Asian Journal of Research in Computer Science

9(1): 23-38, 2021; Article no.AJRCOS.68284
ISSN: 2581-8260

Medical Images Breast Cancer Segmentation Based on K-Means Clustering Algorithm: A Review

Noor Salah Hassan[1], Adnan Mohsin Abdulazeez[2], Diyar Qader Zeebaree[2] and Dathar A. Hasan[3*]

[1]Duhok Polytechnic University, Kurdistan Region, Iraq.
[2]Research Center, Duhok Polytechnic University, Kurdistan Region, Iraq.
[3]Shekhan Technical Institute, Duhok Polytechnic University, Kurdistan Region, Iraq.

Authors' contributions

This work was carried out in collaboration among all authors. Author NSH prepared and reviewed the first draft of the manuscript. Authors AMA and DQZ managed the analyses of the study. Author DAH managed the literature searches. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJRCOS/2021/v9i130212
Editor(s):
(1) Dr.Shivanand S.Gornale , Rani Channamma University, India.
Reviewers:
(1) Mohamed Abdel-Nasser, University Rovira I Virgili, Spain.
(2) Athiraja A, Rathinam Technical Campus, India.
Complete Peer review History: http://www.sdiarticle4.com/review-history/68284

Review Article

Received 12 March 2021
Accepted 17 May 2021
Published 24 May 2021

ABSTRACT

Early diagnosis is considered important for medical images of breast cancer, the rate of recovery and safety of affected women can be improved. It is also assisting doctors in their daily work by creating algorithms and software to analyze the medical images that can identify early signs of breast cancer. This review presents a comparison has been done in term of accuracy among many techniques used for detecting breast cancer in medical images. Furthermore, this work describes the imaging process, and analyze the advantages and disadvantages of the used techniques for mammography and ultrasound medical images. K-means clustering algorithm has been specifically used to analyze the medical image along with other techniques. The results of the K-means clustering algorithm are discussed and evaluated to show the capacity of this technique in the diagnosis of breast cancer and its reliability to identify a malignant from a benign tumor.

**Corresponding author: E-mail: dathar.hasan@dpu.edu.krd;*

Keywords: Medical images; breast cancer; pre-processing; segmentation; clustering; K-means.

1. INTRODUCTION

In recent periods the statistics have shown that breast cancer shows that the number of new patients is getting worse every year, the number of people affected by this disease is rising at an unprecedented pace [1]. Generally, in recent years, medical image handling has been an exciting and difficult field for researchers, with the emergence of multimodal imaging methods such as mammography, ultrasound, computed breast tomography, and some modern methods [2-4]. The most effective imaging technique, recommended by the radiologist is medical imaging [5]. Therefore, it is important to examine many medical images. Besides, the expertise of radiologists and the reliability of the diagnostic images to be examined play a key role in the accuracy of the diagnosis. The best way to decrease mortality breast cancer should be identified and diagnosed in the early stages [6]. The images first need to be pre-processing for increasing the accuracy of the images by removing the non-related and redundant components of the images to make them ready for further processing. Therefore, the image will be in the segmentation step that is a significant step, can make it easier to analyses the images. Then ROI is also obtained either manually by professionals or using image segmentation techniques automatically in medical imaging applications [7,8]. One of the algorithms used in medical images is clustering that is unsupervised learning based on the form of dissimilarity of the grouping of datasets into subsets data [9,10]. These dissimilarities of clustering can as possible as to compute predicted issues, as predicted cancer [11]. It would have improved to recognize that cancer is the disease induced by the out-of-control growth of uncontrolled cells or cells in the section of the body [12,13]. This paper aims to find the best technique to detect breast cancer in the early stage. One of the clustering techniques, the k-mean algorithm using operation frequently. However, the k-mean clustering is certainly very common among multiple clustering techniques that already exist because of its capacity and effectiveness in clustering the data [14-16]. The rest of the paper is outlined as follows, section 2 contains the methodology and the discussion of each method in a special part, section 3 consists of a literature review of the subject, in section 4 the findings are discussed and finally, in section 5 the conclusion of the review.

2. METHODOLOGY

2.1 Medical Images

The imaging technology in medicine eases the help of seeing the inner portions of the body for diagnostic [17]. Medical images involve several types of images that vary from one to another [18], [19]. Mammography is a type of medical image that is a procedure that helps to diagnose breast cancer early [20]. While mammography has been known as the best technique, it presented to be very difficult to find classifications and spread of cancer in the woman body. For the accurate reading of mammography. Expert radiologists are needed to predict breast mass and type of mass. In this study, partition-based techniques for data mining have been widely used, one type of techniques was k-mean clustering [21,22]. The image of mammography helped to include certain interventions to help doctors determine if their disease mass (normal or abnormal) [23]. This study aims to identify cancer in the breast and specify the affected area by separating the images into clusters based on their difference in strength and color, some appropriate stages of the clustering process have established the cancer area [24]. The results of the clusters are then confirmed, by use classification techniques [25]. The major variations in the cell structure of normal, abnormal cells are shown in Fig. 1. Abnormal cells expand uncontrollably and begin to split the cells where the form and size of normal cells are precise. Using the abnormal chromosome number and size of the nucleus, abnormal cells will be morphologically distinct from regular cells as shown in Fig. 1. Mammograms do not cure breast cancer, but they can save a patient's life by detecting breast cancer as early as possible. Mammography helps to detect and thereby increase the survival rate of invisible cancers [27,28].

Ultrasound Also known as sonography, ultrasound is also a type of medical imaging as shown in Fig. 2. Healthcare practitioners also use ultrasonic equipment [29]. The use of ultrasound images in medical diagnosis is well known, according to its non-invasive nature, low cost, the capacity to yield real-time images, and continuous enhancement in image quality and accuracy [30].

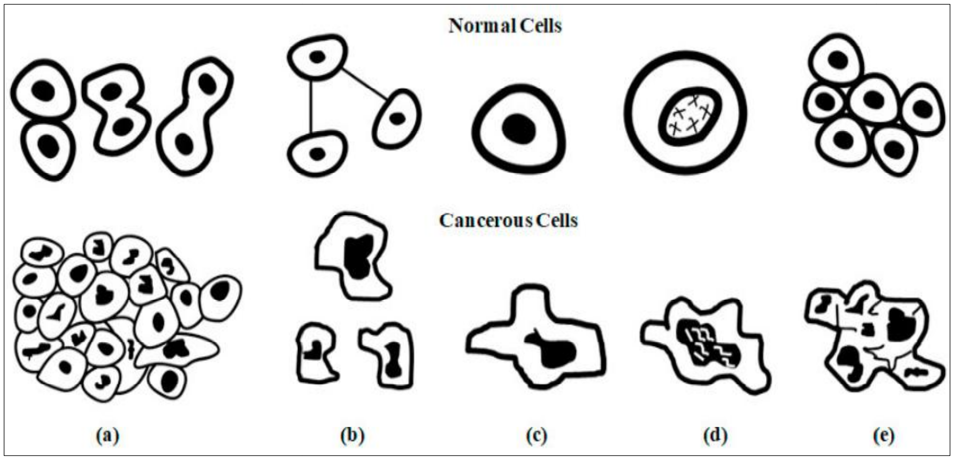

Fig. 1. Type of the cells structures normal and abnormal cells [26]

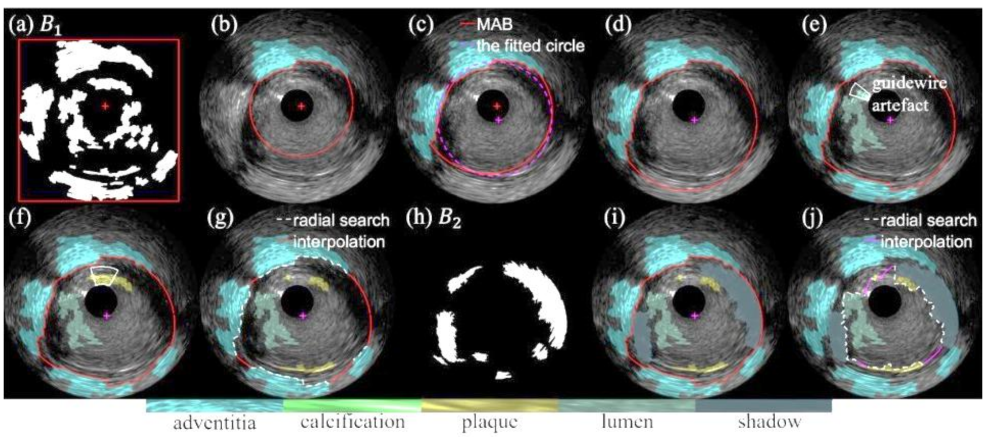

Fig. 2. Ultrasound medical images example [29]

2.2 Breast Cancer

Breast cancer is the most diagnosed cancer worldwide [31]. According to Pal et al. [32], that more than a million women have breast cancer each year and 400,000 of those cases lead to death. This fatal disease not only affects emerging countries but also developed countries [33]. To find a cure, it is important to correctly diagnose the disease and very quickly and treat it based on the type of symptoms that have arisen. Generally, the most effective approach suggested by radiologists for this disease is medical images.

2.3 Pre-Processing

The essential purpose of pre-processing is to enhance the consistency of the image. It prepares the images to be ready for further processing by removing or minimizing in the medical images the nonrelated and redundant components. These methods are classified into the following categories, first to cleaning the data, second to data incorporation, third to data transformation, and reduction data nonrelated, to enhancement the images and reduce noises [34]. The “Region of Interest (ROI)” is the measure involved in the process of the inverse approach and boundary detection method [35]. The accuracy of pre-processing determines, therefore, the expectation of the remaining steps being successful, such as segmentation, classification, etc. [36].

2.4 Segmentation

In the processing of images, image segmentation is the method by separating an image and divided into multiple regions of segmentation that have similar properties. where each of the region's pixels is comparable to the unique characteristics or particular features measured, such as color, strength, or texture. When operating on a large amount of data set created by medical modalities, many segmentation methods are computationally costly [7,37]. In the clinical setting, segmentation of image data before or during the procedure must be quick and accurate. Picture segmentation is also used to segment cancer, brain structures, blood vessels, and bones in medical imaging [38]. Over the past few years, various image segmentation techniques have been improved and further research work has been carried out on fragmentation image [39,40]. So is manual fragmentation of medical images a repetitive, time-consuming, and complex task by an experienced radiologist, it is, however, particularly important to the development of medical image modalities, that need to be examined. Furthermore, it is becoming increasingly important to upgrade image segmentation techniques. In the segmentation process, it is important to delineate and remove the anatomical structure or the region under review in discrete forms by using algorithms that are more precise and need to reduce user interaction, particularly for medical images [41]. However, the clustering-based methods are one of the segmentation methods that may be supervised or unsupervised. They seek to classify some image pixels into multiple clusters such that each cluster pixels have a high intra-class of similarity pixels of the same class and a low inter-class of similarity pixels of another class [42]. The similarity is defined in terms of distance measurements, such as Euclidean distance measurement [43]. A variety of clustering-based approaches have been used in the literature, including K-mean, Fuzzy C-mean, and Markov Random Fields (MRF) [44].

2.5 Clustering

The main purpose of clustering is to categorize a set of objects into relevant groups [45]. The clustering objects is based on the correspondence calculation between pairs of objects using the distance function [46]. Therefore, the product of clustering is a set of clusters, in which the object in one cluster is more like each other than the object in another cluster [47]. The study of clustering has been widely used in numerous applications, including segmentation of medical images, analysis of information, evaluation, and processing of images. In certain applications, clustering is often called image segmentation. The process of grouping a set of physical or abstract objects into groups of similar objects is called clustering [48,49]. With clustering, it is possible to identify dense and sparse regions and thus discover general distribution patterns and interesting similarities between data attributes. Measurement space clustering may also imply the similarity of image regions and can be used for segmentation [50-52].

2.6 K-means Clustering Algorithm

K-means is a technique or method of clustering which could group huge amounts of data with processing time that is relatively fast and more effective [53]. In contrast, the k-means algorithm has a weakness that depends on the initial value cluster that determines the center [54,55]. K-means clustering, give trial results in the form of optimal topical solutions. However, similarities or proximity between the data is expected from the testing process. Thus, this can be divided into multiple clusters, where a high degree of similarity can be obtained across the cluster points [56]. The K-Means algorithm is also multi-sided, according to (celebi et al. 2013) k-means are too simple to modulate at each stage of the process, easy to measure the distance, and are based on requirements for iteration termination. K-mean cluster is a local optimizing, so k-means is sensitive to the initial data point collection from the midpoint of each cluster [57]. The purpose of these adjustments is to achieve the best accuracy and fastest convergence. Moreover, the selection of the initial position from the midpoint of a cluster will restrict the K-mean cluster algorithm to the optimal location [58]. The k-mean cluster method also will choose the style up to the k as the start point from the center randomly [59]. The number of iterations with the centroid cluster will be affected by the initial centroid cluster that is randomly set [60]. Therefore, can be fixed to obtain higher execution by determining the centroid cluster in the high initial data points [61]. Since k-mean clustering is typically implemented, the data point set {x1, x2, ..., xn} is grouped into k clustering. It has high-performance computing and can handle multi-dimensional vectors [62,63]. It thus reduces the distortion measure by reducing the cost function like:

$$\left| x_i^{(j)} - c_j \right|^2 \tag{1}$$

Where xi(j) is a selected measure of the distance between the data point and the cluster centre cj is a measure of the distance between the (n) data points and their respective cluster centres [64]. The algorithm consists of the elements below:

1. Place the k points in the space represented by the clustered objects, these points represented the initial group centroids.
2. Assign the group of each object to the category which has the closest center.
3. Recalculate the places of the k centroids, when all objects will be allocating.
4. Repeat steps 2 and 3 until the centroids no longer move.

This induces a division of the objects into groups from which the metric to be minimized can be calculated [65-67]. There are two known characteristics of the K-means clustering algorithm: one is that it requires a predefined cluster initial number k centroid, as a prerequisite parameter for clustering, but usually, without prior knowledge, we do not know the best initial number of clustering that a data set can generate. The other characteristic is that connecting each point to the nearest cluster [68,69]. The flowchart of the K-means clustering algorithm is shown in Fig. 2 [70,71].

2.6.1 Fuzzy C-mean

Fuzzy c-means (FCM) is an unsupervised clustering technique that is applied to a wide variety of feature analysis, clustering and classification design issues. FCM is widely applied in medical images diagnosis, form analyze and goal recognition [72]. FCM is a data clustering technique. in which a data set clustered into (n) clusters with each data point linked to each cluster in the dataset and which will have a high degree of belonging connection to that cluster, and another data point which will have a low degree of belonging to that cluster far from the center of that cluster [73].

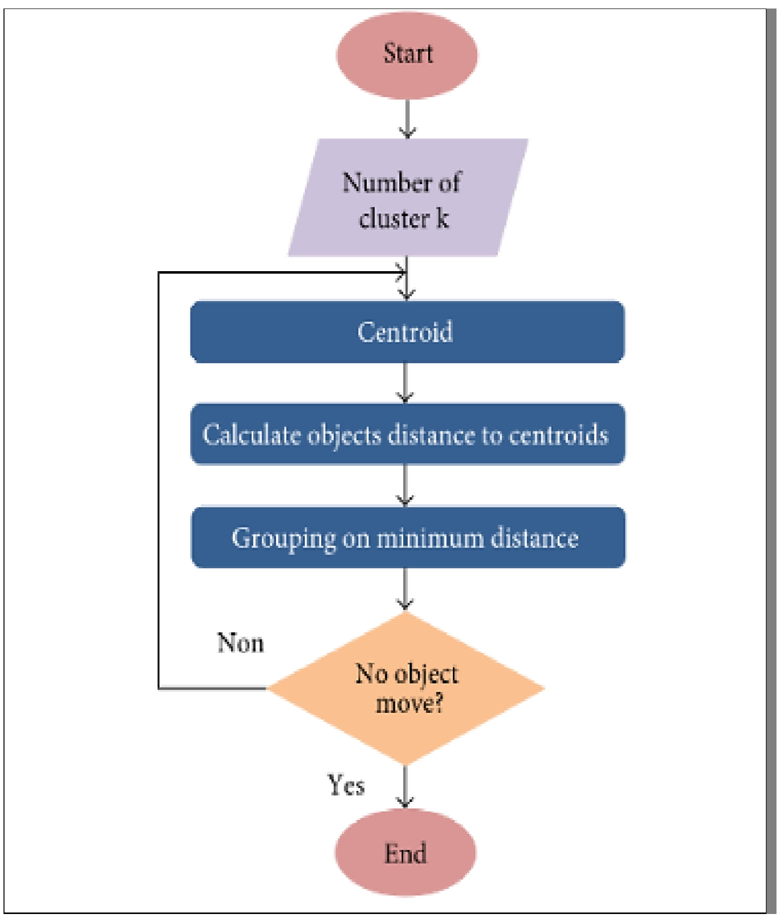

Fig. 3. Steps of the K-mean clustering algorithm [70]

2.6.2 Spherical K-mean

Spherical K-mean (SPKM) is unsupervised learning, and it is a type of clustering algorithms used to group data. SPKM algorithm is found by “Shi Zhong” and is a K-mean algorithm with cosine similarity or measuring the equation between vector data through the inner product, SPKM used the cosine resemblance as a similarity distance measure instead of Euclidean distance [74]. This method aims to improve objective functions [75].

2.6.3 Support vector machine

Support Vector Machine (SVM) is a machine learning algorithm used by Vapnik for classification and regression. SVM offers better performance in applications for orthodox machine learning, pattern recognition to overcome classification glitches [76]. SVM is a valuable tool for an effective nonlinear approximation trick. Structural risk minimization is related to SVM (SRM). For binary classification, SVM was initially used, but it could now be used for multiclass classification. To help nonlinear classification problems where the maximum separation of the hyperplane is built, SVM takes the form of mapping input space into higher dimensional space. The hyperplane is a linear pattern that provides maximum distinction between decision classes with a maximum margin [77].

2.7 Genetic Algorithm (GA)

Genetic Algorithm (GA) is one of the evolutionary algorithms is an algorithm for heuristic stochastic search. It was a strong tool for optimization, searching and machine learning because GA has good worldwide search capacity and can learn the near-optimal solution without the gradient knowledge of error functions. The convergence of evolutionary algorithms and artificial neural network draws much greater interest and forms an area called evolutionary neural networks [78]. Since Genetic Algorithm research has been very active in recent years, a lot of useful conclusions and results have been obtained and some of them have been applied successfully [79].

3. RELATED WORKS

Rostom and Fijri, [80] Proposed a method spherical k-mean algorithm in addressing computational efficiency in terms of quality, resolution, and procedure determination. the authors used modified spherical k-mean cluster to breast cancer Coimbra data collection, from UCI machine learning into several clusters, by using kernel radial base function (RBF) with sigma by internal product measures in spherical k-mean. The highest spherical k-mean kernel accuracy (SPKM) results in a clustering approach with radial base nucleus (RBF) in the Coimbra Dataset (BCC) for breast cancer is equivalent to 72.41%. And when a kernel is applied to the spherical k-means, the precision results are stable. and faster from the use of spherical k-means. However, spherical k-mean achieve greater accuracy in high data training numbers, while kernel spherical k-means in a few data training numbers achieve high accuracy. For the Coimbra dataset classification of breast cancer, the kernel spherical k-means is a good classifier. This technique will allow medical staff to easily identify breast cancer data to predict it faster.

Assir et al., [81] Presented to use of k-means clustering and genetic algorithm for each mammogram by use of Computer-Aided Diagnosis (CAD) systems that enable diagnostic accuracy to be improved by reducing the number of false-positive and false-negative methods. To help radiologists make the right choice by showing them the possibly suspicious area region to early detect cancer detection. Approximately 30% of breast lesions are overlooked screenings during mammogram image screening so, the CAD help to the pre-processing step aims to restrict the ROIs contrast improvement of the ROIs contained in the first phase using the "adaptive local gray level transformation based on variable curve" technique. Finally, The ROIs are segmented using k-mean clustering to optimize the effects of each step, with the genetic algorithm performing both the method of contrast enhancement and the method of segmentation.

Lin and Ji., [82] Proposed to use a k-mean cluster and hybrid algorithm. to show that the hybrid algorithm can correctly cluster the data sets, and the performance of the hybrid algorithm model is better compared to the K means model and SOM neural network model in clustering accuracy and computation speed. The high complexity and low precision of the SOM neural network algorithm and the weakness of the K-means clustering algorithm are due to the need to measure the number of advanced clustering centers and randomly pick initial clustering centers. In this review, the authors proposed a hybrid algorithm combining K-means and the neural network of SOM to provide low complexity and high accuracy. For early breast cancer detection, to save woman lives.

Aswathy and Jagannath, [26] Suggested that a comparison of Three machine learning algorithms, the clustering k-means, the active contour model (ACM), and the clustering algorithm fuzzy c-means. To demonstrate the efficiency of these three segmentation algorithms, the experimental assessment uses different quantitative tests. To classify medical images into benign and malignant, by using a support vector machine classifier. The primary aim of this research is to improve the precision of breast cancer detection using a computer-aided approach to diagnosis. The overall segmentation precision of the k-mean cluster algorithm is 93%. Other efficiency metrics are also superior. The automated classification of breast cancer images is based on GLCM features in the SVM model and achieves 91.1% accuracy. In this study, the SVM classification strategy proposed also enables high sensitivity to be achieved.

Çiklaçandir et al., [83] Proposed a method to help the doctor diagnose a lesion in breast cancer. Ultrasonography is the imaging method used to detect breast cancer. The authors used a k-mean clustering algorithm with three separate filters the Median, Laplace, and Sobel, for early diagnosis. They checked the system for 2×2, 4×4, 8×8, 1×4, 1×8, 4×1, 8×1 partition. The Median filter provided the best result, based on average precision values. Without the filter, near average accuracy rates to the median are also obtained. and the lowest overall accuracy for the Laplace and Sobel filters. Compared to the other six types of partitioning. 8x8 partitioning offers the highest clustering precision. Furthermore, the 4x4 partitioning is presented a second-high accuracy score on average, which is similar to 8x8 partitioning as well. The result suggests using the 8x8 partitioning median filter to get the best result.

ROY1 et al., [84] Worked using three techniques applied on the MRI scanned image of a breast tumor, saliency mapping, and k-means clustering then color space segmentation. the segmentation techniques used, the "salient" or most essential part of the picture is detected in salient mapping was clearly but it also segmentation the noise section, this algorithm specifically detects the tumor part, the clustering of k channels in k-means to segment the image and detect the cancer mass, but also segments the noise part as well. they separated the images based on color spacing segmentation and mask the outer contours. The obtained output image has high clarity in color spacing. Over these segmentation methods, breast cancer is detected, and color spacing segmentation is the best method among the three types as it illustrates images with less error and more consistency.

Velmurugan et al., [85] Presented a comparative study of the two current partitioning clustering algorithms and a hybrid clustering algorithm. Verification of findings using classification algorithms using their accuracy. In this work, they use 310 images, which include three types regular benign and malignant to detect cancer parts from the mammogram. The clustering and classification algorithm performance was carried out based on tumor detection, cluster quality, and other parameters. Some of the well-known classification algorithms, like the clustering algorithm k-Means, FCM, and MCA. Were used to find the accuracy of outcomes. The MCA algorithm was best than the other two algorithms. The accuracy was checked by algorithms of classification J48, JRIP, via its numerous success metrics, SVM, Naive Bayes and CART Among the classification algorithms, the accuracy of CART was found to be higher than the other algorithms. Finally, the result shows that helpful for doctors and radiologists to identify the region of the breast affected by cancer.

Lbachir et al., [86] Presented a method for early detection of breast cancer to save a woman's life. They proposed to use global thresholding and k-mean cluster algorithms in mammogram images to extract suspected lesions, this involves first performing a coarse segmentation (ROI) to give an initial region of interest. Then, a fine segmentation is performed using modified k-means, based on the image histogram. The key inputs are the automatic selection of the number of and the original centroids using a histogram peak analysis algorithm. The algorithm was checked on 170 mammograms in the MIAS dataset. The experimental findings show that the proposed solution methods help to give high accuracy of mammography for lesion detection.

Karthiga and Narasimhan, [87] Presented to overcome the difficulty of a breast cancer diagnosis. and for accurate and early diagnosis, they proposed an automated diagnosis using the k-mean cluster algorithm that is used in this work for cell nuclei segmentation, and Discrete Wavelet Transform (DWT) then a segmented image is added to it. One of the different wavelets used in this paper is Coiflets which has scaling functions with moments of vanishing. They used (BreakHis) The Histopathological Images of Breast Cancer dataset inside it. The outcome of the proposed procedure of SVM gives that result with an accuracy of 93.3% in linear SVM, 92.7% in quadratic SVM, and 91.3% Fine Gaussian SVM. The linear SVM presented the best result and highest accuracy.

Dallali et al., [88] presented the performance of two techniques, K-means and Otsu thresholding, for the image segmentation is a basic task that tests the performance of high-level image treatment criteria for image treatment to in estimating the survival rate and life expectancy of patients with breast cancer, The experimental results, the ROC curve analysis and according to the AUC, calculation show that the Otsu thresholding algorithm is less erroneous and it provides optimum performance with 98.83%, 85.27% and 99.31% the accuracy, sensitivity and specificity, respectively.

Singh and Srivastava, [89] Proposed measures include the cropping of a mammogram to find the region of interest (ROIs), extraction of characteristics using the full local binary pattern based on wavelet (W-CLBP) and K-mean clustering. Strong texture features of the mammogram are captured from two decomposed level ROI Dimensional Discrete Wavelet Transform (2D-DWT) to all detailed coefficients using CLBP (LH, HL, and HH). Accordingly, K-means generates the clusters based on this texture similarity of mammograms, and query mammogram features are matched with all cluster members to find the nearest cluster. Finally, using the Euclidean distance similarity measure, images from this nearest cluster are retrieved. Therefore, the query mammogram is only searched in a small sub-set at the search time based on the cluster size, suggesting a superior response fast time with good accuracy performance. Benchmark (MIAS) database tests confirm that the suggested solution has a greater say in terms of four other variants of texture features. The result showed the structure outperforms suggested for W-CLBP with clustering of K-means.

Samundeeswari et al., [90] Proposed a variant by combining ACO and the regularization parameter of K-mean. proposed to use regularized K-Mean (ReKM) method yields better low-error PRI values for the BUS images dataset to detect the lesion component for further diagnosis of breast cancer than the clustering method. The result showed good segmentation with the use of K-mean clustering show a 96% PRI value and 4% boundary displacement error. The proposed method is intended to be part of the CAD system to help define the lesion portion for further diagnosis of breast cancer.

Prakash et al., [91] Presented three segmentation techniques K-Means, Fuzzy C-Means (FCM), and Gaussian Mixture Model - Expectation-Maximization (GMMEM) to segment and compare the infrared images. These techniques are used to classify cancer tissues by conversion of color space from RGB to L*a*b* to enhance the color analysis for the classification of images into benign and malignant cases. the result shows that the segmentation of the FCM gives a good indication and accuracy of the disease, but the drawback of K-means segmentation is that, in some cases, it results in empty clusters.

Okur et al., [92] Suggested the standard K-Means algorithm with the Ant Colony Optimization. To enable experts or radiologists to find the cancer area with computer assistance, segmentation was attempted on digital mammography images. In addition to conventional methods of segmentation with MIAS dataset, the method proposed is aimed to enhance early detection of the lesion portion for early diagnosis of breast cancer.

Jamal et al., [93] Presented a comprehensive study on WBC datasets for dimensionality reduction. that dimensionality reduction using PCA and K-Means cluster for Breast Cancer Prediction article has shown that the number of breast cancer classification features from the original WBC data set can be reduced by extracting best features, namely transforming original data using decomposition of the main component eigenvector and also using the clustering technique of K-means, based on the Euclidian distance from each cluster centroid, the feature extraction of the function is done by converting data from the original dimension to a new dimension. The metric calculation shows that the reduction in dimensionality using the K-means cluster is almost as good as the PCA with at least two clusters with a reduced number of features. Using only one cluster in k-means clustering yields an incorrect true positive rate classification model sensitivity. The most important indicator for the early detection of breast cancer is sensitivity, as defined the proportion of breast cancer patients correctly.

Table 1. Summary of the reviewed papers

Ref.	Year	Dataset	Techniques	Pros	Cons	Result and Accuracy
[80]	2020	Breast Cancer Coimbra	o Spherical k-means o Kernel-Spherical k-means	This technique will allow medical staff to easily identify breast cancer to predict it faster and give higher accuracy.	Spherical k-means need the high number of data training to get high accuracy.	The result shows that without kernel data Training 80%, time 0,16s, and Accuracy 81,82%. But with kernel data training 20% and 60%, time 0,89s, and Accuracy 72,41%
[81]	2020	Mammogram images	o K-means o Genetic	The development of CAD systems helps to increase breast cancer detection by radiologists and to reduce the number of false-negative medications.	The system needs to be more accurate in automatic classification to detect the type of abnormality.	In this paper the result with specific that CAD improves sensitivity by 21% and reduce the number of false-negative products 5% to 15%. 90% accuracy.
[82]	2020	Wisconsin Breast Cancer UCI	o K-means o SOM hybrid algorithm	K-means are helpful for comprehension and simple calculation. The SOM neural network reflects all points with high precision.	Still need more iterations and reduce the time to get the result faster and more accuracy	Improved overall performance in accuracy and running with time 4,62s, and Accuracy 95%.
[26]	2020	UCSB Images	o K-means o ACM o Fuzzy c-means o SVM	Improve the accuracy of a breast cancer diagnosis with a computer-aided support approach to diagnosis. And It is useful to decrease image noise amplification	Using k-means give higher accuracy but low specificity, and in ACM give low accuracy with high specificity, while Fuzzy c-mean gives low accuracy and specificity from them	K-mean has the maximum segmentation accuracy of 93%. And based on SVM achieved another accuracy of 91%.
[83]	2019	Ultrasonography images	o K-means o filters	Use filters to get a more accurate result was a good advantage	The accuracy increases if just the number of partitioning increased.	The use 8x8 partitioning median filter to get the best result and accuracy of 83.33%.
[84]	2019	MRI images	o K-means o Saliency-mapping	Algorithms were applied to a scanned MRI image of a breast tumor, and the	K-means algorithm detects the tumor part, but often it also segments	Color space segmentation illustrates an image with less error and better accuracy.

Ref.	Year	Dataset	Techniques	Pros	Cons	Result and Accuracy
				findings were clearly shown without noise	the noise part. The precision of this technique is thus reduced.	
[85]	2019	Mammogram images	○ K-mean ○ FCM ○ MCA ○ SVM ○ Naive Bayes CART	The images that have been analyzed by k-Means, FCM .and the MCA helped to identify the affected breast cancer area early on.	The time taken to analyze the images was high and with many images' dataset	CART gives the best accuracy of 92.30% SVM 87.95% NAÏVE BAYES has 84.28% JRIP 82.60% J48 has the least accuracy of 75.58%
[86]	2018	Mammogram from MIAS	○ K-means ○ Otsu thresholding	The method proposed proved effective with pictures Database for MIAS. They give good results.	It needs to reduce the false positive per image FPPI to give good results.	The results show that the proposed methods are easy and achieves 92.93% high sensitivity with a decrease in 1.98 of FPPI with high accuracy.
[87]	2018	BREAKHIS	○ K-means clustering ○ SVM	Use technique SVM because even with large numbers of input data fast training and input images can be labelled as benign or malignant using the extracted features	The higher time takes to analyze the images.	In the case of linear SVM, the best result is obtained compared to the proposed methods, give an accuracy of 93.3%.
[88]	2018	Mammographic Images	○ K-means ○ Otsu thresholding	To segment the mammographic images, they used the Otsu thresholding algorithm and applied a morphological operator to avoid the small noise to reduce false positives.	For further diagnosis, the technique does not identify all the suspect areas.	The ROC curve analysis and the AUC calculation show that with 98.83% 85.27% and 99.31% accuracy, sensitivity, and specificity, the Otsu thresholding algorithm is less erroneous and provides optimum performance.
[91]	2017	infrared images	○ K-means ○ FCM	Improve color analysis for classification of images into benign and malignant cases in early stage	K-Mean's segmentation results in empty clusters in some cases.	The result of FCM segmentation provides a good indication of the disease, with high accuracy.

4. DISCUSSION

Table 1 presents the summary of the reviewed papers. The mammography imaging module datasets have been utilized in the studies that are proposed by [81,85,86,88] for detecting the breast cancer. They utilized clustering algorithms specially K-means with many other algorithms for medical images analysis. The ultimate aim is to improve the diagnosis accuracy that helped the radiologists make the right decision to detect cancer in the early stages. In other studies, like in [80], they explained the kernel spherical k-means is a strong classifier for the Coimbra dataset classification of breast cancer, and when adding a kernel to the spherical k-means improved the training rate of 20%-60% gave an accuracy of 73.41% within 0.89 seconds. This technique enable the doctors to easily identify breast cancer to predict it faster. In [82] the authors proposed a hybrid algorithm combining k-mean clustering and SOM neural network to provide low complexity and high precision. Performance for the early breast cancer detection that depended on Wisconsin Breast as a data set. In [86,88] To extract suspected lesions, the authors suggested applying global thresholding and K-means algorithms to mammogram images. It used segmentation (ROI) to offer an initial region of interest, using modified k-means to perform a fine segmentation based on the images. These experimental results show that the proposed method outperforms other methods of mammography. For this study, they achieved the best result with the highest accuracy in real-time. The study of [26] Three machine learning algorithms, such as the k-means clustering, active contour model (ACM), and fuzzy c-means clustering algorithm, should be used for comparison. To demonstrate the efficiency of these three segmentation algorithms, the experimental assessment uses different quantitative tests to classify images into benign and malignant classifier with the highest accuracy. Additionally, [87] proposed an automated diagnosis also using the k-means clustering that is used in this work for cell nuclei segmentation, and DWT (Discrete Wavelet Transform) is applied to the segmented images that can classify breast cancer into benign or malignant. In [83] presented to use three different types of filters Median, Laplace, and Sobel were performed, for early diagnosis, The Median filter gave the best results and the 8x8 partitioning offers the highest clustering precision. The results suggest that these techniques are effective in the early prognosis especially in [82] has got higher results compared to other techniques, and can help to increase the early detection of breast cancer in the next years.

5. CONCLUSION

Breast cancer is one of the main causes of death among women worldwide, this review paper presents the studies that reveal the effectiveness of medical images that are used for the detection of breast cancer. Also, it demonstrates the benefits and drawbacks of conventional medical imaging techniques such as mammography and ultrasound, as well as some new imaging techniques, in order to increase diagnosis accuracy and assist radiologists in making the right decision. This study relied on the use of k-mean clustering and classification techniques, which provide the most accurate results and reduce the number of false positives per picture (FPPI). The findings show the clustering techniques with classifier techniques that are used in medical images will make it easier for a doctor to diagnose breast cancer at an early stage to identify the total area affected by cancer, and these techniques are more effective in early prognosis than other techniques, and that they may help to improve early detection of breast cancer in the coming years.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Al Nahid, Y Kong. Involvement of machine learning for breast cancer image classification: A Survey, Comput. Math. Methods Med. 2017;2017(i).
2. Abdulazeez AM, Zeebaree DQ, Abdulqader DM. Wavelet Applications in Medical Images: A Review. Transformation (DWT). 2020;21,22.
3. Zeebaree DQ, Haron H, Abdulazeez AM, Zebari DA. Trainable model based on new uniform LBP feature to identify the risk of the breast cancer, in 2019 International Conference on Advanced Science and Engineering (ICOASE). 2019; 106–111.
4. Zeebaree DQ, Haron H, Abdulazeez AM, Zeebaree SRM. Combination of K-means clustering with Genetic Algorithm: A review, Int. J. Appl. Eng. Res. 2017; 12(24):14238–14245.

5. Saeed JN, Abdulazeez AM. Facial Beauty Prediction and Analysis Based on Deep Convolutional Neural Network: A Review.
6. Rabidas R, Midya A, Chakraborty J, Arif W. A study of different texture features based on local Operator for Benign Malignant Mass Classification. 2016;93:389-395.
7. Zeebaree DQ, Haron H, Abdulazeez AM, Zebari DA. Machine learning and Region Growing for Breast Cancer Segmentation. In 2019 International Conference on Advanced Science and Engineering (ICOASE). IEEE. 2019;88-93.
8. Arunkumar N, Mohammed MA, Mostafa SA, Ibrahim DA, Rodrigues JJPC, de Albuquerque VHC. Fully automatic model-based segmentation and classification approach for MRI brain tumor using artificial neural networks. Concurrency Computat Pract Exper. 2018;e4962. Available:https://doi.org/10.1002/cpe.4962
9. Jahwar AF, Abdulazeez AM. Meta-heuristic algorithms for K-means clustering: A Review. PalArch's Journal of Archaeology of Egypt/Egyptology. 2020;17(7):12002-12020.
10. Adeen IMN, Abdulazeez AM, Zeebaree DQ. Systematic Review of Unsupervised Genomic Clustering Algorithms Techniques for High Dimensional Datasets.
11. Susheelamma KH, Brahmananda SH. A survey on clustering and feature selection algorithm for quickly predicting engineering students' academic performance. International Journal of Research in Computer Science Engineering and Technology. 2018;4(1):2394-409.
12. Salih AA, Abdulazeez AM. Evaluation of Classification Algorithms for Intrusion Detection System: A Review. Optimization (PSO); 42,43.
13. Louise CMT, Jodrell N. Screening and breast cancer: the role of breast awareness. Journal of Cancer Nursing. 1997;1(2):76-80.
14. Gayathri R, Cauveri A, Kanagapriya R, Nivetha V, Tamizhselvi P, Kumar KP. A novel approach for clustering based on bayesian network. In Proceedings of the 2015 International Conference on Advanced Research in Computer Science Engineering & Technology (ICARCSET 2015). ACM. 2015;60.
15. Othman G, Zeebaree DQ. The applications of discrete wavelet transform in image processing: A Review. Journal of Soft Computing and Data Mining. 2020;1(2):31-43.
16. Omar N, Abdulazeez AM, Sengur A, Al-Ali SGS. Fused faster RCNNs for efficient detection of the license plates. Indonesian Journal of Electrical Engineering and Computer Science. 2020;19(2):974-982.
17. Hasan DA, Abdulazeez AM. A Modified Convolutional Neural Networks Model for Medical Image Segmentation. Learning. 2020;20,22.
18. Saleem SI, Abdulazeez AM, Orman Z. A new segmentation framework for arabic handwritten text using machine learning techniques, Computers, Materials & Continua. 2021;68(2):2727–2754.
19. Goel N, Yadav A, Singh BM. Medical image processing: A review, 2016 Second International Innovative Applications of Computational Intelligence on Power, Energy and Controls with their Impact on Humanity (CIPECH), Ghaziabad. 2016;57-62. DOI: 10.1109/CIPECH.2016.7918737
20. Laffont V, Durupt F, Birgen MA, Bauduin S, Laine AF. Detection of masses in mammography through redundant expansions of scale. In 2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE. 2001;3:2797-2800.
21. Akila K, Sumathy P. Early breast cancer tumor detection on mammogram images. International Journal Computer Science, Engineering and Technology. 2015;5(9):334-336.
22. Joshi S, Priyanka Shetty SR. Performance analysis of different classification methods in data mining for diabetes dataset using WEKA tool. International Journal on Recent and Innovation Trends in Computing and Communication. 2015;3(3):1168-1173.
23. Young CM. High-speed volumetric in vivo medical imaging for morphometric analysis of the human optic nerve head (Doctoral dissertation, Applied Science: School of Engineering; 2011.
24. Senthilkumar B, Umamaheswari G. Combination of Novel Enhancement Technique and Fuzzy C Means Clustering Technique in Breast Cancer Detection; 2013.

25. Ramani R, Valarmathy S, Vanitha NS. Breast cancer detection in mammograms based on clustering techniques-a survey. International Journal of Computer Applications. 2013;62(11).
26. Aswathy MA, Jagannath M. Performance Analysis of segmentation algorithms for the detection of breast cancer. Procedia Computer Science. 2020;167:666-676.
27. Adam A, Omar K. Computerized breast cancer diagnosis with Genetic Algorithm and Neural Network, in Proc. of the 3rd International Conference on Artificial Intelligence and Engineering Technology (ICAIET). Malaysia. 2006;22–24:533–538.
28. Denny J, Shamsida A, Sobha T. Efficient Segmentation Method for ROI Detection in Mammography Images Using Morphological Operations.
29. Zeebaree DQ, Haron H, Abdulazeez AM. Gene selection and classification of microarray data using convolutional neural network. In 2018 International Conference on Advanced Science and Engineering (ICOASE). IEEE. 2018;145-150.
30. Vaishali Kumbhakarna, Vijaya R.Patil, Dr. Seema Kawathekar. Review on Speckle Noise Reduction Techniques for Medical Ultrasound Image Processing. I. J. of Computer Techniques. 2015;2(1).
31. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015, CA: A Cancer Journal for Clinicians. 2015;65(1):5–29.
32. Pal NR, Bhowmick B, Patel SK, Pal S, Das J. A multi-stage neural network aided system for detection of microcalcifications in Digitized Mammograms Neuro Computing. 2008;2625–2634.
33. Duffy SW, Tabar L, Chen HH, Holmqvist M, Yen MF, Abdsalah S, et al. The impact of organized mammography service screening on breast carcinoma mortality in seven swedish counties, Cancer: 95(3).
34. Naveen A, Velmurugan T. A Novel Layer Based Logical Approach (LLA) Clustering Method for Performance Analysis in Medical Images. International Journal of Control Theory and Applications. 2016;9:4647-4660.
35. Kaur G, Chhabra A. Improved J48 Classification Algorithm for the Prediction of Diabetes. International Journal of Computer Applications. 2014;98:13-17.
Available:https://doi.org/10.5120/17314-7433
36. Velmurugan T, Santhanam T. Computational Complexity between K-Means and K-Medoids Clustering Algorithms for Normal and Uniform Distributions of Data Points. Journal of Computer Science. 2010;6:363-368.
Available:https://doi.org/10.3844/jcssp.2010.363.368
37. Abdulazeez AM, Hajy DM, Zeebaree DQ, Zebari DA. Robust watermarking scheme based LWT and SVD using artificial bee colony optimization. Indonesian Journal of Electrical Engineering and Computer Science. 2021;21(2):1218-1229.
38. Kalaiselvi T, Sriramakrishnan P, Somasundaram K. Survey of using GPU CUDA programming model in medical image analysis. Informatics in Medicine Unlocked. 2017;9:133-144.
39. Dhanachandra N, Manglem K, Y Jina Chanu. Image segmentation using K-means Clustering Algorithm and Subtractive Clustering Algorithm, in Procedia Computer Science. 2015;54:764 – 771.
40. Prabha DS, Kumar JS. Assessment of banana fruit maturity by image processing technique. Journal of Food Science and Technology. 2015;52(3):1316–27.
41. Dinesh D. Patil, Ms. Sonal G. Deore, Medical Image Segmentation, inI JCSMC. 2013;2(1):22–27.
42. Jalalian A, et al. Foundation and methodologies in computer-aided diagnosis systems for breast cancer detection, EXCLI J. 2017;16:113-137.
43. Sadat Fasihi M, Mikhael WB. Overview of current biomedical image segmentation methods," 2016 Int. Conf. Comput. Sci. Comput. Intell. 2016;803-808.
44. Norouzi A, et al. Medical image segmentation methods, Algorithms and Applications, IETE Tech. Rev. 2014;31(3):199-213.
45. Kashyap KL, Bajpai MK, Khanna P. Breast cancer detection in digital mammograms. IEEE International Conference on Imaging Systems and Techniques. 2015;1-6.
46. Jiang D, Tang C, Zhang A. Cluster analysis for gene expression data: A survey, IEEE Transactions on Knowledge and Data Engineering. 2004;16:1370-1386.
47. Zeebaree DQ, Abdulazeez AM, Zebari DA, Haron H, Nuzly H. Multi-level fusion in ultrasound for cancer detection based on uniform lbp features, Computers, Materials & Continua. 2021;66(3):3363–3382.

48. Zebari Dilovan Asaad, Diyar Qader Zeebaree, Adnan Mohsin Abdulazeez, Habibollah Haron, and Haza Nuzly Abdull Hamed. Improved Threshold Based and Trainable Fully Automated Segmentation for Breast Cancer Boundary and Pectoral Muscle in Mammogram Images. IEEE. 2020;8:203097-203116.
49. Sulaiman DM, Abdulazeez AM, Haron H, Sadiq SS. Unsupervised learning approach-based new optimization K-means clustering for finger vein image localization, 2019 International Conference on Advanced Science and Engineering (ICOASE), Zakho - Duhok, Iraq. 2019;82-87.
DOI: 10.1109/ICOASE.2019.8723749
50. Velmurugan T. Performance based analysis between K-means and Fuzzy C-means Clustering Algorithms for Connection Oriented Telecommunication Data. Applied Soft Computing. 2014;19:134-146.
51. Amygdalos I. Detection and classification of gastrointestinal cancer and other pathologies through quantitative analysis of optical coherence tomography data and goniophotometry. PhD Dissertation, Department of Medicine, Imperial College, London; 2014.
52. Velmurugan T, Venkatesan E. Effective Fuzzy C Means Algorithm for the Segmentation of Mammogram Images of Identify Breast Cancer. International Journal of Control Theory and Applications. 2016;9:4647-4660.
53. Bataineh KM, Naji M, Saqer M. A comparison study between Various Fuzzy Clustering Algorithms. Jordan Journal of Mechanical and Industrial Engineering (JJMIE). 2011;5:335.
54. Dhanachandra N, Manglem K, Chanu YJ. Image segmentation using K-means clustering algorithm and subtractive clustering algorithm. Procedia Computer Science. 2015;54:764-771.
55. Aimi Salihai Abdul, Mohd Yusuff Masor, Zeehaida Mohamed. Colour image segmentation approach for detection of malaria parasiter using various colour models and k-Means Clustering, In WSE ASTransaction on Biology and Biomedecine. 2013;10.
56. Bain KK. Customer segmentation of SMEs using K-Means clustering method and modeling LRFM, International Conference on Vocational Education and Electrical Engineering, Universitas Negeri Surabaya; 2015.
57. Haralick RM, Shapiro LG. Image segmentation techniques, Computer Vision, Graphics, and Image Processing. 1985;29(1):100–132.
58. Celebi ME, Kingravi HA, Vela PA. A comparative study of efficient initialization methods for the k-means clustering algorithm, Expert Systems with Applications. 20140:200.
59. Garcia AJ, Flores WG. Automatic clustering using nature-Inspired Metaheuristics: A Survey. Appl. Soft Comput; 2016.
Available:https://doi.org/10.1016/j.asoc.2015.12.001
60. Das S, Abraham A, Konar A. Automatic clustering using an improved differential evolution Algorithm. IEEE Trans. Syst. Man, Cybern. –Part A Syst. Humans. 2008;38:218–237.
Available:https://doi.org/10.1109/TSMCA.2007.909595
61. Neda S, Farzad MS. Customer Segmentation of Bank Based on Discovering of Their Transactional Relation by Using Data Mining Algorithms, Modern Applied Science. 2016;10:283.
62. Shehroz S Khan, Amir Ahmad. Cluster centre initialization algorithm for k-means cluster, In P Attern Recognition Letters. 2004;1293–1302.
63. Kaur A. Comparative Analysis of Segmentation Algorithms for Brain Tumor Detection in MR Images.
64. Yang MS, Sinaga KP. A feature-reduction multi-view K-means clustering algorithm. IEEE. 2019;7:114472-114486.
65. Lin H, Ji Z. Breast cancer prediction based on K-Means and SOM Hybrid Algorithm. In Journal of Physics: Conference Series. IOP Publishing. 2020;1624(4):042012.
66. Aswathy MA, Jagannath M. Performance Analysis of Segmentation Algorithms for the Detection of Breast Cancer. Procedia Computer Science. 2020;167:666-676.
67. Çiklaçandir FGY, Ertaylan A, Bınzat U, Kut A. Lesion Detection from the Ultrasound Images Using K-Means Algorithm. In 2019 Medical Technologies Congress (TIPTEKNO). IEEE. 2019;1-4.
68. Bottou L, Lin C-J. Support vector machine solvers. Large Scale Kernel Mach. 2007;3(1):301–320.
69. Tang T, Chen S, Zhao M, Huang W, Luo J. Very large-scale data classification based

on K-means clustering and multi-kernel SVM. Soft Computing. 2019;23(11):3793-3801.
70. Dhanachandra N, Manglem K, Chanu YJ. Image segmentation using K-means clustering algorithm and subtractive clustering algorithm, Procedia Computer Science. 2015;54:764–771. View at: Publisher Site | Google Scholar.
71. Et-taleby A, Boussetta M, Benslimane M. Faults detection for photovoltaic field based on K-Means, Elbow, and Average Silhouette Techniques through the Segmentation of a Thermal Image. International Journal of Photoenergy; 2020.
72. Yong Y, Chongxun Z, Pan L. A novel fuzzy C-Means clustering algorithm for image thresholding, Measurement Science Review. 2004;4(1).
73. Chen S, Zhang D. Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure, IEEE Transactions on Systems, Man and Cybernetics. 1998; 34:1907-1916.
74. Duwairi R, Abu-Rahmeh M. A novel approach for initializing the spherical k-means clustering algorithm, Simulation Modelling Practice and Theory. 2015; 54:49-63.
75. Zhong S. Efficient online spherical k-means clustering, Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN). 2005;18:790-798.
76. Mittal M, Kumar K. Network lifetime enhancement of homogeneous sensor network using ART1 neural network. In Proceedings of the Sixth International Conference on Computational Intelligence and Communication Networks (CICN), Bhopal, India. 2014;14–16.
77. Cristianini N, Taylor JS. An introduction to support vector machines and other kernelbased learning methods, Cambridge University Press; 2000.
78. Yao X, Xu Y. Recent advances in evolutionary computation [J]. J Comput Sci Technol 2006;21(1):1–18.
79. Ding N, Su SC, Yu J. An optimizing BP neural network algorithm based on genetic algorithm. Artificial Intelligence Review. 2011;36(2):153-162.
80. Rustam Z, Fijri AL. Breast cancer clustering using modified spherical K-Means. In Journal of Physics: Conference Series, IOP Publishing. 2020;1490(1): 012028.
81. Assir A, Harmouchi M, Lyazidi A, Rattal M, Mouhsen A. Fully automatic omputer-aided detection of breast cancer based on genetic algorithm optimization. In 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), IEEE. 2020;1-6.
82. Lin H, Ji Z. Breast cancer prediction based on K-Means and SOM Hybrid Algorithm. In Journal of Physics: Conference Series. IOP Publishing. 2020;1624(4): 042012.
83. Çiklaçandir FGY, Ertaylan A, Bınzat U, Kut A. October lesion detection from the ltrasound Images Using K-Means Algorithm. In 2019 Medical Technologies Congress (TIPTEKNO). IEEE. 2019;1-4)
84. Roy K, Ghosh S, Mukherjee A, Sain S, Pathak S, Chaudhuri SS, Chakraborty M. Breast tumor segmentation using image segmentation algorithms. In 2019 International Conference on Opto-Electronics and Applied Optics (Optronix). IEEE. 2019;1-5.
85. Velmurugan T, Venkatesan E. A hybrid multifarious clustering algorithm for the analysis of memmogram images. Journal of Computer and Communications. 2019;7(12):136-151.
86. Lbachir IA, Daoudi I, Tallal S. Automatic detection of suspicious lesions in mammograms by histogram-peak-analysis based K-means. In 2018 9th International Symposium on Signal, Image, Video and Communications (ISIVC). IEEE. 2018;16-21.
87. Karthiga R, Narasimhan K. Automated diagnosis of breast cancer using wavelet based entropy features. In 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA). IEEE. 2018;274-279.
88. Dallali A, El Khediri S, Slimen A, Kachouri A. Breast tumors segmentation using Otsu method and K-means. In 2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). IEEE. 2018;1-6.
89. Singh VP, Srivastava R. Content-based mammogram retrieval using wavelet based complete-LBP and K-means clustering for the diagnosis of breast cancer. International Journal of Hybrid Intelligent Systems. 2017;14(1-2):31-39.

90. Samundeeswari ES, Saranya PK, Manavalan, R., Mar Segmentation of breast ultrasound image using regularized K-means (ReKM) clustering. In 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). IEEE. 2016;1379-1383.
91. Prakash RM, Bhuvaneshwari KM, Divya KJ Sri, Begum AS. Segmentation of thermal infrared breast images using K-means, FCM and EM algorithms for breast cancer detection, 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore. 2017;1-4.
DOI: 10.1109/ICIIECS.2017.8276142
92. AG Hİ Okur, ÖÜP Görgel, Sertbas A. Segmentation on digital mammogram images using hibrid K-means clustering and ant colony algorithms, 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), Istanbul. 2018;1-4.
DOI: 10.1109/EBBT.2018.8391422
93. Jamal A, Handayani A, Septiandri AA, Ripmiatin E, Effendi Y. Dimensionality reduction using pca and k-means clustering for breast cancer prediction. Lontar Komputer: Jurnal Ilmiah Teknologi Informasi. 2018; 192-201.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/68284