Automatika

Journal for Control, Measurement, Electronics, Computing and Communications

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/taut20

Computer-aided diagnostic system for breast cancer detection based on optimized segmentation scheme and supervised algorithm

S. Balaji, T. Arunprasath, M. Pallikonda Rajasekaran, G. Vishnuvarthanan & K. Sindhuja

To cite this article: S. Balaji, T. Arunprasath, M. Pallikonda Rajasekaran, G. Vishnuvarthanan & K. Sindhuja (2023) Computer-aided diagnostic system for breast cancer detection based on optimized segmentation scheme and supervised algorithm, Automatika, 64:4, 1244-1254, DOI: 10.1080/00051144.2023.2244307

To link to this article: https://doi.org/10.1080/00051144.2023.2244307

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 19 Sep 2023.

Submit your article to this journal

Article views: 426

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at https://www.tandfonline.com/action/journalInformation?journalCode=taut20

AUTOMATIKA
2023, VOL. 64, NO. 4, 1244–1254
https://doi.org/10.1080/00051144.2023.2244307

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Computer-aided diagnostic system for breast cancer detection based on optimized segmentation scheme and supervised algorithm

S. Balaji[a], T. Arunprasath[b], M. Pallikonda Rajasekaran[a], G. Vishnuvarthanan[b] and K. Sindhuja[c]

[a]Department of Electronics and Communication Engineering, Kalasalingam Academy of Research and Education, Krishnankoil, India; [b]School of Computing Science and Engineering, VIT Bhopal University, Madhya Pradesh, India; [c]Department of Electronics and Communication Engineering, Theni Kammavar Sangam College of Technology, Theni, India

ABSTRACT
Breast cancer is a serious threat to the womankind and it leads the susceptible kinds of cancer for women. The mortality rates due to breast cancer increases every single year and the World Health Organization (WHO) aims to reduce the occurrence of breast cancer by at least 2.5% per year. The occurrence of breast cancer can be minimized only when periodical screening is carried out. Mammography is one of the effective screening procedure, which can effectively locate earlier signs of breast cancer. As an aid, this work aims to present a system for the breast cancer detection and classification. This work is segregated into four phases and all these phases aim to enhance the classification performance. The efficiency of the proposed work is evaluated against the state-of-the-art approaches and the proposed contribution to the medical science. The computer-aided diagnostic system (CADS) proves 98.2% accuracy, with minimal false positive and false negative rates in a reasonable period of time.

ARTICLE HISTORY
Received 29 April 2023
Accepted 25 June 2023

KEYWORDS
Breast cancer detection; classification; CADS; optimization; supervised algorithm

1. Introduction

Cancer is one of the dangerous diseases being observed all through the world. The cancerous growth is commonly detected in different organs such as oral cavity, lung, oesophagus, cervix and so on. Among all of these types of cancers, breast cancer is most perceptible among women. Recent statistical reports claim that the breast cancer was detected in about 2.3 million women and the global mortal rates due to this cancer have reached around 6,85,000 in the year 2020 [1]. The report states that about 14% of the cancer in women is observed to be breast cancer. The entire women population suffers from this incessant disease.

The initial point of this disease does not show up any noticeable symptom and the symptoms are observed only when the severity of the cancer progresses. Hence, periodical screening is quite necessary for womanhood to escape from this crucial disease. The early detection of breast cancer can enhance the quality of life and the lifespan of the patient can also be increased significantly. One of the best imaging techniques for breast cancer screening is mammography, which is highly capable of identifying pre-cancerous features such as malignant lumps, microcalcifications and thickened skin.

The group of calcifications is the first sign of breast cancer. Microcalcification is a microscopic calcium deposit seen in breast tissue. This calcium deposit might be spherical, lobular, specular, or uneven in shape. These microcalcifications appear as tiny granules on a mammogram and are extremely difficult for the physician to correctly detect. According to a study, between 10% and 40% of microcalcifications are overlooked by physicians [2,3]. However, computers can quickly identify anomalies, if the system is properly trained.

Recognizing the seriousness of this dreadful disease, the computer technology has rendered remarkable contribution to the medical science. The computer-aided diagnostic system (CADS) helps in assisting the healthcare professional for diagnosing and treating various diseases [4]. Though there are numerous related works concerning the CADS for breast cancer detection, the medical field always welcome computerized system with minimal false positive and false negative rates.

This article intends to present a "CADS" for breast cancer detection and classification by employing advanced image processing techniques. The suggested CADS is organized around four critical phases including pre-processing of mammography images, segmentation, feature extraction and classification. Preprocessing prepares the mammography image for following processes, whereas segmentation identifies the region of interest. The potential features are extracted from the segmented region and are utilized to train the

CONTACT T. Arunprasath arunprasatht.be.kare@gmail.com Department of Biomedical Engineering, Kalasalingam Academy of Research and Education, Krishnankoil 626126, India

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

CADS. Finally, the classifier differentiates between the normal and abnormalities in the mammogram images.

This work employs Lifting Wavelet Transform (LWT) for pre-processing the mammogram images, as LWT requires minimal time and memory. The Lion Optimization Algorithm (LOA) with Fuzzy C Means (FCM) algorithm is utilized to extract region of interest and the Gabor along with Local Ternary Pattern (LTP) features are extracted for training the Support Vector Machine (SVM) classifier. The contributions of this work are as follows.

- The LWT is employed to pre-process the mammograms, owing to its minimal time and memory requirements.
- The segmentation accuracy is improved, as it is performed by the combination of LOA and FCM algorithm. The optimal initial seed point is chosen by the LOA algorithm, which enhances the segmentation efficiency.
- The combination of Gabor and LTP features is effective in determining the significant features.
- SVM classifier requires minimal time for classification, as the training process is effectively done.

The remaining parts of the paper are organized as follows. Section 2 reviews the related works with respect to breast cancer detection and classification. The proposed CADS is explained in Section 3 and Section 4 analyses the performance of the work. Finally, the conclusions are drawn in Section 5.

2. Review of literature

This section reviews the related literature concerning the breast cancer detection and classification.

The authors of ref. [5] reduced the noise coefficient in the input image by utilizing the method of the median filter. The authors segment images using Gaussian Mixture Models (GMMs), a well-known clustering technique, and then using a Probabilistic Neural Network (PNN). A classifier based on the features of the coincident matrix algorithm. Gray area (GLCM) might be benign, cancerous, or normal in nature.

The authors of ref. [6] detected breast cancer using Intelligent Artificial Bee Colonization and the Improved Monarch Butterfly Optimization Technique (IABC-EMBOT). The approach is quick and precise. The accuracy of classification is 97.53%, the sensitivity varies from 96.75 to 97.04%, and the average time required for processing is 113.42 s. In ref. [7], a technique is suggested for predicting breast cancer recurrence using feature selection. Particle Swarm Optimization (PSO-RM) is the selection approach, which utilizes three separate classifiers: k-Nearest Neighbour (k-NN), Naïve Bayes (NB) and a rapid decision tree. The suggested technique selects the best quadratic method from the 34 features and increases the accuracy of each of the three classifiers. k-NN accuracy increased from 70% to 80%, NB accuracy increased from 76% to 80%, and the rapid decision tree accuracy increased from 66% to 75%.

The work presented in ref. [8] employs Deep Neural Network (DNNs). They classified breast cancer data using deep neural networks with several layers. The accuracy of this system is 97.66%, and sensitivity is little under 0.98, according to experimental data. This study's deep network is only relevant to breast cancer datasets. The authors of ref. [9] estimated the basal chest muscle boundary using a generic thresholding approach and then morphological techniques were used to modify the obtained area borders, and a mean filter was used to eliminate noise. The property is extracted using the Grey Level Co-occurrence Matrix (GLCM) technique, and a genetic algorithm is used to choose a subset of these characteristics that performs the best in terms of classification rate. Finally, SVM is utilized to distinguish benign and malignant tumours.

The work proposed in ref. [10] offers a generic thresholding technique for obtaining chest limits from pictures that converts images to binary using a constant threshold value of 18. The chest region is defined as each component with a substantial amount of pixels linked. The region border is then smoothed using morphologically based filtering procedures on a disc with a radius of five pixels.

The authors of ref. [11] have presented a method for determining the breast borders using a generic thresholding technique. The mean filter reduces the 8-bit picture noise, while the Contrast Limited Adaptive Histogram Equalization (CLAHE) method improves the image contrast. Following that, the images are converted to binary pictures using a preset threshold and then morphologically filtered to remove minor background items. The study's results indicated that the highest accuracy was 97.1%, the sensitivity was 98.8% and the specificity was 95.4%.

An efficient classification system is presented in ref. [12] for predicting breast cancer risk using mammographic picture features. 500 data points were examined, 50% of which were classified as high danger, and 50% as low risk. They presented a Linear Programming Problem (LPP) model based on a number of factors in order to make the feature space smaller in order to forecast the risk of cancer detection. Unlike conventional feature selection approaches, which choose a selection of optimum features from the principal feature, LPP generates a new optimum feature array that incorporates non-primary characteristics in the feature pool, improving hazard prediction accuracy by 9.7%.

The work presented in ref. [13] performed research to categorize breast cancer mammographic pictures. To extract texture-type characteristics from pictures, the GLCM method was employed. Then, in addition to the entire collection of features, the individual created a smaller set. 60% of data is used for training, 20% for validation and 20% for testing. The findings of this research demonstrated 99% accuracy in the picture identification method while using a neural network as a classifier.

The work proposed in ref. [14] extracted the chest region, mammography pictures, dark masking and mild filtering to map worrisome abnormalities using a method for adjusting partial differential equations. The algorithm for FCM clustering is employed. This work has rotated the local binary pattern and calculated local binary patterns to identify the textural characteristics of probable fragmented masses. Finally, regions of suspicion are classified using SVM, a polynomial kernel and radial basis function, as well as a multilayer and linear perceptron. This work rotates the Local Binary Pattern (LBP) and calculated LBP to identify the textural characteristics of probable fragmented masses.

In ref. [15], a fuzzy histogram method is employed to pre-process mammographic images and used the fuzzy c-mean bounded probability (IPFCM) technique to address the limitations of earlier techniques, such as noise and random grouping. Ultimately, to assist the radiologist in diagnosing tumours, extraction, categorization and validation are performed. In ref. [16], a modified version of AlexNet is utilized that had been pre-trained to a database of CBISDDSM mammography pictures, making extensive modifications. Instead of Rectified Linear Unit (ReLU), the AlexNet network design employed parameters with more complex functionality, such as PReLu. The classification accuracy (ACC) and Area Under the Curve (AUC) of this study are 80.4% and 0.84, respectively.

Inspired by these existing techniques, the suggested "CADS" methodology is an excellent method for detecting and classifying breast cancer with a low percentage of false positive and negative results. The next section elaborates on the recommended strategy.

3. Proposed CADS for breast cancer detection and classification

Mammograms can identify breast cancers at an early stage and assist physicians in disease diagnosis and treatment. Certain grayscale-based segmentation techniques have been shown to be successful at extracting the precise boundaries of homogeneous grayscale areas. Breast cancers do not have a consistent look in their early stages and as a result, clinicians may be unable to detect problems. In these instances, the automated method enables clinicians to readily spot anomalies. A tumour identification algorithm must be capable of identifying the lesion and must be accurate with a low probability of false negatives. The overall flow of the proposed work is depicted in Figure 1.

With this idea, the proposed work intends to differentiate between the malignant and benign types of cancer. Following the rule "The better the pre-processing, the better are the results", this work pre-processes the mammogram images with the help of LWT and the segmentation is carried out with the help of LOA in association with FCM. The Gabor and LTP features are then extracted from the region of interest and SVM classifier is employed for differentiation.

3.1. Mammogram pre-processing by LWT

Wavelet transform preserves spatial information, however, suffers from computational and memory complexity. The LWT requires minimal memory with minimal time consumption. The lifting technique enables the generation of an unlimited number of discrete biorthogonal wavelets from a single entity. By starting with the wavelet transform's nature, this lifting technique creates excellent reconstruction filter banks. Split, predict and update are the three fundamental steps that define a single lifting step.

The split phase is divides the signal into many discontinuous components, as shown in Equation (1). Additionally, the predict operation might be referred to as a dual lifting step. This replaces the samples of the odd polyphase component with the difference between the odd polyphase component and its anticipated value, as shown in Equation (2). The update phase is sometimes referred to as the primal lifting phase. The update component is the even polyphase component, as defined in Equation (3). It is based on the linear combination of the sample difference acquired in the predict step.

$$Sp_E(i,j) = SP(i,2j); SP_O(i,j) = SP(i,2j+1) \quad (1)$$

$$HI(i,j) = SP_O(i,j) - \Pr[SP_E(i,j)] \quad (2)$$

$$LW(i,j) = S_E(i,j) + UP(H(i,j)) \quad (3)$$

In the above equations, $Sp_E(i,j)$ and $SP_O(i,j)$ denote even and odd samples respectively. The high pass and low pass coefficients are represented by $HI(i,j)$ and $LW(i,j)$ respectively. Pr[.] and UP[.] are the predict and update operators. Hence, the LWT is applied and the approximation information of the image is taken into account for further processing, which results in reduced time consumption with better efficiency. The pre-processed image is then treated with image segmentation phase.

3.2. Mammogram segmentation by LOA optimized FCM

This section presents the background information of FCM and LO algorithm followed by the description of the proposed segmentation algorithm.

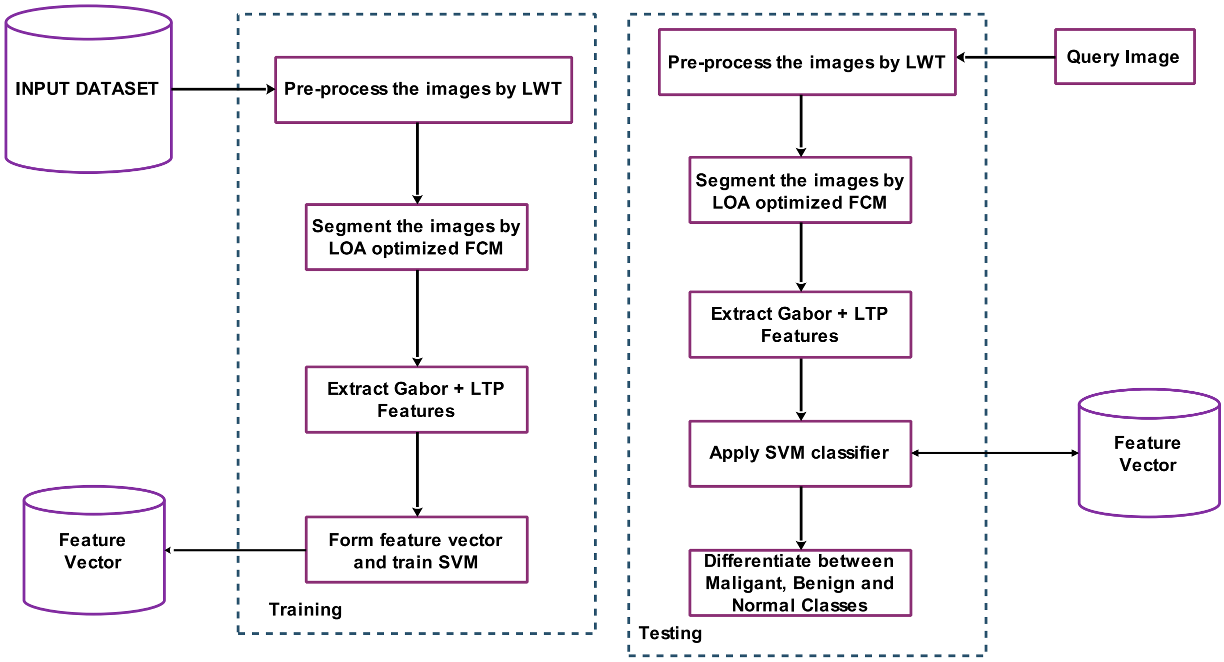

Figure 1. Overall flow of the proposed CADS.

3.2.1. FCM algorithm

The FCM algorithm allots each and every entity with a representation degree of 0 to 1. Hence, the summation of representation of each entity in a cluster is 1. This problem can be seen as a minimization problem and the values are determined by means of the degree of representation. The functional value of FCM is computed by the variance between the entities and the centroid of the cluster. As this function needs to be minimized, non-linear optimization problem has to be solved and it can be attained by any metaheuristic algorithm. The clustering operation of FCM is performed by the following equation [17,18].

$$CL_{FCM} = \sum_{i=1}^{K} \sum_{j=1}^{C} \mu_{ij}^{f} \|a_i - c_j\|^2 \tag{4}$$

In the above equation, μ_{ij} is the fuzzy partition matrix and it ranges from 0 to 1. i is the entity where, j is the cluster's centroid and f is the cluster's fuzziness. The squared Euclidean distance between the entity and the centroid is computed by $\|a_i - c_j\|^2$.

3.2.2. Lion optimization algorithm

LO algorithm is a bio-inspired algorithm that duplicates the behaviour of original lions. On studying the life policy of lions, it can be noticed that the lions organize themselves as resident and nomadic lions [19,20]. The resident lions live in a group named as pride and each pride consists of about four to five female lions, their cubs and one or two male lions. Eventually, the cubs grow and they reach maturity. At this juncture, the young male lions are expelled from the pride. As these young lions do not have any company, they wander without any specific principle. However, a nomadic lion can become a part of a pride, when it can defeat the matured lion inside a pride. At this point, swap happens inside a pride and this may happen at any point of time. Hence, the stay at a pride is impermanent for any lion. The overall process of the LOA is presented as follows.

The activities of lionesses are quite different from the male lions and the prey hunting is usually performed by the lionesses and not lions. Each pride of lions is confined to a selective location. The lionesses encircle the targeted prey to initiate the attack. The input parameters required for this algorithm are initial population with the associated percentage of nomadic and residential lions. In the meantime, the lioness reproduces cubs and the cycle of getting inside and outside of a pride continues.

The optimal solution to a problem is found by arranging the nomadic lions based on their fitness value. When the fitness value of the lions is lower, they are eliminated from consideration. Hence, the lion has to be fitter for capturing a place in the pride. Additionally, it can be considered that the fitter lions alone can have place in the pride. This process is continued until the best possible solutions are attained.

3.2.3. Proposed mammogram segmentation algorithm

This work proposes a segmentation algorithm based on FCM and LOA. FCM is an efficient clustering and the main concern of FCM is the selection of the initial points. When the initial point selection is carried out by some other algorithm, FCM can prove its best. For this sake, this work utilizes LOA which proves better performance. LOA is proven with speedy convergence and better global optimality. This is why LOA was chosen, and the proposed method is as follows.

Proposed Mammogram Segmentation Algorithm with LO optimized FCM

Input: Entities, m, termination condition
Output: Segmented Mammograms (μ_{fsbl})
Begin
  Choose initial points by LOA; $\mu_0 := LOA(entities, m)$
  Do
    Compute $c_j = \frac{\sum_{i=1}^{K} \mu_{ij}^f a_i}{\sum_{i=1}^{K} \mu_{ij}^f}$;
    if (current solution is better than existing $\vee$ *met termination condition)*
      End the process;
    Else
      Compute $\mu_{ij} = 1/\sum_{p=1}^{C} \left(\frac{\|a_i - c_j\|}{\|a_i - c_p\|}\right)^{\frac{2}{f-1}}$
    End if;
  $\mu_{f\,sbl} := \mu_{i,j}$;
  While (termination condition)
End;

The traditional FCM algorithm randomly chooses the initial points, which results in huge count of iterations for reaching the final feasible solution. This in turn consumes more resource, which is not recommendable. This shortcoming of FCM is addressed by means of LOA in which the centroids are chosen by the LOA. Therefore, the combination of FCM and LOA reduces the count of iterations considerably, which in turn conserves the resources and enhances the performance of the algorithm. As the optimal centroids are chosen by LOA before being passed to FCM, faster convergence is experienced. Hence, the area of interest is extracted and the features are then extracted from the segmented images, as presented below.

3.3. Gabor and LTP feature extraction

Gabor filters are excellent descriptors of an image's texture. The Gabor filter is quite common for extracting picture characteristics, particularly texture features. Gabor filters are well-known for their ability to resolve ambiguity in both the spatial and frequency domains. Additionally, they are utilized as scaled and tuned edge detectors. It reduces the combined uncertainty in terms of space and frequency. The Gabor filter is created using the formula below. A two-dimensional Gabor function $g(x, y)$ is applied on the images as follows.

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{1}{2}\left[\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right]+2\pi jWx\right] \tag{5}$$

The modulation frequency is denoted by W and the fourier transform in two dimensions is given by

$$G(u,v) = \exp\left[-\frac{1}{2}\left[\frac{(u-w)^2}{\sigma_u^2}+\frac{v^2}{\sigma_v^2}\right]\right] \tag{6}$$

σ_u is $(2\pi\sigma_x)^{-1}$ and σ_v is $(2\pi\sigma_y)^{-1}$. Consider $g(x, y)$ is the mother wavelet, and many filters with varying rotations can be constructed by (7).

$$g_{mn}(x,y) = a^{-m} g(x', y') \tag{7}$$

Where $x' = a^{-m}(x\cos\theta + y\sin\theta)$,

$y' = a^{-m}(-x\sin\theta + y\cos\theta);\ a > 1,$

$\theta = n\pi/N;\ \ n = 0, 1, 2, ..N-1;\ \ m = 0, 1, 2, ..M-1$

where$n$ and m are the orientation and scale of the Gabor wavelet. In this work, we consider 3×3, 5×5, 7×7, 9×9 as different sizes of Gabor, theta values considered by this work are 15, 45, 75, 135 and 180 degrees. Finally, each picture pixel is composed of two components: real and imaginary. The real component denotes magnitude, whereas the imaginary part denotes phase characteristics. This is followed by the extraction of LTP features.

Three multi-resolution pictures are used to extract LTP characteristics using a two-dimensional circular symmetric Gaussian filter bank.

$$G(a,b,\mu) = \frac{1}{2\pi\mu^2} e^{-(a^2+b^2)/2\mu^2} \tag{8}$$

Gaussian pictures are generated using

$$GI(\mu) = G(a,b,\mu) * I(a,b) \tag{9}$$

The scales in the preceding equations are denoted by μ, the scaling factors are indicated by a,b, and the convolutional operator is denoted by $*$. Three distinct multiresolutional pictures are produced in a three-dimensional grid, and the eight neighbours of a central pixel are calculated in each of the five directions D. After that, the feature vector is constructed by concatenating all the histograms. The so computed feature vector is utilized for training the SVM classifier, as explained in the next section.

3.4. SVM classification

SVM is a supervised classification method that uses a boundary to categorize objects. Binary SVM, on the other hand, is not practical for tasks with many categories. In this instance, a multiclass SVM is used. Multiclass SVM is used in this study because it examines three distinct classes: Normal, Benign and Malignant. Hence, $\frac{n(n-1)}{2}$ classifiers are employed and the final decision of each classifier is considered. Finally, the results are declared with the maximum-voting policies [21]. Thus, by solving the following equation, all of the distinct classes are handled concurrently.

$$\min_{nh,b,sv}\frac{1}{2}\sum_{y=1}^{q} nh_y^p nh_y + c\sum_{i=1}^{r}\sum_{y\neq s_i} sv_{i,y} \tag{10}$$

Here, nh appears to be normal to the hyperplane in this case, b denotes the bias, sv denotes the slack variable, and $i = 1, 2, \ldots r$ is the number of training samples and y denotes the number of classes. The following equation makes the definitive conclusion.

$$\text{decn} = \max_y(w_y^p \beta(x_i) + b_y) \tag{11}$$

In this method, all classifiers are used on every pair of classes. Consider the following object obj, which must be classified into one of three distinct classes (say, x, y, z). This is achieved by superimposing all classifiers on a picture. When a classifier classifies an item as belonging to class x, the value of class x is increased by one. The ultimate classification choice is made based on the class with the most votes. This categorization method results in an accurate choice in an acceptable amount of time. By following this way, the mammogram image is classified into either of three classes and the performance of the proposed approach is analysed in the following section.

4. Results and discussion

This work is simulated in a stand-alone computer with 16 GB RAM using Matlab R2017B upon the benchmark dataset MIAS. This dataset mainly involve three classes such as normal, malignant and benign [22]. The MIAS dataset is composed of 326 images, which fall into three categories such as fatty, fatty and dense glandular. Table 1 presents the dataset information.

Table 1. Dataset information.

Dataset	MIAS
Normal	207
Benign	68
Malignant	51
Total Images	326

All the images are eight-bit gray scale images of size 1024 × 1024. This work utilizes 60% of the images for training and the remaining images for testing. The performance of the proposed work is evaluated for highlighting the effectiveness of segmentation by FCM + LOA, feature extraction by Gabor + LTP and classification techniques.

Classification accuracy is the statistic used to assess the suggested approach's correctness. Classification accuracy focuses on the best classification of a mammography image in this work. Classification accuracy is

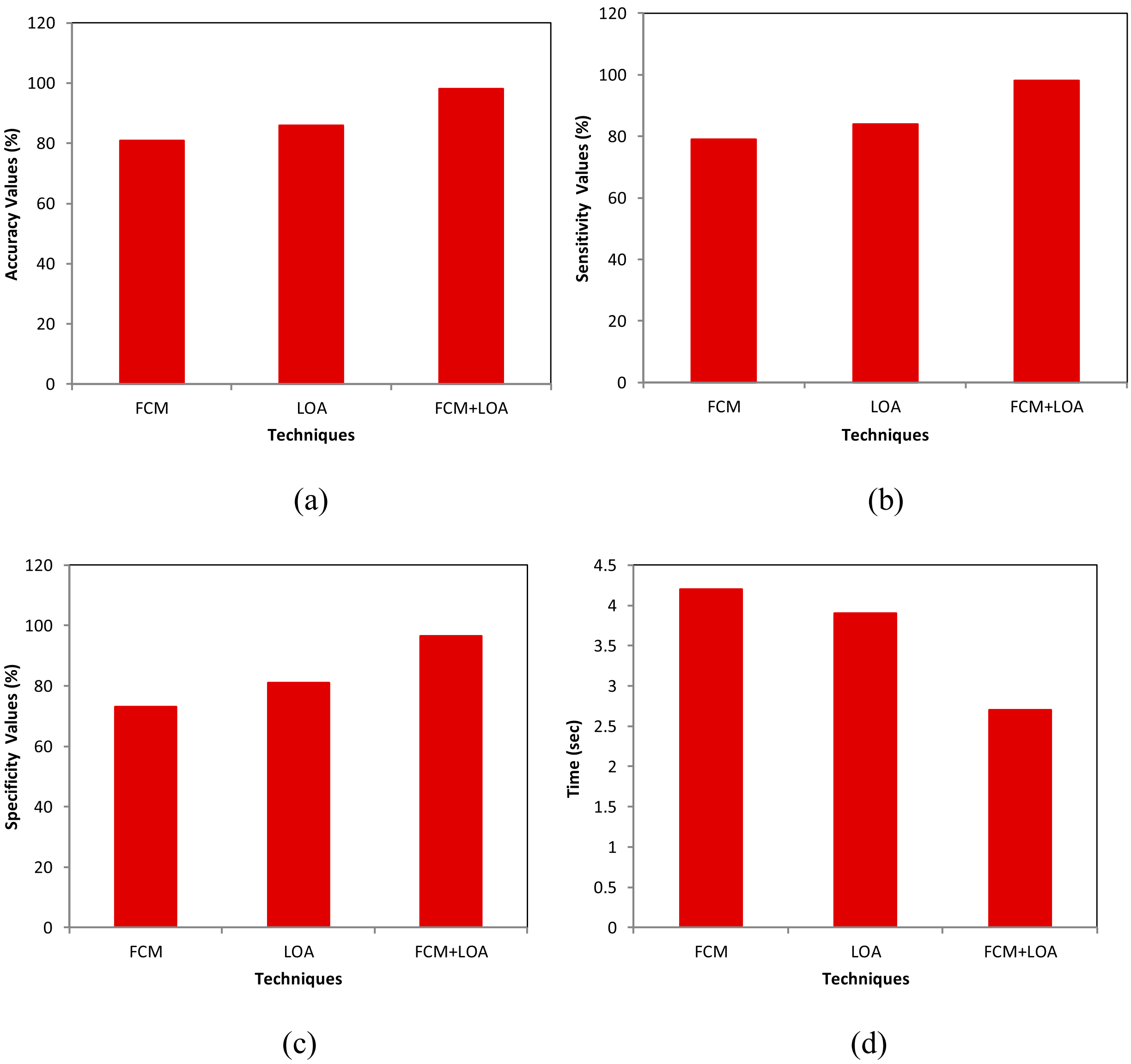

Figure 2. Analysis on segmentation performance (a) Accuracy (b) Sensitivity (c) Specificity (d) Time consumption.

the most critical measure, since it determines the system's efficiency. The accuracy rates are calculated as follows.

$$A_r = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (12)$$

Sensitivity and specificity are two more critical criteria that are difficult to achieve. The reason for this is that the majority of classification systems are designed to enhance accuracy, which results in a rise in FP and FN rates. This is a significant issue, particularly for healthcare systems, because FP rates imply that the system incorrectly classifies a normal mammography as malignant. Similarly, when the system considers a malignant picture to be normal, FN rates increase. The rates of sensitivity and specificity are highly correlated with the rates of FN and FP, respectively. When FN rates grow, the system's sensitivity rate decreases. Similarly, when the FP rates of the system are decreased, the specificity of the system rises. The suggested approach's sensitivity and specificity are calculated using

$$sn_r = \frac{TP}{TP + FN} \times 100 \quad (13)$$

$$sp_r = \frac{TN}{TN + FP} \times 100 \quad (14)$$

In the above equations, TP, TN, FP and FN are True Positive, True Negative, False Positive and False Negative rates respectively.

4.1. Performance validation of segmentation approach

Segmentation is the most important process for image processing application, specifically when it deals with detection system. Hence, this work focuses more on segmentation, in order to arrive at better classification results. A better segmentation algorithm results in minimal time consumption and computational complexity. The supremacy of the combination of FCM and LOA is proven by the attained results, which is illustrated in Figure 2.

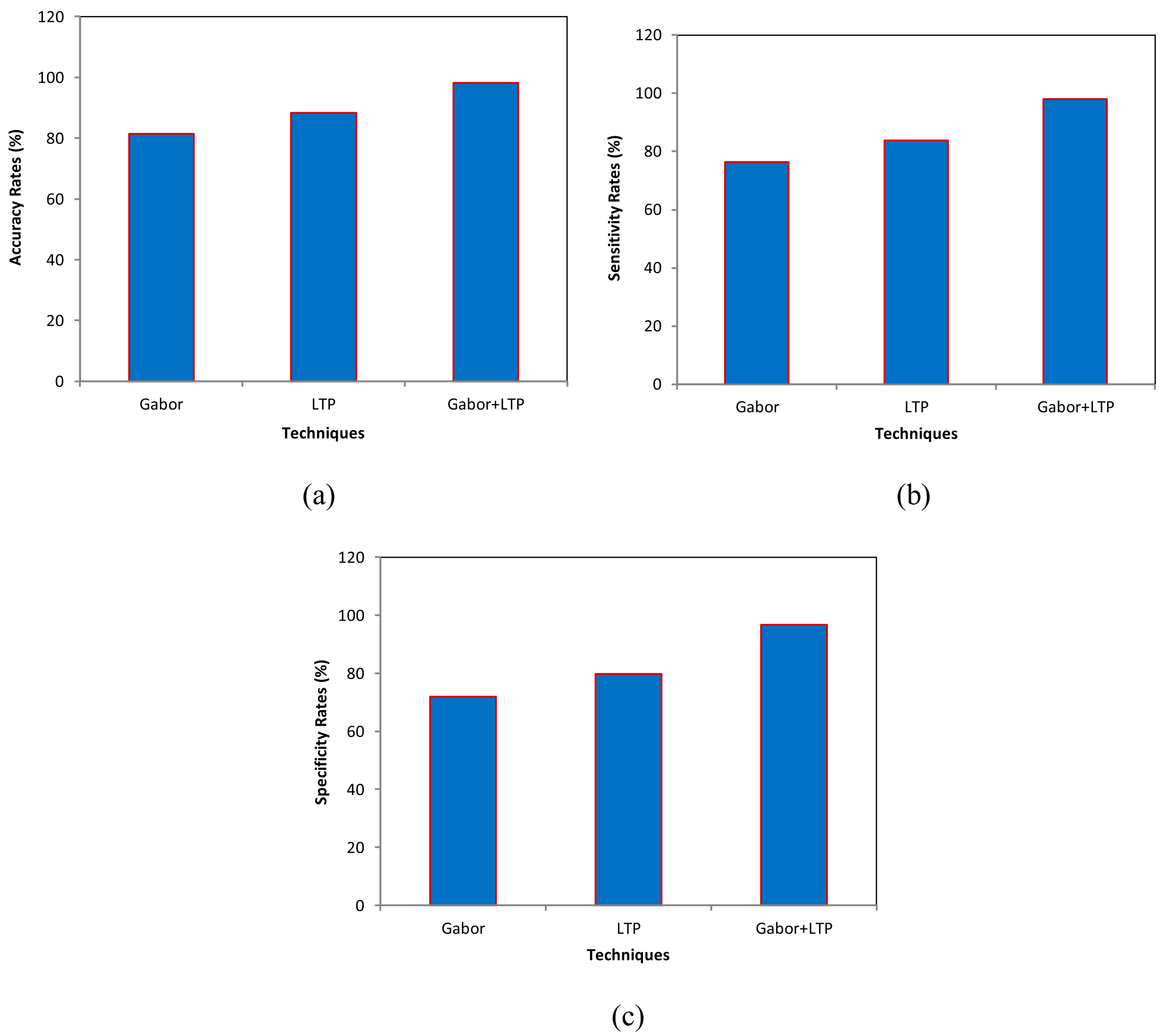

Figure 3. Analysis on feature extraction performance (a) Accuracy (b) Sensitivity (c) Specificity.

From the experimental analysis, it is evident that the performance of the combination of FCM and LO algorithm performs well than when they are deployed alone. FCM has proven itself for clustering but suffers from the initial cluster point selection and this issue is addressed by employing LO algorithm. The LO algorithm chooses the best initial seed point for clustering and thus, the efficiency of FCM is enhanced. The sensitivity and specificity rates of the proposed work are 98 and 94%, respectively. This indicates that the false positive and false negative rates of the proposed work are minimal, when compared to the comparative techniques. The following section analyses the work performance with respect to the feature extraction techniques.

4.2. Performance evaluation of feature extraction technique

Though the areas of interest are successfully extracted, it is quintessential to extract significant features, as these features are utilized for training the system. This work extracts the Gabor and LTP features for training the CADS and the efficiency of the employed feature extraction technique is illustrated in Figure 3.

The experimental results show the efficiency of the combination of Gabor and LTP with maximal accuracy, sensitivity and specificity rates. These results are computed by training the SVM classifier with Gabor features, LTP features and the combination of both. Hence, the reason for the choice of these feature extraction techniques is justified and the following section proves the performance of the classifier.

4.3. Performance evaluation of the classifiers

Classification is the most important stage of any CADS, as it declares the final result of any problem. In spite of the inclusion of efficient segmentation and feature extraction techniques, the choice of classifier is crucial, as the learning ability of the classifier determines the effectiveness of the system. For comparison, the classifiers being considered are k-NN [23] and Relevance Vector Machine (RVM) [24]. The results with respect

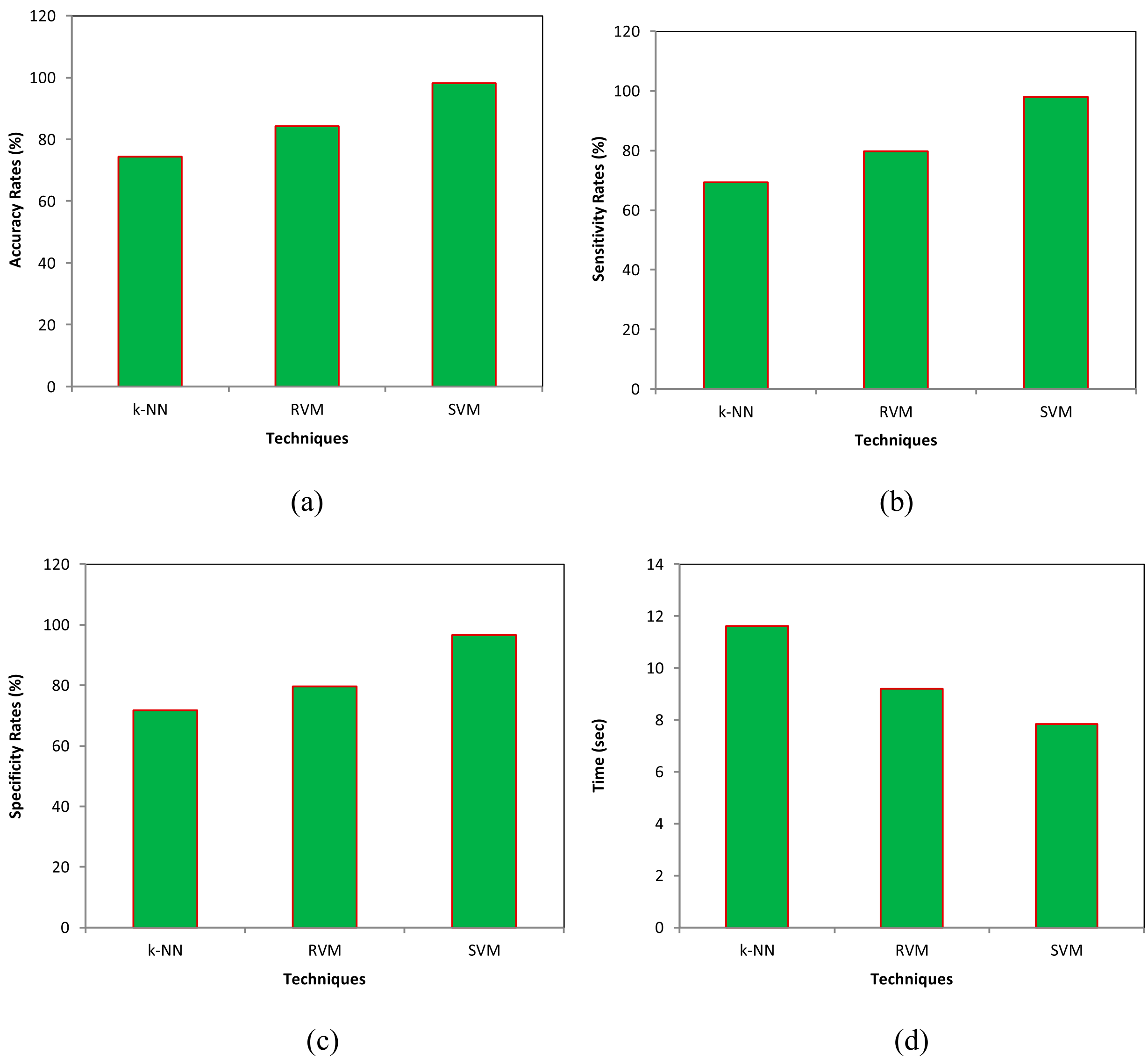

Figure 4. Analysis on classification performance (a) Accuracy (b) Sensitivity (c) Specificity (d) Time consumption.

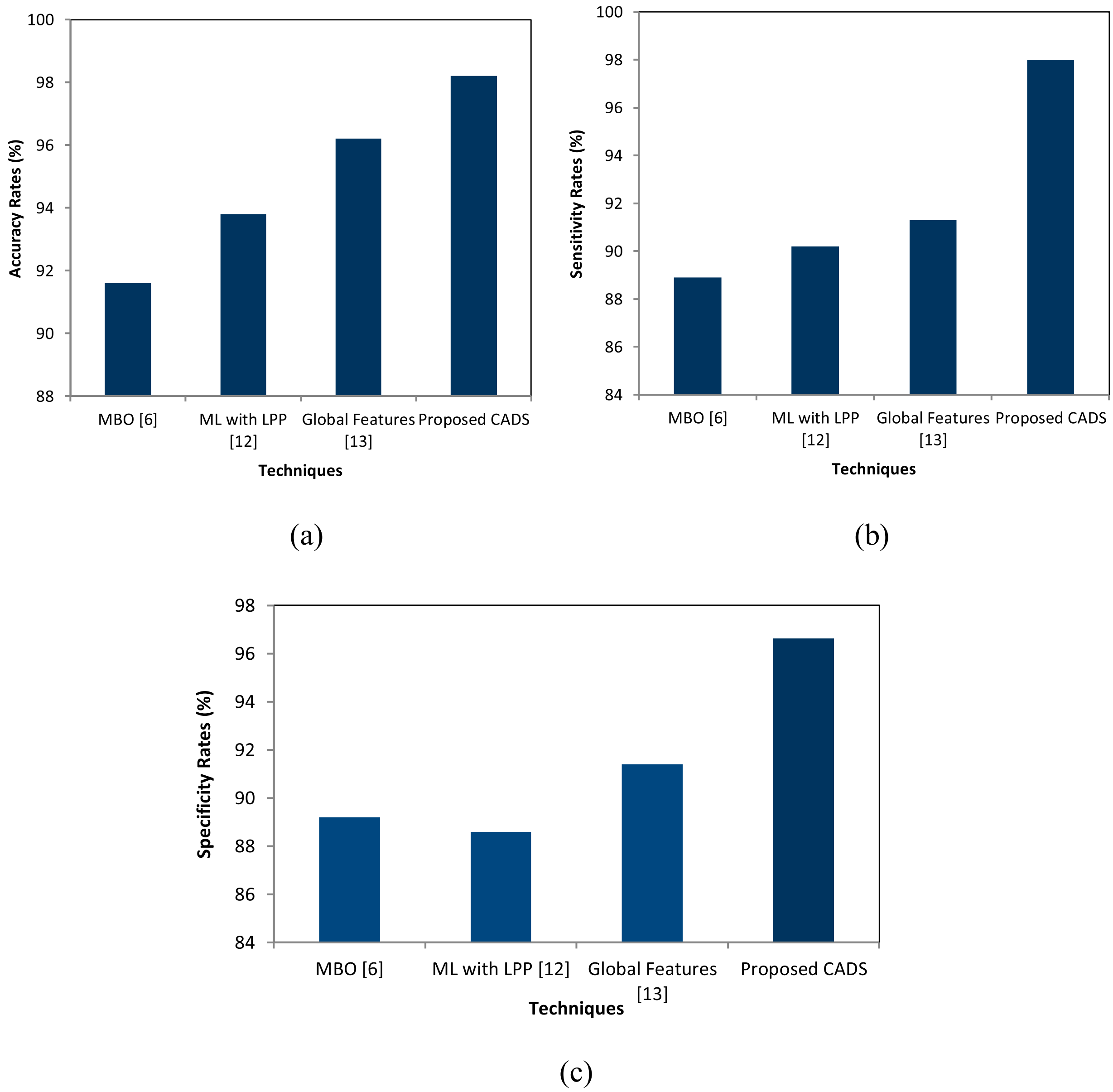

Figure 5. Comparative analysis with existing approaches (a) Accuracy (b) Sensitivity (c) Specificity.

to the classification techniques are depicted as follows (Figure 4).

Observation demonstrates that SVM outperforms k-NN and RVM classifiers. SVM outperforms k-NN and RVM in terms of accuracy, sensitivity and specificity. On the other hand, the SVM classifier consumes significantly less time than the k-NN and RVM classifiers. The reason for this is because k-NN classification depends entirely on the distance measure to discriminate between images, and so consumes more time than any other classifier. RVM is based on probability, and hence performs better than k-NN but not as well as SVM.

4.4. Performance evaluation with state-of-the-art approaches

The performance of the proposed CADS is compared with the existing approaches such as Monarchy Butterfly Optimization (MBO) technique [6], ML with LPP [12] and global features [13]. The average results attained by these works are shown in the following Figure 5. Table 2 tabulates the results.

From the experimental results, it is evident that the proposed "CADS" performs better than the existing approaches. The reason behind the better performance of the proposed work is the effective pre-processing activity by LWT, optimized segmentation procedure performed by LO optimized FCM algorithm, better features extracted by Gabor with LTP and SVM classification. In this work, optimized segmentation plays a major role which effectively extracts the area of interest and the potential features are extracted by which the SVM classifier is trained. The greatest accuracy rate attained by the proposed work is 98.2%, which is greater than the existing works. All the computed results are the average of five rounds of execution.

Table 2. Performance results.

Techniques/perf.measures	Accuracy (%)	Sensitivity (%)	Specificity (%)	Time consumption (ms)
MBO [6]	91.6	88.9	89.2	2438
ML with LPP [12]	93.8	90.2	88.6	2108
Global features [13]	96.4	92.1	91.4	1985
Proposed CADS	98.2	97.9	96.6	1796

5. Conclusions

This article presents a "CADS" for mammogram images for breast cancer detection and classification. The complete work is segregated into four major phases, where each phase plays a significant role in achieving better outcome of the proposed work. LWT is employed for pre-processing and the segmentation is carried out by the LOA optimized FCM algorithm. The Gabor and LTP features are extracted from the segmented images and the classification is performed by SVM. The performance of the proposed CADS is observed to be satisfactory with greater accuracy, sensitivity and specificity rates in acceptable time period. In future, the images to be processed can be increased and the optimized machine learning algorithm can be incorporated for analysis. This work can be employed in real-time environment for aiding the healthcare professionals.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Authorship contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by T. Arunprasath. The first draft of the manuscript was written by S. Balaji and the co-authors have assisted on the preparation of the manuscript. All authors read and approved the final manuscript.

References

[1] https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[2] Yasmin M, Sharif M, Mohsin S. Survey paper on diagnosis of breast cancer using image processing techniques. Res J Recent Sci. 2013;2(10):88–98.

[3] Mohan H. Textbook of pathology. 7th ed. New Delhi, India: Jaypee Brothers Medical Publishers (P) Ltd.; 2014.

[4] Cheng H, Cai X, Chen X, et al. Computer-aided detection and classification of microcalcifications in mammograms: a survey. Pattern Recognit. 2003;36(12): 2967–2991.

[5] Veena M, Padma MC. Detection of breast cancer using digital breast tomosynthesis. In: Sridhar V, Sheshadri HS, Padma MC, editors. Emerging research in electronics, computer science and technology. Singapore: Springer; 2019. p. 721–730. (Lecture notes in electrical engineering). doi:10.1007/978-81-322-1157-0

[6] Punitha S, Amuthan A, Suresh Joseph K. Enhanced monarchy butterfly optimization technique for effective breast cancer diagnosis. J Med Syst. 2019;43(7):206 p. doi:10.1007/s10916-019-1348-8

[7] Sakri S, Abdul Rashid NB, Muhammad Zain Z. Particle swarm optimization feature selection for breast cancer recurrence prediction. IEEE Access. 2018;6:29637–29647. doi:10.1109/ACCESS.2018.2843443

[8] Karthik S, Srinivasa Perumal R, Mouli C. Breast cancer classification using deep neural networks. In: Margret Anouncia S, Wiil UK, editors. Knowledge computing and its applications. Singapore: Springer; 2018. p. 227–241. doi:10.1007/978-981-10-6680-1_12

[9] Unni A, Nidheep E, Vinod S, et al. Tumour detection in double threshold segmented mammograms using optimized GLCM features fed SVM. Proceedings of the 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI); IEEE, Bangalore, India; September 2018. p. 554–559.

[10] Selvathi D, Aarthy Poornila A. Deep learning techniques for breast cancer detection using medical image analysis. In: Hemanth J, Balas VE, editors. Biologically rationalized computing techniques for image processing applications. Berlin, Germany: Springer; 2018. p. 159–186. (Lecture notes in computational vision and biomechanics 25). doi:10.1007/978-3-319-61316-1_8

[11] Sasikala S, Ezhilarasi M. Fusion of k-Gabor features from medio-lateral-oblique and craniocaudal view mammograms for improved breast cancer diagnosis. J Cancer Res Aerapeutics. 2018;14(5):1036 p.

[12] Heidari M, Hollingsworth AB, Danala G, et al. Prediction of breast cancer risk using a machine learning approach embedded with a locality preserving projection algorithm. Phys Med Biol. 2018;63(3):Article ID 035020. doi:10.1088/1361-6560/aaa1ca

[13] Tariq N, Abid B, Ali Qadeer K, et al. Breast cancer classification using global discriminate features in mammographic images. Int J Adv Comput Sci Appl. 2019;10(2). doi:10.14569/IJACSA.2019.0100250

[14] Kashyap KL, Bajpai MK, Khanna P, et al. Meshfree based variational level set evolution for breast region segmentation and abnormality detection using mammograms. Int J Numer Method Biomed Eng. 2018;34(1):Article ID e2907. doi:10.1002/cnm.2907

[15] Chowdhary CL, Acharjya DP. Segmentation of mammograms using a novel intuitionistic possibilistic fuzzy c-mean clustering algorithm. In: Panigrahi BK, Hoda M, Sharma V, et al., editors. Nature inspired computing. Singapore: Springer; 2018. p. 75–82. (Advances in intelligent systems and computing 652) doi:10.1007/978-981-10-6747-1_9

[16] Rampun A, Scotney BW, Morrow PJ, et al. Breast mass classification in mammograms using ensemble convolutional neural networks. Proceedings of the 2018, IEEE 20th International Conference on E-Health Networking, Applications and Services (Healthcom). IEEE, Ostrava, Czech Republic, September 2018; p. 1–6.

[17] Bezdek JC, Ehrlich R, Full W. FCM: The fuzzy c-means clustering algorithm. Comput Geosci. 1984;10(2–3):191–203. doi:10.1016/0098-3004(84)90020-7

[18] Zhou K, Fu C, Yang S. Fuzziness parameter selection in fuzzy c-means: the perspective of cluster validation. Sci China Inf Sci. 2014;57(11):1–8.

[19] Mccomb K, Pusey A, Packer C, et al. Female lions can identify potentially infanticidal males from their roars. Proc R Soc Lond Ser B: Biol Sci. 1993;252(1333):59–64. doi:10.1098/rspb.1993.0046

[20] Schaller GB. The Serengeti lion: a study of predator–prey relations. Wildlife behavior and ecology series. Chicago (IL): University of Chicago Press; 1972.

[21] Hsu C-W, Lin C-J. A comparison of methods for multiclass support vector machines? IEEE Trans Neural Netw. 2002;13:415–425. doi:10.1109/72.991427

[22] https://www.kaggle.com/datasets/kmader/mias-mammography

[23] Warfield S. Fast k-NN classification for multichannel image data. Pattern Recognit Lett. 1996;17(7):713–721. doi:10.1016/0167-8655(96)00036-0

[24] Tipping ME. Sparse Bayesian learning and the relevance vector machine. J Mach Learn Res. Jun 2001;1:211–244.