

The Improvement of Breast Cancer Diagnosis Rate in Magnetic Resonance Imaging (MRI) using Fusion of Super Pixels and Fuzzy Connectedness

Fatemeh Bakhshi Zade¹, Mehran Emadi^{2*}

1- Faculty of Computer Engineering, Mobarakeh Branch, Islamic Azad University, Mobarakeh, Isfahan, Iran.
Email: Fatibakhshi65@gmail.com.

2- Faculty of Electrical Engineering, Mobarakeh Branch, Islamic Azad University, Mobarakeh, Isfahan, Iran.
Email: emadi.mehran49@gmail.com. (Corresponding Author)

Received: June 2021

Revised: August 2021

Accepted: October 2021

ABSTRACT:

Precise segmentation of tumors in the breast is one of the most significant steps for MRIs and diagnosis tools using computers. Segmentation of the breast tumor is a demanding task due to some factors including partial volume effect, the similarity of the brightness of tumor texture with other surrounding non-tumor textures, variety in shape size and location of the tumor in different patients. Due to its vitality, the process of segmentation is carried out manually by specialists and its disadvantages are long computation time, and high cost. To overcome these issues, algorithms are required to segment images with high accuracy and no need for user intervention. This study presents a new method based on fuzzy connectedness algorithm and super pixels for tumor segmentation in magnetic resonance imaging (MRI). The proposed method is applied to a dataset built by the respected researchers on Matlab. The suggested method has been compared using two commonly used methods of clustering and morphological operators in tumor segmentation in magnetic resonance imaging (MRI). Mean average precision of 98.33 and the Dice similarity coefficient of 98.06 signifies the prominence of the suggested method in comparison with other methods compared using clustering algorithm 90.33 and morphological algorithm 91.83.

KEYWORDS: Tumor Segmentation, Magnetic Resonance Imaging (MRI), Fuzzy Connectedness, Super Pixel Algorithm.

1. INTRODUCTION

Cancer is a sort of disease involving the abnormal growth of cells in one part of the body. The generated cells are accumulated and form a mass or tumor. Cancer can be categorized into two types of benign and malignant. In a benign tumor, cancer cells are immovable but in a malignant tumor, these cells are spread to other parts of the body and allow the regrowth of cancer cells. Breast cancer is one of the most prevalent cancers which is often seen among women. Determining the danger of a tumor in the breast is done considering criteria such as tumor size, its spread through neighboring organs as well as its development in the breast. The direct connection between the danger and the development rate of a tumor emphasizes the importance of early diagnosis of the existence of the tumor or the chance of tumor formation. Nearly 200,000 women and 1,700 men develop breast cancer each year, with more than 40,000 women and 450 men dying. In fact, breast cancer is the

abnormal growth of breast cells which causes the tumor to be formed in the breast. Age is one of the most significant factors in determining breast cancer. In addition to age, race, geographical location, family background, etc are among factors affecting the chance of developing breast cancer. One in eight women in the United States and one in six women in Europe have breast cancer. Despite an increase in the development rate of breast cancer, its death rate has dropped which can be attributed to new treatment methods and new diagnosis methods such as new MRI systems. One of the diagnosis methods for breast cancer is carrying out MRI tests. Over recent years, new methods for identification of breast cancer tumors based on image processing have caused the need for human resources and subsequently human error to be decreased. More importantly, it has resulted in the reduction of treatment costs and played a significant role in medicine as a profession.

MRI can be utilized for early diagnosis of cancer before any symptom appears or a tumor is felt in the breast. Nonetheless, examining magnetic resonance imaging (MRI) and diagnosis of benign and malignant tumors is challenging for a radiologist. Many researchers believe the automatic analysis of magnetic resonance imaging (MRI) increases the rate of early diagnosis. One of the methods for automatic analysis of magnetic resonance imaging (MRI) used for cancer diagnosis is employing image processing and machine learning techniques of which several methods with high diagnosis rates have been provided and some of them are introduced in [1] and [2]. However, there are some challenges which are mentioned below:

- Changes in the gray areas are not noticeable in different parts of the image which causes the tumor segmentation only through gray areas to be difficult.
- A tumor is not always evident.
- High-frequency elements and different levels of noise exist in magnetic resonance imaging which roughly makes up 10% to 15% of pixel existence.

Due to the little variations in the gray areas of different parts of magnetic resonance imaging (MRI), segmentation of the breast including the tumor is difficult. Therefore, noise and extra information are removed in the preprocessing phase and normally methods such as wavelet transform, multi-resolution analysis, region growing method, median filter, adaptive median filter, Wiener filter and etc [8]. are employed to promote and improve the contrast of images for segmentation to be carried out in a higher quality in the following steps. In the segmentation step, the image is divided into several zones without overlap to select respected objects. This step is of great significance because the tissues suspected to containing tumors are separated from the main tissue. Suspected zones usually are brighter than other areas. Although these boundaries are completely random, their orderly shape and monotonous density can be identified compared to other tissues. Since the properties of the masses are different from one image to another, the segmentation phase is complex and difficult [9]. Median filter, tissue analysis, and nearest neighbor algorithm are methods that can be used in this phase. Based on the results obtained from the previous steps including preprocessing phase, magnetic resonance imaging (MRI) are segmented based on the cancer being benign or malignant and feature extraction in the identified zone in MRI is done as the last step of classification and the identified zone is classified into benign or malignant. Methods used in this classification in this stage are artificial neural network, support vector machine, nearest neighborhood, fuzzy clustering, etc. Symptoms of breast cancer in magnetic resonance imaging (MRI) can be categorized into two

main categories: 1. Mass 2. Micro calcification. These two signs exist in magnetic resonance imaging (MRI) however, are hard to detect and require the radiologist to be highly concentrated. Therefore, this study insists on improving the quality of magnetic resonance imaging (MRI) of a standard database to reach a favorable quality for detecting such phenomena. Afterward, a method is employed for the precise detection of these tumors. In the third step, the desirable features of the Super pixel algorithm for training a fuzzy connectedness to identify the cancerous tumors identified from each image are presented in the following, methods including super pixels and fuzzy connectedness will be addressed in section 2 of this study. Section 3 will present the suggested method and the proposed method will be evaluated and in the end, the result of the study will be presented.

1.1. LITERATURE REVIEW

This section provides information about fuzzy algorithm and super pixels and relative works and advantage and disadvantage of them.

1.1.1. Relative Works

Several studies have been done on this matter and each of them has got its advantages and disadvantages. Diagnosis of breast cancer in digital magnetic resonance imaging (MRI) is done in four steps. These four steps are: preprocessing, segmentation, feature extraction, and classification [1]. The output of this process helps radiologists obtain more precise results for breast cancer diagnosis. Santiga et al, have employed a combination of effective methods for segmentation of medical images for identification of micro calcification in breast. Although the suggested method follows the complete pattern of classification process, there is a main difference which is utilizing clustering in segmentation. After clustering, statistical features are extracted from each clusters. Eventually, these feature vectors are classified using Support Vector Machine. The only innovation in the proposed method is using clustering in the segmentation step and feature extraction of these areas [3]. Marina et al, made use of the Bayesian neural network for the segmentation of micro calcification. The suggested method makes a logical connection between low-level features and high-level features (for instance the relation between calcium fine particles and large volumes in the image). This connection occurs using probability distribution in a Bayesian neural network. The generated logical connection is utilized for training of the Bayesian neural network. Eventually, micro calcifications are identified [4]. In their study, Gili et al, have proposed a new method searching for fine particles in breast tumors. In the suggested method, the

topology of each calcium fine particle is first obtained using fuzzy clustering. Afterward, a graph is created from all the obtained fine particles and a general graph is created using features extracted from fine particles. Multi scale topological feature vector is generated from graphs of micro calcifications and finally, is classified using the k-nearest neighbor classifier. The proposed method is simulated on the database built by the respected researcher. The accuracy of the aforementioned method is 95%. The defect of the proposed method is that region of interest for selection of micro calcification has opted manually [5]. In the research done by Haili et al, a combination of k-means algorithm and fuzzy c-means clustering is used for the detection of breast cancer. In the preprocessing phase, Median filter and discrete wavelet transform are used to remove noise and high-frequency elements. Then, texture features are extracted and breast cancer diagnosis is done employing fuzzy k-means algorithm and fuzzy c-means clustering [6]. In 2012, Chen et al, suggested the combination of particle swarm optimization (PSO) and SVM for the detection of cancer tumors. The suggested method makes use of features based on image texture such as scattering wavelet as well as a neural network. Afterward, relevant features are selected through PSO. To better balance, the local and global search in the PSO algorithm, the time-varying acceleration coefficient is exploited in the suggested method. Furthermore, optimization of SVM parameters including the windows used in this classifier is done to increase the precision of the classification. In this research, comparing the accuracy of the proposed method with other methods, the proposed method is not performed with SVM combination and a basic variant of the PSO algorithm [7].

1.1.2. Fuzzy Algorithm and Super Pixels

Fuzzy logic was first introduced by Professor Lotfi Zadeh in his proposal of "Fuzzy Set Theory-Information and Control" in 1965 and was developed and put into practice in the 70s. production of fuzzy controllers for real systems was one of the biggest events of the decade. Fuzzy logic is one of the successful applications of fuzzy sets in which variables are linguistic rather than numerical. Fuzzy logic is opposed to binary or Aristotelian logic which evaluates everything as yes or no, black or white, zero or one. This logic changes between zero and one. Fuzzy connectedness system is a knowledge-based system that takes linguistic variables (good, average, poor) as input and converts them to mathematical equations using a set of rules and is able to extract the correct conclusion. The whole fuzzy connectedness system is based on "if..then" which makes different connections between inputs and outputs. Fuzzy set theory is a strong

tool for dealing with uncertainty caused by ambiguity. Although fuzzy systems describe uncertain phenomena, the fuzzy theory, itself, is precise. Super pixel algorithm is a very useful technique for classification and image segmentation. In super pixel algorithm, the image is divided into smaller areas using internal boundaries. Although super pixel algorithm is difficult to be thoroughly described, its important characteristics can be named as follows:

1. Super pixels must be interconnected in a way that the set of all super pixels makes the whole picture.
2. The preprocessing phase is mandatory when using super pixels.
3. The compression in super pixels should be chosen in a way that is suitable for its particular processing applications.

This algorithm is becoming an applied algorithm in diagnostic studies such as tumors. This algorithm is increasingly becoming popular for use in machine vision and image processing applications.

2. The Suggested Method

Since the process of tumor identification in magnetic resonance imaging (MRI) is a costly medical procedure, it seems that providing a cost-effective method for tumor identification and subsequently its removal is necessary. In this regard, presenting new methods based on machine vision as well as image processing for precise identification of tumor boundaries in the breast reduces the cost substantially. A new method of segmentation is presented for accurate identification of the tumor in the breast in magnetic resonance imaging (MRI). In the following, the details of the proposed method are further discussed. Generally, three stages are considered for tumor segmentation and identification which are: preprocessing stage, segmentation stage using a combination method of super pixels and fuzzy connectedness and the stage of displaying the identified zone. In each of these stages, processes are done which have resulted in the improvement of accuracy and increase in the Dice similarity coefficient. Figure 1 shows a block diagram of the proposed method. The proposed method is described in full below. The images used in this study for identification of breast tumor are of database built by the respected researcher. The average age of the patients is 64.2. Whilst the youngest patient is 18 years old, the oldest one is 85 years old. Table 1 shows the characteristics of this database. Images of this database, as well as existing magnetic resonance imaging (MRI), have low quality. The low quality of these images is mainly because of their low contrast and is majorly caused by magnetic resonance imaging (MRI) devices. Since distinguishing among tissues in the image including Adipose tissues,

glandular tissues, and lymphatic glands is a difficult task, enhancing the image quality seems to be necessary. Gaussian filter is utilized to enhance the image quality in this study.

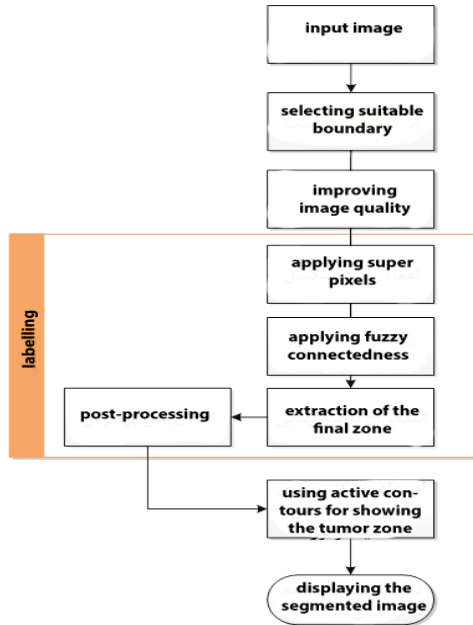


Fig. 1. Block diagram of the suggested method.

Gaussian filter is of low-pass filters which changes high-frequency pixels by transition through low-frequency pixels. A low-frequency pixel is a pixel that has a low brightness difference with its neighboring pixels, in other words, it's calm and still spots in the image. In contrast, a high-frequency pixel is a pixel that has a low brightness difference with its neighboring pixels such as edges and noises. Then Sobel filter is applied upon the image to improve the initial edges. The generated image is an image with better quality (higher contrast) and better edges for identification compared to the original image. Figure 2 displays the improved image as well as its histogram.

Table 1. The collected database.

Database	Database built by the researcher
Number of images	30
Resolution	1024*1024
Format	pgm

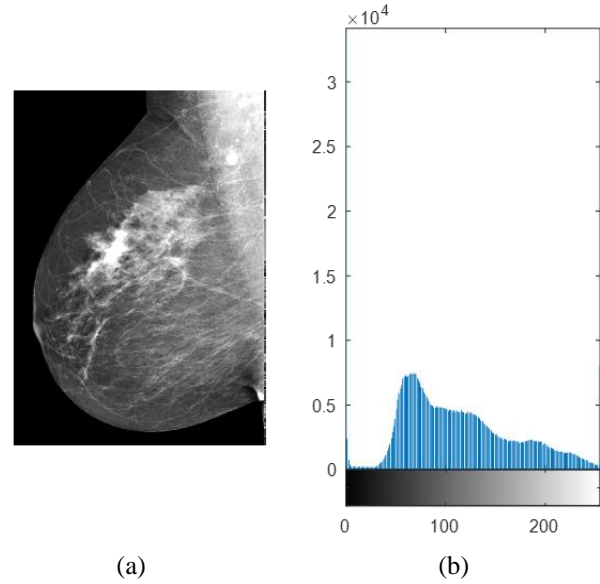


Fig. 1. a) Improved image of the database. b) Relevant histogram.

2.1. Fuzzy Connectedness

The training and testing phase is carried out with the proposed fuzzy connectedness system which has a high capability in training, construction, and classification and also has the advantage of allowing fuzzy rules to be extracted from specific numerical information or knowledge and adaptively, builds a rulebase. Furthermore, it can regulate the complicated transformation of human intelligence into fuzzy systems. The adaptive suggested method receives pre-processed data as input and considers 70% of the samples as training data and 30% as test data. Test and training data all are randomly selected. It should be noted that the proposed fuzzy connectedness system is trained 1000 times per run and is shown in figure 3.

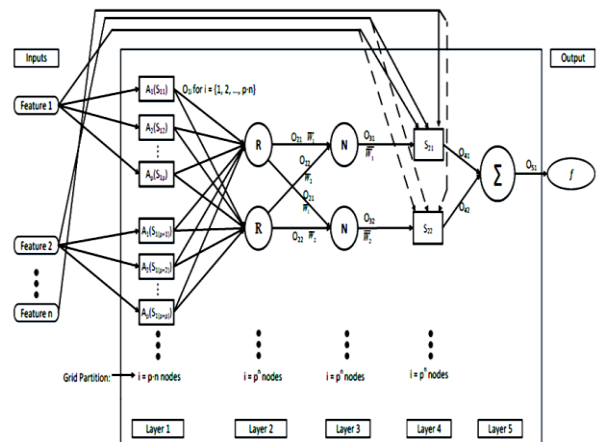


Fig. 3. Overview of the fuzzy connectedness system used in this study.

After the training of input data using various learning functions, the test data are evaluated. The values of different parameters of the adaptive fuzzy connectedness system are displayed in table 2.

Table 2. Range of different parameters in examining the suggested fuzzy connectedness system.

Parameter	Range
membership functions= mfType	gbellmf
	trimf
	trapmf
	gaussmf
	gauss2mf
	pimf
	dsigmf
psigmf	
Train Data	70%
Test Data	30%
Epochs	1000

Different learning functions will be employed for modelling using adaptive fuzzy connectedness system which are shown in table 3.

Table 3. Learning functions in ANFIS .

Learning functions names	Learning functions in Matlab	NO
Gaussian bell MF	gbellmf	1
Triangular MF	trimf	2
Trapezoidal MF	trapmf	3
Gaussian curve MF	gaussmf	4
Gaussian combination MF	gauss2mf	5
Pi MF	pimf	6
Difference between 2 Sigmoid MFs	dsigmf	7
Product of 2 Sigmoid MFs	psigmf	8

For examining this method, different parameters have been utilized and their values have been studied at different intervals. To examine closely, these parameters are put into a loop considering the desired interval. The parameter of the number of neurons in the neural network is a Radial basis function and has been used to estimate data. To examine more closely, data are once again randomly selected per run of the application so that the different conditions would be investigated. The range of parameters used in the radial-basis function of the neural network is chosen concerning table 4. Figure 5 displays the generated fuzzy functions.

Table 4. Range of different parameters used in radial-basis function

Parameter	Range	The best selection
Goal	0-0.20	0.0
Neuron	10-1	Different results obtained using different methods
Train Data	70%	
Test Data	30%	
Epochs	1000-1	

In the complementary processes, the obtained image is pre-processed using morphological operators. Soft areas become connected and are enhanced and areas, where most pixels are connected, are created. Then, the labeling of the areas begins. This image is created to create brighter tumor areas. Figure 6 shows the final binary image.

2.2. Active Contour Algorithm

Active contour algorithm is a useful tool for the identification and display of images. The active contour model directs contour toward the removal boundaries based on the energy of edges. If a powerful edge does not exist in the image, target identification comes to a dead-end. Since the suggested method puts its effort into improving the boundaries of tumor edges, therefore, the respected contour is formed around the tumor and accurately shows the boundaries including the edges of the tumor. Figure 7 displays a tumor identified in the lung image. The obtained flexible boundaries indicate the superiority of the proposed method.

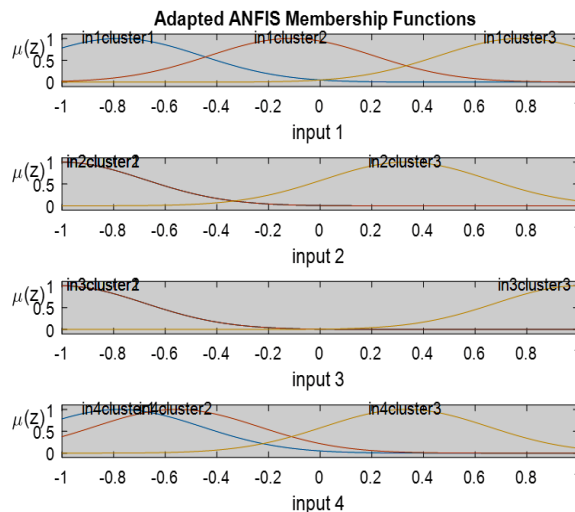


Fig. 4. Generated fuzzy membership functions.

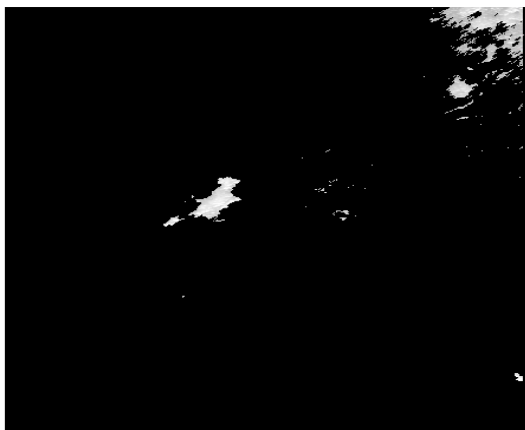


Fig. 5. Final output image.

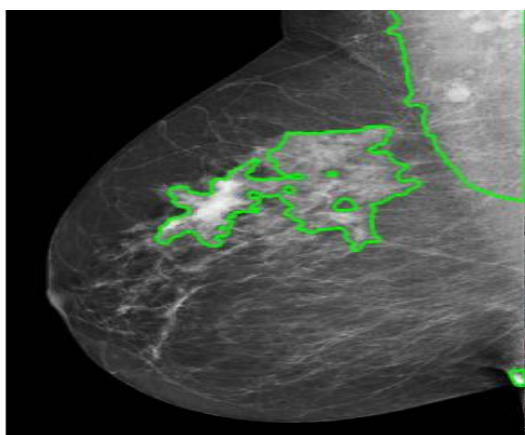


Fig. 6. Contour demonstrates the identified tumor in the breast.

3. Evaluation of the suggested method

Breast tumor detection is one of the most significant steps for MRIs and diagnosis tools using computers. Segmentation of the breast tumor is a demanding task due to some factors including partial volume effect, the similarity of the brightness of tumor texture with other surrounding non-tumor textures, variety in shape size and location of the tumor in different patients. Due to its vitality, the process of segmentation is carried out manually by specialists and its disadvantages are long computation time, and high cost. To overcome these issues, algorithms are required to segment images with high accuracy and no need for user intervention. The final step in a breast tumor segmentation system is the calculation of precision, sensitivity and dice similarity coefficient. The necessary explanations regarding the values of equations 1 to 3 are given. Table 5 describes these parameters.

Table 5. Definition of the evaluation parameters.

Parameter number	Definition of the parameter	Symbol
1	It is tumor and it is diagnosed	TP
2	It is tumor and it is not diagnosed	FN
3	It is not tumor but it is diagnosed as tumor	FP
4	It is not tumor and it is not diagnosed as tumor	TN

$$Precision = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (1)$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100 \quad (2)$$

This similarity criterion is utilized to evaluate and also confirm the superiority of the suggested method. This criterion is achieved through equation 3.

$$Dice = \frac{2 \times TP}{2 \times TP + FN + FP} \times 100 \quad (3)$$

3.1. Qualitative comparison

Number The proposed method aims to identify the precise boundary of a tumor in the breast based on a combination method of super pixels and fuzzy connectedness methods which is compared to methods based on morphological operators as well as clustering algorithms. Figure 8 shows one of the images of the database used in the study which is segmented using the proposed method. Figure 9 shows the same image which is segmented using the method based on the morphological algorithm. The clustering method is one of the well-known morphological image segmentation methods. Figure 10 displays the segmented image using this method. A qualitative comparison of these three methods confirms the prominence of the suggested method. Boundaries of tumor tissue in the breast are detected easily and with high precision. The extracted boundaries of the tumor are really precise and separate the tumor tissue from the rest of the tissues within the breast while the clustering method does not identify the boundaries of some parts of the tumor. The result obtained from morphological operators is even poorer.

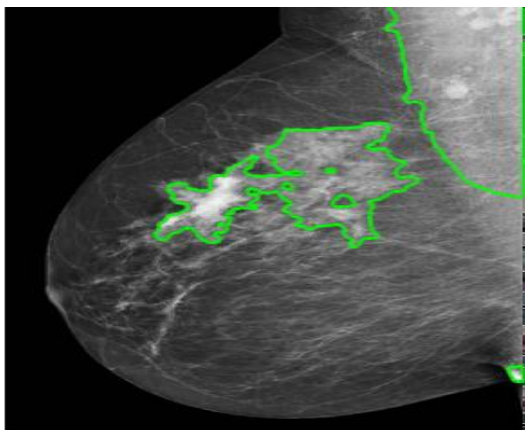


Fig. 7. Extraction of tumor zone in the breast using the suggested method.

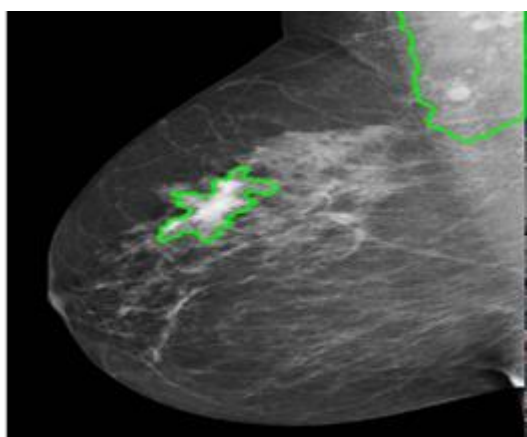


Fig. 8. Extraction of the tumor zone in the breast using methods based on morphological operators.

3.2. Quantitative evaluation

For evaluation of the suggested method, evaluation criteria such as Dice similarity coefficient, sensitivity, and precision have been used. Confidence interval and standard deviation confirmed the obtained result. Using confidence interval means that tumor boundaries are identified using 300 images as input with 300 times of repetition. The purpose of this repetition is for the confidence interval to be calculated for the obtained result. This criterion is intended to prove the quality of the suggested method. The confidence interval is a range that might contain the true value of a parameter. The fewer samples we obtain, the wider this range which reduces the probability of achieving the true value. When the sample size increases to a great extent, we have more certainty to say that the number we have achieved is closer to the truth and CI is narrower. The dice coefficient in table 6 shows 30 segmented images from the database built by the respected researcher in this

study. This table compares the suggested method, clustering-based method as well as the method based on the clustering operators. The advantage of the proposed method in the dice coefficient is noticeable. Mainly, the suggested method is a combination of fuzzy and neural approaches which has caused the tumor area in the breast to be more accurately retrieved and identified. Figure 11 displays a comparison between confidence interval and the mean of 30 images in the dice coefficient. The narrower the interval, the more precise is our estimate. In figure 11, the confidence interval of the suggested method is far less than of other methods such as the clustering method as well as morphological operators which shows the superiority of the suggested method. Although the method of morphological operators has a better result in the dice similarity coefficient than the clustering method, the mean achieved in all the images of the database is equal.

Table 6. Comparison of the dice coefficient in the suggested method, clustering method and method based on morphological operators.

Number of Image	Morphological Operators method (percent) [11]	Clustering method (percent)[12]	The suggested method (percent)
Mean	91.13	91.13	98.67
STD	5.285	1.097	0.7397
Confidence interval	89.23-93.06	90.39-91.87	97.79-98.34

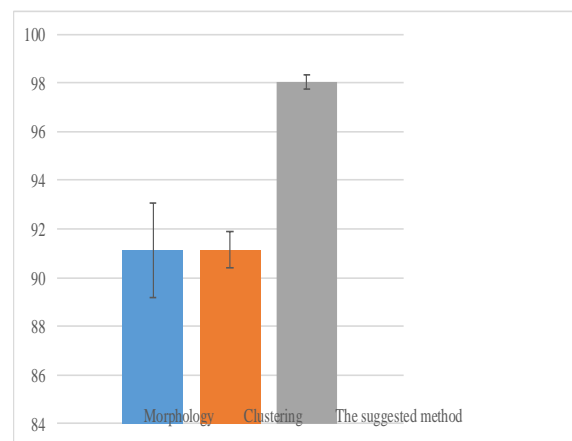


Fig. 9. Comparison of the confidence interval in the suggested method, clustering method and method based on morphological operators.

Table 7. Comparison of precision in the suggested method, clustering method and method based on morphological operators.

Number of Image	Morphological Operators method (percent) [11]	Clustering method (percent)[12]	The suggested method (percent)
Mean	91.83	90.33	98.33
STD	2.930	3.0775	0.6609
Confidence interval	70.93-90.92	18.48-89.91	06.58-98.98

Table 7 compares the precision of the suggested method, clustering method, and morphological operators. The suggested method achieved the best results in breast tumors. The mean precision of the morphological operators is far better than that of the clustering method. However, it should be noted that the computational complexities of the clustering method are by far more than that of the morphological operators. Figure 12 compares the precision criterion in the suggested method to the clustering method and morphological operators. Mean and confidence interval of the suggested method is compared to the other two methods in figure 12. In the suggested method, the confidence interval is much narrower than that of the other two methods which implies the superiority of the suggested method.

4. CONCLUSION

In this section of the study, the proposed new method based on the fuzzy connectedness algorithm as well as super pixels is simulated on the database built by the respected researcher. 30 images of 30 different individuals are selected from this database for evaluation and afterward, are segmented using the suggested method. The suggested method was compared to methods based on morphological operators and clustering methods in precision criteria and dice similarity coefficient. For more evaluation, the confidence interval was also calculated in precision criteria and dice coefficient. The proposed method with a mean precision of 98.06 and an average dice similarity coefficient of 98.06 compared to mean precision and average dice similarity coefficient of 90.33 and 91.33 respectively in clustering method and mean precision and average dice similarity coefficient of 91.83 and 91.13 respectively has indicated its prominence over other methods. Furthermore, the confidence interval in the two examined criteria is much narrower in the proposed method in comparison with the other two methods which confirms the superiority of the proposed method.

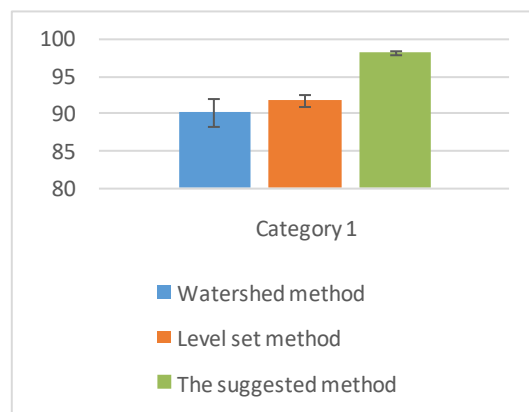


Fig. 10. The comparison of confidence interval of the suggested method and morphological operators and clustering methods.

REFERENCES

- [1] B. Priyanka and D. Kulkarni, "Digital mammography: a review on detection of breast cancer," *Int J Adv Res Comp Common Eng*, Vol. 5, PP. 386-390, 2016.
- [2] E. Aličković and A. Subasi, "Breast cancer diagnosis using GA feature selection and Rotation Forest," *Neural Computing and Applications*, Vol. 28, No. 4, PP. 753-763, 2017.
- [3] R. Sangeetha and K. S. Murthy, "A novel approach for detection of breast cancer at an early stage by identification of breast asymmetry and micro calcification cancer cells using digital image processing techniques," in *Convergence in Technology (I2CT)*, 2017 2nd International Conference for, 2017, PP. 593-596: IEEE.
- [4] M. Velikova, I. Dutra, and E. S. Burnside, "Automated diagnosis of breast cancer on medical images," in *Foundations of Biomedical Knowledge Representation: Springer*, 2015, PP. 47-67.
- [5] Z. Chen, A. Oliver, E. Denton, C. Boggis, and R. Zwigelaar, "Classification of microcalcification clusters using topological structure features," in *Medical image understanding and analysis*, 2012, PP. 37-42.
- [6] D. Hailey and D. Marshall, "The place of magnetic resonance imaging in health care," *Health policy*, Vol. 31, No. 1, PP. 43-52, 1995.
- [7] H.-L. Chen, B. Yang, G. Wang, S.-J. Wang, J. Liu, and D.-Y. Liu, "Support vector machine based diagnostic system for breast cancer using swarm intelligence," *Journal of medical systems*, Vol. 36, No. 4, PP. 2505-2519, 2012.
- [8] 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, P. 47-51, 2013.
- [9] V. Raman, P. Sumari, H. Then, and S. A. K. Al-Omari, "Review on Mammogram Mass Detection by Machine Learning Techniques," *International Journal of Computer and Electrical Engineering*, Vol. 3, No. 6,

- PP. 873-876, 2011.
- [10] Y. Tan, C. Shuai, L. Jiao, and L. Shen, "Adaptive neuro fuzzy inference system approach for urban sustainability assessment: A China case study," *Sustainable Development*, Vol. 26, No. 6, PP. 749-764, 2018.
- [11] P. Vikhe and V. Thool, "Morphological operation and scaled Rényi entropy based approach for masses detection in mammograms," *Multimedia Tools and Applications*, Vol. 77, No. 18, PP. 23777-23802, 2018.
- [12] T. Padmavathy, M. Vimalkumar, and D. Bhargava, "Adaptive clustering based breast cancer detection with ANFIS classifier using mammographic images," *Cluster Computing*, PP. 1-10, 2018.