techniques (expected maximization (EM), Fuzzy c-Means (FCM), K-means (KM), multilevel thresholding (MT), region growing (RG), particle swarm optimization (PSO), and Radiologist 2 (R2))

References

1. Al-Faris AQ, Umi Kalthum N, Matlsa NA, Shuaib IL (2014) Computer-aided segmentation system for breast MRI tumour using modified automatic seeded region growing (BMRI-MASRG). *J Digit Imaging* 27:133–144
2. Arodz T, Kurdziel M, Popielka TJ, Sevre EO, Yuen DA (2006) Detection of clustered microcalcifications in small field digital mammography. *Comput Methods Prog Biomed* 81(1):56–65
3. Banchhor SK, Londhe ND, Saba L, Radeva P, Laird JR, Suri JS (2017) Relationship between automated coronary calcium volumes and a set of manual coronary lumen volume, vessel volume and atheroma volume in Japanese diabetic cohort. *J Clin Diagn Res* 11(6):TC09
4. Breast cancer statistics, How common is breast cancer? American cancer society. Online document available at <https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breastcancer.html>. Accessed 4 Nov 2019
5. Chen Z, Zwiggelaar R A modified fuzzy c-means algorithm for breast tissue density segmentation in mammograms. In: Information technology and applications in biomedicine (ITAB), 2010 10th IEEE international conference on 2010 Nov 3. IEEE, pp 1–4
6. Choi JY, Kim DH, Plataniotis KN, Ro YM (2016) Classifier ensemble generation and selection with multiple feature representations for classification applications in computer-aided detection and diagnosis on mammography. *Expert Syst Appl* 46:106–121
7. Ciecholewski M (2017) Malignant and benign mass segmentation in mammograms using active contour methods. *Symmetry* 9(11):277
8. Costa DD, Campos LF, Barros AK (2011) Classification of breast tissue in mammograms using efficient coding. *Biomed Eng Online* 10:55
9. Dheeba J, Singh NA, Selvi ST (2014) Computer-aided detection of breast cancer on mammograms: a swarm intelligence optimized wavelet neural network approach. *J Biomed Inform* 49:45–52
10. Duijm LE, Louwman MW, Groenewoud JH, Van de Poll-Franse LV, Fracheboud J, Coebergh JW (2009) Inter-observer variability in mammography screening and effect of type and number of readers on screening outcome. *Br J Cancer* 100(6):901
11. DuPrel JB, Röhrig B, Hommel G, Blettner M (2010) Choosing statistical tests: part 12 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 107(19):343
12. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR (1994) Variability in Radiologists' interpretations of mammograms. *N Engl J Med* 331(22):1493–1499
13. Elmoufidi A, El Fahssi K, Jai-Andaloussi S, Sekkaki A (2015) Automatically density based breast segmentation for mammograms by using dynamic K-means algorithm and seed based region growing. In: Instrumentation and measurement technology conference (I2MTC), 2015 IEEE international. IEEE, pp 533–538
14. Fan J, Zeng G, Body M, Hacid MS (2005) Seeded region growing: an extensive and comparative study. *Pattern Recogn Lett* 26(8):1139–1156
15. GokilaDeepa G (2012) Mammogram image segmentation using fuzzy hybrid with particle swarm optimization (PSO). *International Journal of Engineering and Innovative Technology (IJEIT)* 2(6)
16. Guliatto D, Rangayyan RM, Carnielli WA, Zuffo JA, Desautels JL (2003) Segmentation of breast tumors in mammograms using fuzzy sets. *J Electronic Imaging* 12(3):369–379
17. Gumaei A, El-Zaart A, Hussien M, Berbar M Breast segmentation using k-means algorithm with a mixture of gamma distributions. In: Broadband networks and fast internet (RELABIRA), 2012 symposium on 2012 May 28. IEEE, pp 97–102
18. Harrabi R, Braiek EB (2012) Color image segmentation using multi-level thresholding approach and data fusion techniques: application in the breast cancer cells images. *E J Image Video Proc* 2012(1):11
19. He W, Hogg P, Juetta A, Denton ER, Zwiggelaar R (2015) Breast image pre-processing for mammographic tissue segmentation. *Comput Biol Med* 67:61–73
20. Jalalian A, Mashohor S, Mahmud R, Karasfi B, Saripan MI, Ramli AR (2017) Foundation and methodologies in computer-aided diagnosis systems for breast cancer detection. *EXCLI J* 16:113
21. Jen CC, Yu SS (2015) Automatic detection of abnormal mammograms in mammographic images. *Expert Syst Appl* 42(6):3048–3055
22. Karmiliasari SW, Hermita M, Agustiyani NP, Hanum Y, Lussiana ET (2014) Sample k-means clustering method for determining the stage of breast cancer malignancy based on cancer size on mammogram image basis. *IJACSA. Int J Adv Comput Sci Appl* 5(3):86–90
23. Keller B, Nathan D, Wang Y, Zheng Y, Gee J, Conant E, Kontos D (2011) Adaptive multi-cluster fuzzy C-means segmentation of breast parenchymal tissue in digital mammography. In: International conference on medical image computing and computer-assisted intervention. Springer, Berlin, pp 562–569
24. Keller BM, Nathan DL, Gavenonis SC, Chen J, Conant EF, Kontos D (2013) Reader variability in breast density estimation from full-field digital mammograms: the effect of image postprocessing on relative and absolute measures. *Acad Radiol* 20(5):560–568

25. Li Y, Brennan PC, Lee W, Nickson C, Pietrzyk MW, Ryan EA (2015) An investigation into the consistency in mammographic density identification by radiologists: effect of radiologist expertise and mammographic appearance. *J Digit Imaging* 28(5):626–632
26. Malvia S, Bagadi SA, Dubey US, Saxena S (2017) Epidemiology of breast cancer in Indian women. *Asia Pac J Clin Oncol* 13(4):289–295
27. MedCalc:- <https://www.medcalc.org/manual/>. Accessed 3 July 2019
28. Melouah A (2015) Comparison of automatic seed generation methods for breast tumor detection using region growing technique. In: IFIP international conference on computer science and its Applications_x000D_. Springer, Cham, pp 119–128
29. Neto OP, Carvalho O, Sampaio W, Corrêa A, Paiva A Automatic segmentation of masses in digital mammograms using particle swarm optimization and graph clustering. In: Systems, signals and image processing (IWSSIP), 2015 international conference on 2015 Sep 10. IEEE, pp 109–112
30. Ng KH, Muttarak M (2003) Advances in mammography have improved early detection of breast cancer. *J HK Coll Radiol* 6:126–131
31. Nurhasanah, Sampumo J, Faryuni ID, Ivansyah O (2016) Automated analysis of image mammogram for breast cancer diagnosis. In: AIP conference proceedings, vol. 1719, no. 1. AIP Publishing, pp 030036
32. Oliver A, Freixenet J, Marti J, Perez E, Pont J, Denton ER, Zwiggelar R (2010) A review of automatic mass detection and segmentation in mammographic images. *Med Image Anal* 14(2):87–110
33. Pompe E, de Jong PA, De Jong WU, Takx RA, Eikendal AL, Willeminck MJ, Oudkerk M, Budde RP, Lammers JW, Hoessein FA (2016) Inter-observer and inter-examination variability of manual vertebral bone attenuation measurements on computed tomography. *Eur Radiol* 26(9):3046–3053
34. Raja NS, Sukanya SA, Nikita Y (2015) Improved PSO based multi-level thresholding for cancer infected breast thermal images using Otsu. *Procedia Computer Science* 48:524–529
35. Raju NG, Rao NC (2013) Particle swarm optimization methods for image segmentation applied in mammography. *Journal of Engineering Research and Applications* 3(6):1572–1579. ISSN: 2248-9622
36. Ramani R, Valarmathy S, Vanitha NS (2013) Breast cancer detection in mammograms based on clustering techniques-a survey. *Int J Comput Appl* 62(11)
37. Redondo A, Comas M, Macia F, Ferrer F, Murta-Nascimento C, Maristany MT, Molins E, Sala M, Castells X (2012) Inter-and intra radiologist variability in the BI-RADS assessment and breast density categories for screening mammograms. *Br J Radiol* 85(1019):1465–1470
38. Rejani Y, Selvi ST (2009) Breast Cancer detection using multilevel thresholding. arXiv preprint arXiv:0911.0490
39. Saba L, Than JC, Noor NM, Rijal OM, Kassim RM, Yunus A, Ng CR, Suri JS (2016) Inter-observer variability analysis of automatic lung delineation in normal and disease patients. *J Med Syst* 40(6):142
40. Saba L, Banchhor SK, Araki T, Suri HS, Londhe ND, Laird JR, Viskovic K, Suri JS (2018) Intra-and inter-operator reproducibility analysis of automated cloud-based carotid intima media thickness ultrasound measurement. *J Clin Diagn Res* 12(2):KC01-KC11
41. Saba L, Banchhor SK, Araki T, Viskovic K, Londhe ND, Laird JR, Suri HS, Suri JS (2018) Intra-and inter-operator reproducibility of automated cloud-based carotid lumen diameter ultrasound measurement. *Indian Heart J* 70:649–664
42. Sadad T, Munir A, Saba T, Hussain A (2018) Fuzzy C-means and region growing based classification of tumor from mammograms using hybrid texture feature. *J Comput Sci* 29:34–45
43. Saha A, Grimm LJ, Harowicz M, Ghate SV, Kim C, Walsh R, Mazurowski MA (2016) Interobserver variability in identification of breast tumors in MRI and its implications for prognostic biomarkers and radiogenomics. *Med Phys* 43(8 Part1):4558–4564
44. Sathish A, Sundaram JM (2004) A comparative study on K-means and fuzzy C-means algorithm for breast cancer analysis. *International Journal of Computational Intelligence and Informatics*. H. Simpson, Dumb robots, 3rd edn. UOS Press, Springfield, pp 6–9
45. Satyendra SA, Pawar MM (2017) Segmentation of breast images using Gaussian mixture models. *Int J Adv Res Ideas Innov Technol* 3(3):437–441. ISSN: 2454-132X
46. Senthilkumar B, Umamaheswari G, Karthik J A novel region growing segmentation algorithm for the detection of breast cancer. In: Computational intelligence and computing research (ICCC), 2010 IEEE international conference on 2010 Dec 28. IEEE, pp 1–4
47. Sheshadri HS, Kandaswamy A (2005) Detection of breast cancer by mammogram image segmentation. *J Cancer Res Ther* 1(4):232
48. Siegel R, Naishadham D, Jemal A (2013a) Cancer statistics. *CA Cancer J Clin* 63:11–30
49. Spandana P, Rao KM (2013) Novel image processing techniques for early detection of breast cancer, mat lab and lab view implementation. In: Point-of-Care Healthcare Technologies (PHT), pp 105–108
50. Suckling J et al (1994) The Mammographic Image Analysis Society Digital Mammogram Database Excerpta Medica. *Int Congr Ser* 1069:375–378

51. Sujata RB, Dhiman R, Chugh TS (2012) An evaluation of two mammography segmentation techniques. *International Journal of Advanced Computer Research* 2 Number-4(7):2277–7970. (ISSN (print): 2249-7277 ISSN (online)
52. Survey by Indian cancer society (2013) Indian cancer society
53. Thyagarajan R, Murugavalli S (2012) Segmentation of digital breast tomograms using clustering techniques. In: India conference (INDICON), 2012 annual IEEE. IEEE, pp 1090–1094
54. Valarmathie P, Sivakrithika V, Dinakaran K (2016) Classification of mammogram masses using selected texture, shape and margin features with multilayer perceptron classifier. *Biomedical research India*, Special Issue page no. s310–s313
55. Vedanarayanan V (2017) Advanced image segmentation techniques for accurate isolation of abnormality to enhance breast cancer detection in digital mammographs. *Biomed Res* 28(6):2753–2757
56. Vesal S, Ravikumar N, Ellman S, Maier A (2018) Comparative analysis of unsupervised algorithms for breast MRI lesion segmentation. In: *Bildverarbeitung für die Medizin 2018*. Springer Vieweg, Berlin, pp 257–262
57. Wang Y, Li J, Gao X (2014) Latent feature mining of spatial and marginal characteristics for mammographic mass classification. *Neurocomputing* 144:107–118
58. Wang J, Jing H, Wernick MN, Nishikawa RM, Yang Y (2014) Analysis of perceived similarity between pairs of microcalcification clusters in mammograms. *Med Phys* 41(5):051904
59. Yuvaraj K, Ragupathy US (2013) Automatic mammographic mass segmentation based on region growing technique. In: 3rd international conference on electronics, biomedical engineering and its applications (ICEBEA'2013), pp 29–30
60. Zheng Y, Keller BM, Ray S, Wang Y, Conant EF, Gee JC, Kontos D (2015) Parenchymal texture analysis in digital mammography: a fully automated pipeline for breast cancer risk assessment. *Med Phys* 42(7):4149–4160

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Bikesh Kumar Singh, Ph. D., received his doctorate degree in Biomedical Engineering from National Institute of Technology, Raipur, India. He is currently working as an assistant professor in the Department of Biomedical Engineering of National Institute of Technology, Raipur, India. He has more than 11 years of teaching and research experience. He has published more than 50 papers in various international journals and conferences of repute. He is life member of IE, IETE, and CSI India. He has received IETE Gowri Memorial Award 2016, IEI Young Engineer Award 2012, Chhattisgarh Young Scientist Award-2010. His research interests include medical image processing and analysis, biomedical signal processing, soft computing and machine learning.



Pankaj Jain is currently Junior Research Fellow at Department of Biomedical Engineering, National Institute of Technology, Raipur, India. His research interests include medical image processing and analysis and softcomputing.



Sumit K. Banchhor, Ph. D. received his BE degree in Electronics and Telecommunication engineering from the Pt. Ravishankar Shukla University, Bilai, Chhattisgarh in 2007. He received his M. Tech. degree in Digital Electronics from Chhattisgarh Swami Vivekanand Technical University, Bilai, Chhattisgarh in 2011. Later, he received his Ph.D. degree in 2018 from Department of Electrical Engineering of National Institute of Technology, Raipur, India. His research interests include medical image processing and analysis, and softcomputing.



Kesari Verma, Ph. D., She received her master's degree with first class honors in computer applications from Government Engineering College Raipur, India in 1998 and her PhD degree from Pt RSU Raipur, India in 2007. She is currently an assistant professor in the Department of Computer Applications, NIT Raipur, India. She has around 16 years of teaching and research experience. She has around 40 paper in international journal and conferences. Her research interests include digital image processing and analysis, data mining and machine learning.