VOL. 18, NO. 21, NOVEMBER 2023 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

GENETIC ALGORITHM BASED DETECTION OF BREAST CANCER USING LEAST SQUARE-SUPPORT VECTOR MACHINE CLASSIFIER

K. Gowrishankar[1], Sudhakara Pandian[2], G. Narmadha[3], Muthukumar Vellaisamy[4], S. Deivasigamani[5] and Manickam Ramasamy[5]
[1]Department of Electrical and Electronics Engineering, AMET University, Chennai, Tamil Nadu, India
[2]Department of Manufacturing Engineering, VIT University, Vellore Campus, India
[3]Department of Electrical and Electronics Engineering, Sethu Institute of Technology, Virudhunagar, Tamil Nadu, India
[4]Faculty of Computing and Engineering, QUEST International University, Ipoh, Malaysia
[5]Faculty of Engineering, Technology and Built Environment, UCSI University, Kuala Lumpur, Malaysia
E-Mail: gowrishankark@ametuniv.ac.in

ABSTRACT

Breast tumors are a dangerous disease among women worldwide. They are the second leading cause of death among all forms of cancers in women. Their early detection is critical to increasing the survival rate of women. Mammography is a reliable screening technique in the early detection of abnormal breast tissue severity. Radiologist abnormalities in the breast tissue, radiologists employ mammography. However, detecting breast abnormalities through digital diagnostic techniques by a radiologist could be time consuming. Consequently, computerized studying of digital mammography has emerged via the development of CAD systems. Several CAD systems have been developed for breast cancer detection. However, obtaining a satisfactory performance of CAD systems is a challenging task. We propose a CAD architecture for the classification of breast tissues as either benign or malignant using an LS-SVM classifier with various kernels namely linear, quadratic, polynomial, MLP, and RBF kernels. From the experimental outputs, it is clear that GA based LS-SVM classifier with RBF kernel outputs classification accuracy of 94.59% for normal/abnormal case classification is better, when it is compared with all other kernels. It is also stated that GA based LS-SVM classifier with RBF kernel produces a better classification accuracy of 98.26% for benign/malignant case classification when it is compared with other reported works.

Keywords: breast cancer detection, genetic algorithm, shearlet transform, LS-SVM classifier, mammographic image analysis society.

Manuscript Received 25 July 2023; Revised 3 December 2023; Published 10 January 2024

1. INTRODUCTION

Breast malignant growth is a significant health complication that is predominantly basic and has a high mortality rate, particularly among women. According to the International Agency for Research on Cancer (IARC 2012), approximately 1.7 million people perceive bosom malignant growth on a worldwide scale [1]. In 2015, 40290 lady's passing was recorded because of bosom malignant growth [2]. Mammography is the most reliable tool for the recognition of bosom malignant growth during the initial period [3]. It can recognize 85-90 % of all bosom diseases. A significant indication of breast diseases is microcalcification and mass [4]. The size, shape, texture distribution, and margin attributes that are difficult to detect are predominantly utilized to make the decision. Hence, precise detection of abnormal cases is a critical process in CAD layout. Table-5 presents the various abnormalities that form the basis of breast masses that lead to cancer. In healthcare applications, the performance of a diagnosis system is based on significant feature extraction from the medical images available in real-time and the feature vectors fed into the classification system. The classifier performance for breast cancer diagnosis depends on the correct number of feature vectors provided as input. Thus, feature selection reduces the burden on the classification system in terms of complexity and computation time. A considerable reduction in the dimensionality is also required. Feature selection is considered the preprocessing step of the classification system, and it removes noisy, irrelevant, and redundant information. Genetic algorithms (GAs) are stochastic methods that follow the analogy of biological processes. They are used in various real-time optimization problems and have been considered for the rapid advancement of health systems. They follow chromosomal structures and are inspired by the natural evolution process to derive the optimal solution. The GA process involves selection, crossover, and mutation operations capable of eliminating the local optimal solutions to derive the global optimum solutions. A GA works through an iterative process in which each iteration generates better chromosomes. This GA strategy facilitates the determination of the significant features necessary for a successful classification process. The GA process for the selection of better chromosomes at each iteration is depicted in Figure-1.

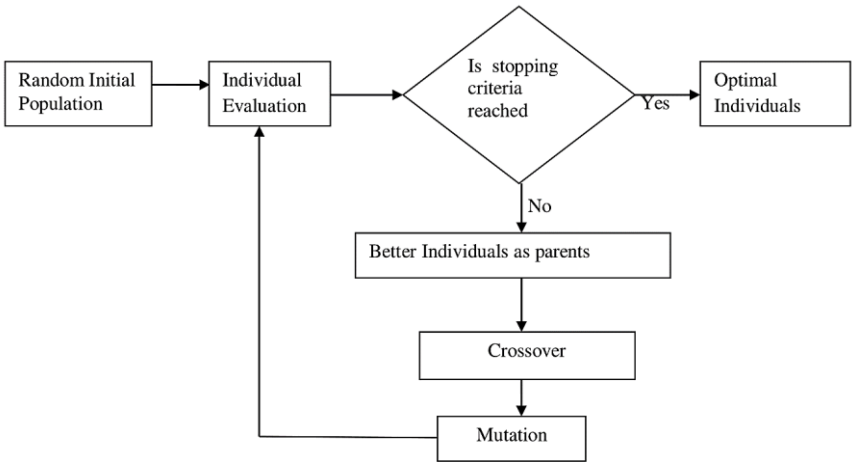

Figure-1. Steps involved in the GA process.

2. RELATED WORKS

Deepa *et al*. [5] proposed a mammogram image classification methodology using the contour let transform and probabilistic neural network (PNN). The region of interest (ROI) image was segmented from the actual image, primarily based on the abnormal details provided by the MIAS information file. Initially, the ROI was decomposed, and then the contour let coefficient was computed. From the contour let, a co-efficient co-occurrence matrix was produced, following which appropriate features were selected. PNN was employed as a classifier for the classification task. In [6], the CAD layout was explained to categorize breast tumors using Zernike moments [6]. Noise suppression is one of the foremost image-preprocessing techniques. The input image was pre-processed to smoothen the image edges using a median filter [7]. Jen *et al*. [8] reported a CAD model for the detection of abnormal mammograms by using principal component analysis and a two-stage classifier.

Various image enhancement techniques were proposed in [9, 10] to increase the visual quality of images. The latest research has revealed that using the wavelet transform highly enhances image quality. The CAD system presented by Hu *et al.* [11] afforded a sensitivity of 91.3%. It used both local and global adaptive thresholding methods for image segmentation. The work was tested using MIAS dataset images, and 172 mammograms were considered. Tang *et al*. [12] summarized an outline of the latest developments in breast tumor discovery strategies and related methods.

Moayedi *et al*. [13] provided breast cancer detection and classification methods using the discrete contour let transform and several machine learning techniques. The ROI was segmented from the original image by removing the pectoral muscles. Subsequently, the contour let transform was applied to the ROI image for decomposition, and then features were extracted from the contour let coefficients. A GA was applied for feature selection. Finally, the mass image was classified using three different classifiers, and the accuracies obtained were compared. Weighted support vector machine (SVM), SVFNN, and SVM classifiers with RBF kernel obtained accuracies of 96.6%, 91.5%, and 82.1%, respectively. Talha [14] reported a CAD system for mammogram identification and classification using genetic programming, various transforms, and an SVM classifier. DWT and DCT were used for feature generation, and genetic programming was used for feature selection. A mammogram image was identified as either normal or abnormal using an SVM classifier. The work was tested using MIAS database mammogram images and achieved good classification accuracy for normal and abnormal cases. A deep CNN for breast cancer screening and classification was explained in [15], and it obtained an AUC of 0.895. In [16], three different algorithms, namely, J48, NB, and SMO were used on two different datasets and achieved accuracies of 74.82%, 75.53%, and 72.66%, respectively. The naive Bayes (NB) classifier and k-nearest neighbor (KNN) approach employed in [17] proved that KNN yields a better result than NB. In [18], the backpropagation classification accuracy, neural network model, and logistic regression models were evaluated. A softmax discriminant classifier and linear discriminant analysis are utilized [19] to classify breast tumors. Five different machine-learning techniques were explained [20] for tumor prediction, and it was concluded that the support vector machine (SVM) was the best model. The image was segmented using the region growing method [21]. Various features were extracted from the segmented area, and an artificial neural network was then used as a classifier. Additionally, the classification results were compared with those of other competing classifiers. The work in [22] improved the ROI contrast/distinction by employing nonlinear polynomial filters. A multi-model image fusion method using wavelet transforms and the contour let transform was proposed in [23]. The method was mainly developed to enhance the contract of features. The CAD designed in [24] was used to classify abnormalities as either malignant or non-malignant. Initially, mammogram images were pre-processed, and then the ROI images were obtained. From that ROI, Zernike moments were extracted, and an SVM was used to classify the CAD images presented in [25]. It

was implemented using SVMs with different kernel functions. In this study, feature subset selection is performed by a GA. This work compares the performance of different kernels used in SVM and compares the SVM performance with that of a neural network. Five different feature extraction methods were compared in [26]. Additionally, three different classifiers were employed to classify mammogram images and compare them using evaluation metrics. An MIAS database image was used for evaluation, and it was concluded that GLDM is the best feature extraction method among the five mentioned methods. The CAD system developed in [27] was used to differentiate mass images from normal ones. This study was tested using DDSM database images. First, the images were pre-processed, and then the ROI was extracted. Various features were generated from the ROI images. Sequential forward selection and sequential floating forward selection methods were employed to select the feature vector moments. Finally, different classifiers were used for classification, and their performances were compared. In [28], CAD was implemented using a Curvelet transform to differentiate normal images from abnormal types, and then the abnormal images were further categorized as either malignant or benign tissue. Feature selection was performed using PCA and LDA, and then SVM and KNN classifiers were utilized to calculate the precision value. The work evaluated 200 mammogram images taken from the MIAS database. Automated segmentation and mass classification in a mammogram approach was demonstrated in [29]. First, the ROI was separated from the actual image using chain codes, and it was reinforced using a rough set. Subsequently, the ROI was segmented. Second, 32 features were derived from the partitioned mass region. RFT was used for the classification task. The work output was compared with those of an SVM, a GA-based SVM, a particle swarm optimization-based SVM, and a decision tree. Matthew's correlation coefficient was used as the performance metric. This method was evaluated using two databases, namely, MIAS and DDSM. The authors of [30] introduced a procedure for breast mass detection in asymmetric regions by using different functions formed with a spatial description with a variogram and cross-variogram.

A topological modeling-based algorithm was used in [31] to detect and identify microcalcifications using the KNN classifier. Topological features were generated using multiscale morphology. A group of topological-based features was formed from a graph built using the connectivity of the clusters of the microcalcifications. Classification of breast masses using correlated rule mining was explained in [32]. The ROI was obtained from a pre-processed image. The statistical and textural features were generated using the GLCM method. The breast masses were then classified as normal, benign, or malignant by applying correlated association rule mining. Subsequently, features are fed as input to the KNN classifier to categorize the microcalcifications as benign or malignant. The AdaBoost-based method was proposed in [33] to identify malignant masses, in which the ROI was separated from the processed original image. The stellate features were further extracted from three sub-regions: the core, inner, and outer parts, using statistical characteristics for individual sub-regions. Finally, AdaBoost was employed to group the image as that of a normal, benign, or malignant tumor.

3. PROPOSED METHOD

The proposed methodology for breast cancer recognition is executed progressively in different stages. The proposed framework is shown in Figure-2. Various stages of the proposed framework begin with the extraction of the ROI, which contains the abnormal regions. Subsequently, a shearlet transform-based ROI image transformation is performed to extract the statistical texture moments from the shearlet coefficients. Furthermore, feature reduction is performed using GA to select the most significant features. The last stage is classification using LS-SVM with various kernels.

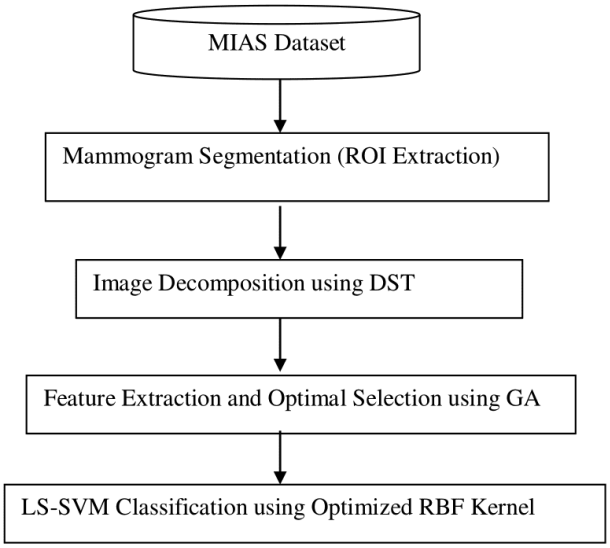

Figure-2. Proposed GA-optimized diagnosis system.

3.1 Extraction of ROI

The mammogram images available in the MIAS database are large (1024 × 1024) and contain noise and different unwanted details, including labels and artifacts. Here the region of interest (ROI) image was segmented from the actual image, based on the abnormal details provided by the MIAS dataset information file. Consequently, an ROI image of size 256×256 was segmented by eliminating the unwanted background details through manual cropping. Figure-3 shows the original mammogram and the corresponding ROI image.

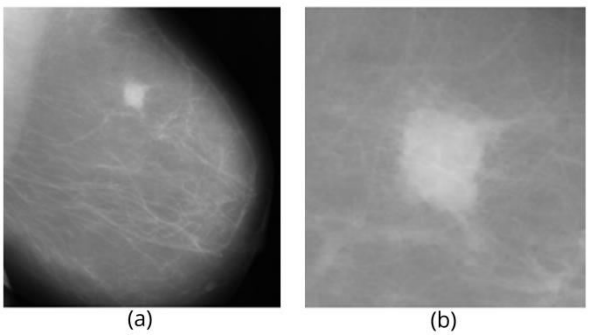

Figure-3. (a) Original image. (b) ROI image.

3.2 Image Decomposition Using DST

DST is a mathematical structure combined with a multiresolution analysis technique and it also converts this digital information into its equivalent frequency domain by partitioning the image pixel matrix into blocks of size depending upon the type of image. The theory of composite wavelets, introduced by Guo *et al*. (2006) [13], presents an efficient approach to combining geometry and multiresolution analysis with affine system concepts. The shearlet transform is predominantly well-established to implement a highly well-organized representation of images with edges. These transform elements represent a form of well-localized waveforms that have numerous variant locations, scale orientations, and anisotropic shapes. Consequently, the shearlet representations are especially well-fitted for representing the edges and other anisotropic objects, and powerful features in typical images. The DST is depicted in Figure-4.

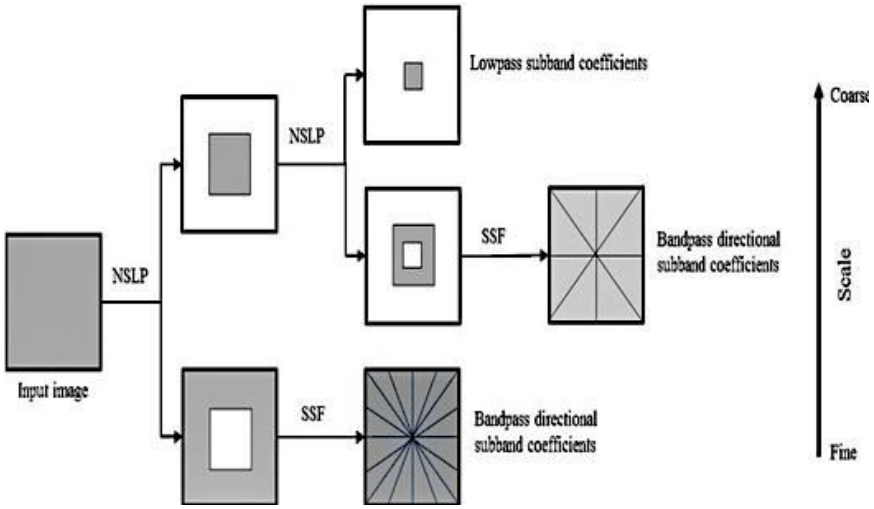

Figure-4. Diagrammatic representation of DST.

Using DST [17], the extracted ROI is decomposed into shearlet coefficients, known as sub-bands. Image decomposition using DST is shown in Figure-5. Each shearlet coefficient is a component of the ROI image at a particular scale and direction. For example, if the ROI image is decomposed using the shearlet transform with level 2 and direction 2, 5 different shearlet coefficients are obtained (see Figure-6). The lower sub-band design tests the approximation ROI image, and the remaining four specify the detailed information of that ROI image.

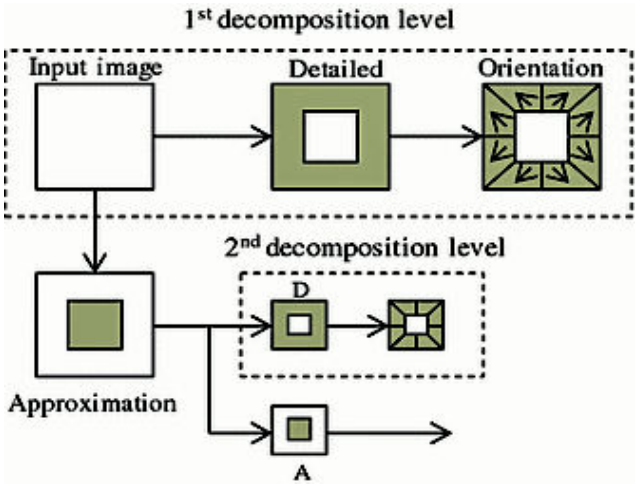

Figure-5. Image decomposition using two-level DST.

3.3 Feature Generation and Optimal Feature Set Selection

Feature extraction is an essential component of CAD development. Various shearlet moments are derived from the decomposed image by a statistical approach. Statistical texture features are primary and are frequently used in medical image analysis. GLDM is the best feature extraction method which is used in the proposed work to extract features in mammogram images for analysis and identification of microclassification.

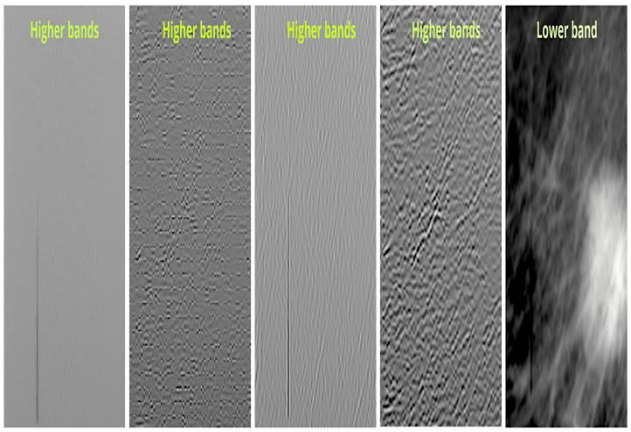

Figure-6. The output of two-level DST of an MIAS image.

GA is an evolutionary computation method superior to the existing conventional approaches. It is a meta heuristic search technique that randomly builds an initial solution or initial population. Subsequently, each member of the initial population is evaluated using a fitness function. Based on the results of the fitness function, a new population is generated. This method then continually seeks -out the most effective feature set, and these steps are shown in Figure-6. The optimal feature subset selection technique is intended to find the discriminant feature by eliminating irrelevant and redundant features. In this study, a GA was used as a feature-reduction algorithm to find the optimal subset features by eliminating insignificant features. The algorithm for the GA-based feature selection process involved in this study is presented in Algorithm 1.

Algorithm 1: Feature selection using a GA

a. Representation of the initial population

The solution (population) of the feature selection process follows a binary form of representation in 0 and 1. The binary digit 0 indicates that the features are considered for selection, and the binary digit 1 indicates that the features are not considered for selection. Initially, X bits were considered to represent the total size of the available features before the selection process.

b. Decoding process

In the decoding process, the optimal features selected by the GA algorithm are decoded, and the features that are not selected are removed from the entire feature set. Only the selected features are kept for the classification process.

c. Training process

The features selected by the GA algorithm are fed as input to the SVM classifier employed for the classification of benign and malignant tumors. During the training process, a validation set is used to avoid overfitting problems. The SVM classifier performance is measured during the iterative process, and if the performance degrades or does not upgrade, the training process is stopped.

d. Fitness evaluation

The validation set is used for calculating the fitness of the trained network.

e. GA operations

The GA process involved is as follows:

Elitism process: The better chromosomes are copied to the next generation, eliminating the worst chromosomes according to the "survival of the fittest" rule. An appropriate elitism size, which is three, is used for the elitism process.

Roulette wheel selection: Using a probabilistic method, the best parental chromosomes are selected and given to the mating pool. Furthermore, among the parents selected, only those with high fitness are eligible for admission to the mating pool.

Crossover and mutation processes: Chromosomes are selected according to fitness, and then crossover operations are performed on the first and second better chromosomes. It proceeds through an iterative process until all the chromosomes are cross-covered. The resultant chromosomes were mutations based on the mutation rate.

Generation of the new population: Better chromosomes are selected from the resultant pool after the elitism process, which eliminates the worst chromosomes. The feature selection measures such as information gain, gain ratio, and index result in over fitting problem. Whereas the Genetic algorithm is naturally inspired and provides a stochastic optimization. GAs use probabilistic transition rules rather than deterministic rules. As the genetic algorithm is a stochastic optimization method, the genes of the individuals are usually initialized at random. Genetic algorithms operate on a population of individuals to produce better and better approximations. They use processes of selection, cross-over, and mutation to get to optimal solutions. The subsets of variables selected by genetic algorithms are generally more efficient than those obtained by classical methods of feature selection, with large features. When compared with other feature selection techniques, the genetic algorithm results in better performance and can manage data sets even with few features, and GA itself is a parallelized algorithm to further speed up the feature selection process.

In this work, GA is applied to select the optimal features with the help of the tournament selection method and the size of the tournament is 2. The input value assigned for population size, population type, and the number of generations is 20, bit string and 20 respectively. Then the uniform mutation and arithmetic crossover operations are performed, and the probability of mutation and the probability of crossover are 0.1 and 0.8 respectively.

LS-SVM Performance evaluation: The testing set was used to determine the accuracy of the SVM classification process using statistical measures, namely, accuracy, sensitivity, and specificity. The optimal

solutions (chromosomes) selected after the GA process were used to evaluate the LS-SVM classifier performance by feeding the mas input to the classifier.

3.4 LS-SVM Classification Process

3.4.1 LS-SVM Training phase

SVM and LS-SVM are employed to solve classification problems using machine learning techniques. Both are non-probabilistic binary classifiers and construct a hyperplane that separates the two classes. SVM advancement is called LS-SVM, which is utilized to solve linear equations and find a learning model for categorization. Despite this, SVM was used to solve quadratic problems. An LS-SVM classifier is more affordable when it is matched with an SVM classifier. Compared with SVM, LS-SVM is naturally simple. Only a few parameters must be tuned in LS-SVM. Additionally, LS-SVM can quickly perform linear and nonlinear multivariate classifications when correlated with multivariate classifiers such as NB and NN. To classify the breast abnormalities, the proposed system uses an RBF exponential kernel as follows:

$$\text{RBF Kernel: } K(x, x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right) \quad (1)$$

where $\alpha$is represents the spreading parameter. The parameter $\ddot{I}f$, which affects data transformation, can be tuned such that the SVM performance can be improved. RBF Kernel is something like a low-band pass filter, which selects smooth solutions. The proposed system uses a GA to obtain the optimal values of α. The optimized α can be used to efficiently calculate breast abnormalities using the optimal features. The process of the GA-optimized LS-SVM is shown in Figure-7. The figure shows the process of deriving the optimal parameters of the RBF kernel used in the LS-SVM classifier. The proposed system performs classifications by utilizing the training and testing datasets. Once the optimal features are selected by the GA-based feature-selection process, the RBF kernel parameters are also optimized by the GA process. In the LS-SVM training phase, the training dataset is used to train the SVM classifier with the optimal features and SVM parameters.

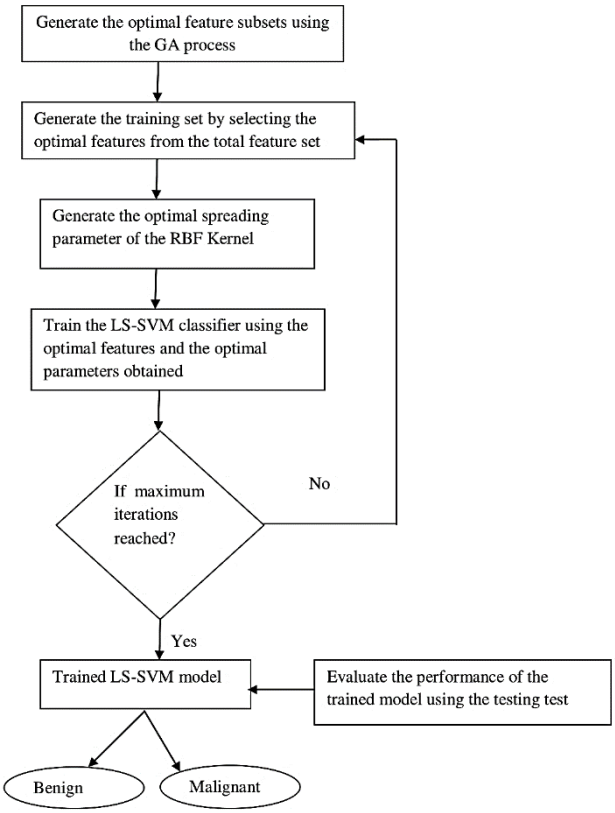

Figure-7. Proposed GA-optimized LS-SVM Working Model.

3.4.2 LS-SVM Evaluation phase

In the trained SVM model, the optimally generated features are fed as inputs. Abnormal classes in the mammograms were detected, and then the LS-SVM classification phase was implemented. The trained SVM can now compare these features with the features of its entries produced in the training step, to correctly perform benign or malignant tumor classification.

4. EXPERIMENTAL SETUP

The MIAS dataset comprises 322 images, which a categorized as normal, microcalcification, and mass images. Among the 322 images, 207 normal and 115 abnormal images were obtained. Among the 115 abnormal images, 25 were microcalcification images, and 90 mass images. Of the 115 abnormal images, 64 were benign and 51 malignant. We took all 115 abnormal images and 70 randomly chosen normal images for the experimental task. The performance was evaluated using various performance evaluation metrics, namely, true positive (TP), true negative (TN), false positive (FP), false negative (FN), sensitivity, specificity, precision, false-positive value, and classification accuracy. Tables 1 and 2 present the confusion matrix and various measures applied for performance evaluation. The basic evaluation measures of the confusion matrix are shown in Figure-8.

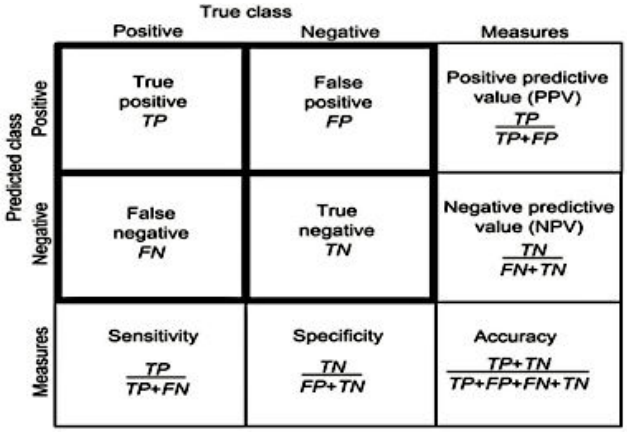

Figure-8. Basic evaluation measures of the confusion matrix.

Table-1. Benign and malignant classification of breast cancer.

Shape and boundary Features of breast masses	Classification based on the severity
Round or Oval	Benign (non-cancerous)
Smooth, Circumscribed	Benign (non-cancerous)
Irregular	Malignant (cancerous)
Speculated, Blurred, and Rough	Malignant (cancerous)

Table-2. Confusion matrix and measures after ten runs for benign/malignant classification.

Methods of comparison	Actual	Number of cases	Test outcome-Predicted	
			Malignant	Benign
Linear kernel	Malignant	510	490(TP)	40(FN)
	Benign	640	20(FP)	600(TN)
RBF kernel	Malignant	510	500 (TP)	20(FN)
	Benign	640	10(FP)	620 (TN)
GA based RBF kernel	Malignant	510	505(TP)	5(FN)
	Benign	640	5(FP)	635(TN)

Table-3. Performance of LS-SVM with shearlet attributes for normal/abnormal case.

Kernels / Measures	TN	TP	FN	FP	Recall (%)	Specificity (%)	Precision (%)	False Positive Value (%)	Classification Accuracy (%)
Linear	83	41	29	32	58.5	72.17	56.16	27.83	67.32
Quadratic	73	27	43	42	38.57	63.47	39.13	36.53	54.03
Polynomial	83	41	29	32	58.5	72.17	56.16	27.83	67.32
MLP	52	30	40	63	42.82	45.21	32.25	54.79	44.42
RBF	107	66	4	8	94.28	93.04	89.19	6.95	93.04

Table-2 presents the comprehensive performance of the LS-SVM classifier with various kernels for the classification of normal/abnormal images using shearlet attributes. As shown in Table-3, the LS-SVM classifier with RBF kernel yields 93.04% classification accuracy for normal/abnormal case classification, and the accuracy is superior to those achieved with the other kernels. Among the various kernels, the RBF kernel is the most popular one and it is mainly used to map data nonlinearly into a higher dimensional space which leads to better classification accuracy. Table-2 presents the result of the GA-based LS-SVM classifier with different kernels for image classification as either normal or abnormal. From the table, it is clear that the GA-based LS-SVM classifier with RBF kernel achieves a classification accuracy of 94.59% for normal/abnormal image classification. The GA-based LS-SVM classifier with RBF kernel yields improved results compared with the LS-SVM classifier with RBF kernel (see Table-1) and all other kernels for normal/abnormal case classification. Table-3 shows the execution of the LS-SVM classifiers with different kernels for the classification of abnormal images as benign or malignant using statistical moments. By utilizing the estimations in Table-2, the LS-SVM classifier with RBF kernel outputs a classification accuracy of 97.39% for benign/malignant case classification, and the accuracy is higher than those achieved for all other kernels for benign/malignant case classification using statistical texture features. Furthermore, Tables 2 and 4 show that an LS-SVM classifier with linear and polynomial kernels, respectively, provides the same accuracy for both classification cases.

Table-4. Performance of LS-SVM with Shearlet attributes and GA for normal/abnormal classification.

Kernels/ Measures	TN	TP	FN	FP	Recall (%)	Specificity (%)	Precision (%)	False Positive Value (%)	Classification Accuracy (%)
Linear	85	50	20	30	71.42	73.91	62.5	26.09	72.72
Quadratic	78	36	34	37	51.42	67.82	49.32	32.17	61.62
Polynomial	85	50	20	30	71.42	73.91	62.5	26.09	72.72
MLP	78	36	34	37	51.42	67.82	49.32	32.17	61.62
RBF	108	67	3	7	95.71	93.91	90.54	6.09	94.59

Table-5. Performance of LS-SVM with Shearlet attributes for benign/malignant case.

Kernels/ Measures	TN	TP	FN	FP	Recall (%)	Specificity (%)	Precision (%)	False Positive Value (%)	Classification Accuracy (%)
Linear	38	51	13	13	79.68	74.5	79.68	25.49	77.39
Quadratic	34	37	27	17	57.81	66.66	68.51	33.34	61.7
Polynomial	38	51	13	13	79.68	74.5	79.68	25.49	77.39
MLP	23	32	32	28	50	45.1	53.33	54.9	47.82
RBF	51	61	3	0	95.32	100	100	0	97.39

Table-6. Performance of LS-SVM with Shearlet attributes and GA based benign/malignant classification.

Kernels/ Measures	TN	TP	FN	FP	Recall (%)	Specificity (%)	Precision (%)	False Positive Value (%)	Classification Accuracy (%)
Linear	38	57	7	13	89.06	74.5	81.42	25.5	82.61
Quadratic	34	35	29	17	54.68	66.66	67.3	33.34	60
Polynomial	38	57	7	13	89.06	74.5	81.42	25.5	82.61
MLP	27	40	24	24	62.5	52.94	62.5	47.06	58.26
RBF	50	63	1	1	98.44	98.04	98.44	1.96	98.26

Table-7. Overview of existing research for mammogram image classification using SVM classifiers.

S. No.	Author Name & Year	Techniques Used	Classification Accuracy
1	Moayedi F *et al*. [13]	Contourlet Features, GA, SVM classifier with RBF kernel	Normal / Abnormal: 82.31%
			Benign / Malignant: 85.10%
2	Muhammad Talha [14]	Discrete Cosine Transform, GP, SVM	Normal / Abnormal: 93.39%
3	Alfonso [34]	Dynamic-programming-based method and a constrained region-growing method, SVM	Sensitivity is0.6 and Specificity is 0.8
4	T.S. Subashini, *et al* [35]	Gray level thresholding and connected component labeling, SVM RBF Kernel	The algorithm has a classifier accuracy of 95.44%
5	BASAVARAJ *et al* [36]	Gaussian Filters, Local Binary Patterns features, RBF Kernel	Classification Accuracy of 89.33%
6	W. Borges *et al* [37]	Shape, texture using geostatic function, SVM	Classification Accuracy of 80%
7	Wang *et al.* [38]	Curvilinear, GLCM, Gabor, Multi-resolution statistical features, Structured SVM	Classification Accuracy of 91.4%
8	Y.Ireaneus Anna Rejani, *et al.* [39]	Shape Feature-based DWT, SVM Classifier	Classification Accuracy 88.75%
9	Ioan B. *et al*. 2011 [40]	Gabor wavelets and directional features SVM	Classification Accuracy 84.37%
10	Proposed work	DST, GA, LS-SVM classifier with RBF kernel	Normal / Abnormal:94.59%
			Benign / Malignant: 98.26%

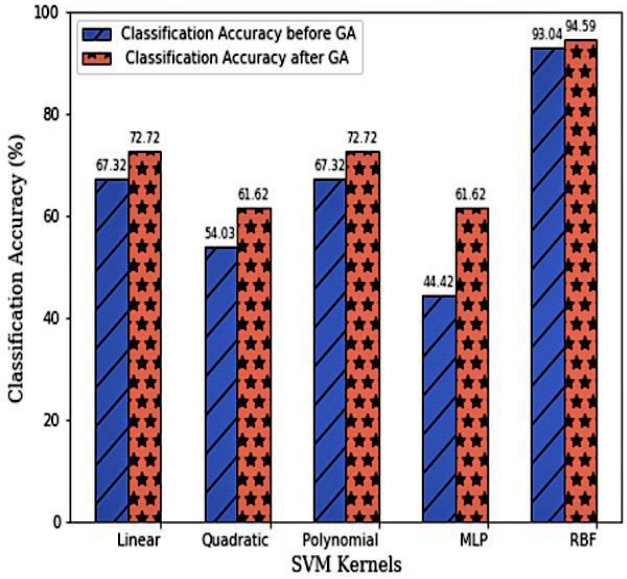

Figure-9. Accuracies obtained for normal and abnormal case classification with and without GA, respectively.

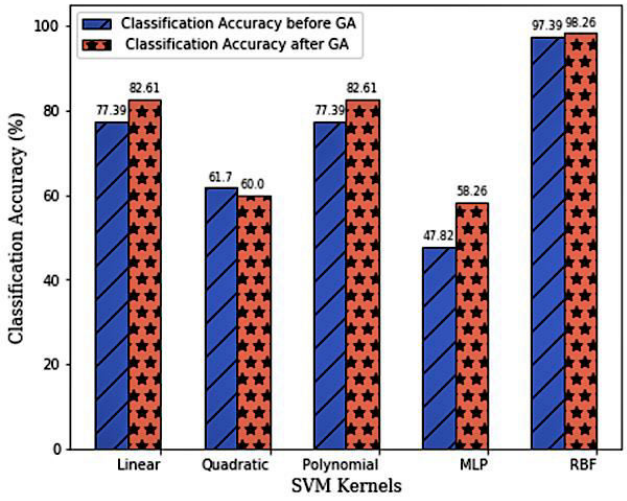

Figure-10. Accuracies obtained for benign/malignant type classification with and without GA, respectively.

Table-5 presents the execution of the GA-based LS-SVM classifier with different kernels for the classification of abnormal images as benign or malignant type. From Table-4, it is clear that the GA-based LS-SVM classifier with RBF kernel achieves an accuracy of 98.26% for benign/malignant type classification. From Table-5, the GA-based LS-SVM classifier with RBF kernel yields higher classification accuracy compared with the LS-SVM classifier with RBF kernel (see Table-3) for benign/malignant cases using statistical shearlet features. Table-6 shows that the accuracy of the projected system is higher than that of the other existing strategies. Figure-9 shows the accuracies for normal/abnormal case classification without and with GA, respectively. Figure-10 shows the accuracies for benign/malignant case classification with and without GA, respectively. From Figures 9 and 10 the GA-based LS-SVM classifier achieves better classification accuracy for both cases.

5. CONCLUSIONS

A CAD model was developed to detect and classify breast cancer using LS-SVM, GA, and shearlet moments. In addition to GA, the proposed scheme utilizes an LS-SVM classifier with different kernels, such as linear, quadratic, polynomial, MLP, and RBF, and discrete shearlet transform, for breast lesion detection. This study was performed in two different steps. The first step was to find the abnormal condition. In the second step, the abnormal tumor was distinguished as either benign or malignant. The performance of the proposed approach was evaluated using shearlet texture features. Features were selected from the shearlet-based feature set by using GA. The experimental results clarified that among the various kernels, the RBF kernel provides better classification accuracy with GA-based features, and testing was performed using a tenfold cross-validation method. In conclusion, the GA-based LS-SVM classifier with RBF kernel is superior to the LS-SVM classifier with RBF kernel. In the future, the proposed algorithm can be used for high-dimensional datasets. Additionally, it can be implemented for the classification of abnormalities present in MRI and CT scan images. Moreover, Real time datasets also should be tested using this approach.

AUTHORS' CONTRIBUTIONS

Authors proposed CAD architecture for the classification of breast tissues as either benign or malignant, yielding a classification accuracy of 98.26%.

REFERENCES

[1] F. Bray, J.-S. Ren, E. Masuyer and J. Ferlay. 29013. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. International journal of cancer. 132(5): 1133-1145.

[2] C. DeSantis, R. Siegel, and A. Jemal. 2015. Breast cancer facts & figures 2015-2016. Am cancer soc. Vol. 44.

[3] M. Dong, X. Lu, Y. Ma, Y. Guo, Y. Ma and K. Wang. 2015. An efficient approach for automated mass segmentation and classification in mammograms. Journal of digital imaging. 28(5): 613-625.

[4] D. Kufe, R. Pollock, R. Weichselbaum, R. Bast and T. Gansler. 2013. Holland-frei cancer medicine. hamilton (on): Bc decker.

[5] S. Deepa and V. S. Bharathi. 2013. Textural feature extraction and classification of mammogram images using cccm and pnn. IOSR Journal of Computer Engineering (IOSR-JCE). 10(6): 07-13.

[6] S. Sharma and P. Khanna. 2015. Computer-aided diagnosis of malignant mammograms using zernike moments and svm. Journal of digital imaging. 28(1): 77-90.

[7] R. Ramani, N. S. Vanitha and S. Valarmathy. 2013. The pre-processing techniques for breast cancer detection in mammography images. International Journal of Image, Graphics and Signal Processing. 5(5): 47.

[8] C.-C. Jen and S.-S. Yu. 2015. Automatic detection of abnormal mammograms in mammographic images. Expert Systems with Applications. 42(6): 3048-3055.

[9] H. Wang, J.-B. Li, L. Wu, and H. Gao. 2013. Mammography visual enhancement in cad-based breast cancer diagnosis. Clinical imaging. 37(2): 273-282.

[10] Ibrahima sory keita, Ir. Pratap Nair, Haarindra Prasad, Sudhakara pandian, S. Deivasigamani. 2022. Classification of Benign and Malignant MRIs Images using SVM Classifier for Brain Tumor Detection. International Journal of Engineering Trends and Technology. 70(2): 221-228, ISSN:2231-5381/https://doi:10.14445/22315381/IJETT-V7012P226.

[11] K. Hu, X. Gao and F. Li. 2010. Detection of suspicious lesions by adaptive thresholding based on multiresolution analysis in mammograms. IEEE Transactions on Instrumentation and Measurement. 60(2): 462-472.

[12] J. Tang, R. M. Rangayyan, J. Xu, I. El Naqa and Y. Yang. 2009. Computer aided detection and diagnosis of breast cancer with mammography: recent advances. IEEE transactions on information technology in biomedicine. 13(2): 236-251.

[13] F. Moayedi, Z. Azimifar, R. Boostani and S. Katebi. 2010. Contourlet-based mammography mass classification using the svm family. Computers in biology and medicine. 40(4): 373-383.

[14] S. K. Uma Maheswaran, S. Deivasigamani, K. Joshi, D. Verma, S. Kumar Rajamani and D. S. ROSS. 2022. Computational Intelligence Approach to Improve the Classification Accuracy of Brain Tumor Detection. 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, pp. 1192-1196, doi: 10.1109/SMART55829.2022.10047792.

[15] N. Wu, J. Phang, J. Park, Y. Shen, Z. Huang, M. Zorin, S. Jastrzebski, T. F´evry, J. Katsnelson, E. Kim *et al*. 2019. Deep neural networks improve radiologists performance in breast cancer screening. IEEE transactions on medical imaging. 39(4): 1184-1194.

[16] S. A. Mohammed, S. Darrab, S. A. Noaman and G. Saake. 2020. Analysis of breast cancer detection using different machine learning techniques. in International Conference on Data Mining and Big Data. Springer. pp. 108-117.

[17] M. Amrane, S. Oukid, I. Gagaoua and T. Ensar˙I. 2018. Breast cancer classification using machine learning. in 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT). IEEE. pp. 1-4.

[18] K., G., V., M., R., S. P., S., D. and Ang, C. K. 2023. A Novel SVM and K-NN Classifier Based Machine Learning Technique for Epileptic Seizure Detection. International Journal of Online and Biomedical Engineering (iJOE). 19, 07, pp. 99-124. DOI: https://doi.org/10.3991/ijoe.v19i07.37881.

[19] S. Prabhakar and H. Rajaguru. 2017. Performance analysis of breast cancer classification with softmax discriminant classifier and linear discriminant analysis. in International Conference on Biomedical and Health Informatics. Springer. pp. 197-201.

[20] A. Ivaturi, A. Singh, B. Gunanvitha and K. Chethan 2020. Soft classification techniques for breast cancer detection and classification. in 2020 International Conference on Intelligent Engineering and Management (ICIEM). IEEE. pp. 437-442.

[21] R. Rouhi, M. Jafari, S. Kasaei and P. Keshavarzian. 2015. Benign and malignant breast tumors classification based on region growing and cnn segmentation. Expert Systems with Applications. 42(3): 990-1002.

[22] V. Bhateja, S. Urooj and M. Misra. 2014. Technical advancements to mobile mammography using nonlinear polynomial filters and ieee 21451-1 ncap information model. IEEE Sensors Journal. 15(5): 2559-2566.

[23] V. Bhateja, H. Patel, A. Krishn, A. Sahu and A. Lay-Ekuakille. 2015. Multimodal medical image sensor fusion framework using cascade of wavelet and

contourlet transform domains. IEEE Sensors Journal. 15(12): 6783-6790.

[24] S. Sharma and P. Khanna. 2015. Computer-aided diagnosis of malignant mammograms using zernike moments and svm. Journal of digital imaging. 28(1): 77-90.

[25] M. Hussain, S. K. Wajid, A. Elzaart and M. Berbar. 2011. A comparison of svm kernel functions for breast cancer detection. in 2011 Eighth International Conference Computer Graphics, Imaging and Visualization. IEEE. pp. 145-150.

[26] I. Kitanovski, B. Jankulovski, I. Dimitrovski and S. Loskovska. 2011. Comparison of feature extraction algorithms for mammography images. in 2011 4th International Congress on Image and Signal Processing, vol. 2. IEEE, pp. 888-892.

[27] M. E. Elmanna and Y. M. Kadah. 2015. Implementation of practical computer aided diagnosis system for classification of masses in digital mammograms. in 2015 International Conference on Computing, Control, Networking, Electronics and Embedded Systems Engineering (ICCNEEE). IEEE. pp. 336-341.

[28] N. Gedik and A. Atasoy. 2013. A computer-aided diagnosis system for breast cancer detection by using a curvelet transform. Turkish Journal of Electrical Engineering & Computer Sciences. 21(4): 1002-1014.

[29] M. Dong, X. Lu, Y. Ma, Y. Guo, Y. Ma and K. Wang. 2015. An efficient approach for automated mass segmentation and classification in mammograms. Journal of digital imaging. 28(5): 613-625.

[30] D. R. Ericeira, A. C. Silva, A. C. De Paiva and M. Gattass. 2013. Detection of masses based on symmetric regions of digital bilateral mammograms using spatial description with variogram and cross-variogram functions. Computers in biology and medicine. 43(8): 987-999.

[31] Z. Chen, H. Strange, A. Oliver, E. R. Denton, C. Boggis and R. Zwiggelaar. 2014. Topological modeling and classification of mammographic microcalcification clusters. IEEE transactions on biomedical engineering. 62(4): 1203-1214.

[32] A. K. Mohanty, M. Senapati, S. Beberta and S. K. Lenka. 2013. Retracted article: Mass classification method in mammograms using correlated association rule mining. Neural Computing and Applications. 23(2): 273-281.

[33] D. H. Kim, J. Y. Choi and Y. M. Ro. 2015. Region based stellate features combined with variable selection using adaboost learning in mammographic computer-aided detection. Computers in biology and medicine. 63: 238-250.

[34] A. R. Dom´ınguez and A. K. Nandi. 2009. Toward breast cancer diagnosis based on automated segmentation of masses in mammograms. Pattern Recognition. 42(6): 1138-1148.

[35] T. Subashini, V. Ramalingam and S. Palanivel. 2010. Automated assessment of breast tissue density in digital mammograms. Computer Vision and Image Understanding. 114(1): 33-43.

[36] B. Hiremath and S. Prasannakumar. 2015. Automated evaluation of breast cancer detection using svm classifier. International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR), 5, 20(1): 11.

[37] W. B. Sampaio, E. M. Diniz, A. C. Silva, A. C. De Paiva and M. Gattass. 2011. Detection of masses in mammogram images using cnn, geostatistic functions and svm. Computers in Biology and Medicine. 41(8): 653-664.

[38] D. Wang, L. Shi and P. A. Heng. 2009. Automatic detection of breast cancers in mammograms using structured support vector machines.Neurocomputing. 72(13-15): 3296-3302.

[39] Y. Rejani and S. T. Selvi. 2009. Early detection of breast cancer using svm classifier technique. arXiv preprint arXiv: 0912.2314.

[40] I. Buciu and A. Gacsadi. 2011. Directional features for automatic tumor classification of mammogram images. Biomedical Signal Processing and Control. 6(4): 370-378.