

Mammogram Classification Using Genetic Algorithm based Feature Selection

Neeti Arora¹, Gaurav Aggarwal²

¹Jagannath University, NCR, Haryana, India

²Jagannath University, NCR, Haryana, India
(E-mail: neetiarorapl@gmail.com)

Abstract—The breast cancer is one of the most dangerous and well-known cancer in the world especially for women. The digital mammogram is the most suitable tool for the detection of breast cancer. In this research work, Genetic Algorithm–Support Vector Machine (GA-SVM) method introduced for detecting the breast cancer. Initially, the input mammogram images were taken from the dataset: mammographic Image Analysis Society (MIAS). Then, Fast Local Laplacian (FLL) filter was used to find the area of edges in mammogram images. The pre-processed mammogram images were used for segmentation by employing Adaptively Regularized Kernel based on Fuzzy C-Means (ARKFCM) that is a flexible high-level machine learning technique to localize the object in complex template. A hybrid feature extraction: Histogram of Oriented Gradient (HOG), energy and homogeneity was performed on the segmented image in order to improve the feature subsets. In this work, the feature selection process was performed by using Kernel Nearest Neighbor (KNN) based GA in order to acquire the best feature values. These values were given as the input for Support Vector Machine (SVM) classifier. The SVM technique was used for classifying the normal and abnormal regions in the mammogram images. In the experimental results, the GA-SVM method distinguishes the normal and abnormality of the breast cancer by means of sensitivity, specificity, accuracy, False-Measure (F-measure). The experimental outcome shows that the GA-SVM methodology improved the accuracy in breast cancer detection up to 2-4% compared to the existing methods: Neural Network (NN) and Naïve Bayes.

Keywords—Breast Cancer Detection, Feature selection, Genetic algorithm, Histogram of Oriented Gradient and Support Vector Machine.

I. INTRODUCTION

Nowadays, Cancer is the most dangerous and leading cause of death in the world. Several types of cancers occur in the different organs of a human body, such as blood cancer, lung cancer, brain cancer, breast cancer and so on. The breast cancer mostly affects women, which is the second most common cancer. Breast cancer occurs when the cells of breast become abnormal and uncontrollably distributed or divided. That abnormal cell forms a huge lump of tissues that consequently become tumor [1]. Hence, a special attention is required to overcome this breast cancer disease. The Mammography can be

used to detect cancer cell in the breast, which have a specific type of mammogram breast image. [2]. Imaging methodologies are playing the most significant role in early breast cancer detection and mammography approach has been mostly accepted as diagnostic and screening mammography that reduces the number of breast cancer deaths among the women aged from forty to seventy years [3]. Most of the conventional systems have some issues because of poor image quality. Some traditional methods did not perform well in case of noises due to low radiation and poor quality images. Breast cancer detection is one of the major challenge in medical fields [4]. The breast cancer can be identified by performing a biopsy, where abnormal tissue is removed and studied under a microscope. The histopathologist who detects abnormal cell, do the diagnosis based on his/her qualification. However, if the histopathologist is not well trained, it may lead to a wrong diagnosis [5]. The traditional mammogram enhancement techniques use the transform domain filtering; it possibly provides some artifacts in images [6].

The micro-calcification clusters and masses are the main early signs of the breast cancer. It is very difficult to differentiate abnormalities from normal breast tissues due to their precise appearance and uneven margins [7]. An automatic classification of the breast masses for the diagnosis of breast cancer in the digital Mammograms are done by utilizing the NN technique. The NN technique classifies the segmented Region of Interest (RoI) into normal and abnormal RoI. However, the learning time is high, which is the major drawback of the NN technique [8]. Manual recognition technique takes more time for segmenting the cancer and non-cancer regions and it is applicable only for a limited time mammogram image [9]. Hence, an automatic breast cancer detection system has created an attention among the research community. In automated breast cancer system, segmentation and classification are the two main key steps [10]. To conquer this problem, the FLL filter is used for preprocessing, which is performed to identify the area of edges in the Mammogram images. The pre-processed Mammogram images are used for the segmentation process by using the ARKFCM technique. The ARKFCM technique provides robustness to the clustering parameters, which helps to reduce the computational complexity. Moreover, a hybrid feature extraction such as HOG, energy and homogeneity are used for extracting the feature values. This feature extraction helps to reflect the depth and smoothness of mammogram image texture architecture. In this research work, the SVM technique is used for classifying normal and abnormal breast cancer regions in the mammogram images.

This research paper is composed as follows; Section-2 presents an extensive survey of recent papers on BCD techniques. Section-3 briefly described an effective GA-SVM-BCD methodology. Section-4 shows the comparative experimental result for the existing and GA-SVM-BCD method. The conclusion is made in Section -5.

II. LITERATURE SURVEY

The researchers implemented several methods on breast cancer, segmentation, feature extraction and classification techniques. This section presented a brief evaluation of some significant contributions to this field.

Kumar *et al.* [11] implemented a hybrid hierarchical technique used for classifying the density of breast cancer using Digitalized Mammogram Images (DMIs). Five types of classification techniques such as Principal Component Analysis- Kernel Nearest Neighbor (PCA-KNN), PCA-Artificial NN (PCA-ANN), PCA-SVM, PCA Probabilistic Neural Network (PCA-PNN) and PCA Neuro-Fuzzy Classifier (PCA-NFC) used in this paper. The PCA-NFC outperformed the conventional techniques in terms of overall classification accuracy. However, the performance of hybrid hierarchical technique highly degraded due to a number of samples in the training set, which is main drawback of this method.

Sakri *et al.* [12] proposed the Particle Swarm Optimization (PSO) feature selection for breast cancer recurrence prediction. The main aim of this research work was to compare the accuracy of some conventional data mining algorithms that predict breast cancer recurrence. The PSO algorithm used feature selection in 3-renowned classifiers such as Bayes, KNN and Fast Decision Tree Learner (FDTL). The naïve Bayes produced better accuracy of 70% and 81.3% without and with PSO algorithm compared to other two classifiers. The PSO algorithm is not suitable for combinatorial problems.

Abubacker *et al.* [13] illustrated Genetic Association Rule Miner (GARM) with the learning capacity of NN technique used for classification. In this research work, the feature selection was carried out by using a multivariate filter to eliminate the inappropriate feature. The multivariate filter approach helped to maximize the classification efficiency. The experimental results were carried out on MIAS dataset for exhibiting the robustness of the GARM technique. It was very complex to find the projection of the space in the proposed architecture, which is the major limitation of the GARM technique.

Zhang *et al.* [14] implemented integrating feature selection and feature extraction techniques with deep learning to predict the clinical outcome of breast cancer. This literature work has established an additional classifier with same classifier learning strategy (Ada-boost algorithm) in order to perform as a baseline to the proposed method where the training inputs only differs. An Area Under the receiver operating characteristic Curve (AUC) index, Matthews Correctional coefficient (MCC) index was tested based on many independent breast cancer database, and compared with the specific gene sign based algorithms including baseline technique. In the NN classifier, the constructed model is not easy to analyze, which is a drawback of the NN classifier.

Bardou *et al.* [15] proposed the comparison of two machine-learning techniques for the automatic classification of breast cancer histology images into malignant and benign sub-classes. The first technique was based on the extraction of a set of handcrafted features encoded by 2-coding model and the data were trained by using SVM. While the second technique was designed based on conventional NN. This paper achieved an accuracy of 96.15% and 98.33% for binary classification, 83.31% and 88.23% for the multi-class classification. However, this literature method has required more run time.

To overcome the above-mentioned limitations, the feature selection based BCD system is implemented for improving the performance of the BCD system.

III. GENETIC ALGORITHM BASED FEATURE SELECTION USING BREAST CANCER DETECTION METHODOLOGY

The dimensions of the medical data include many features and each feature contains different kinds of values. The data quality related problems are outliers, presence of noise, duplicate data and so on [16]. Preprocessing is applied to the raw data for further analysis. The feature selection technique is used to solve some of the issue related to data quality [17].

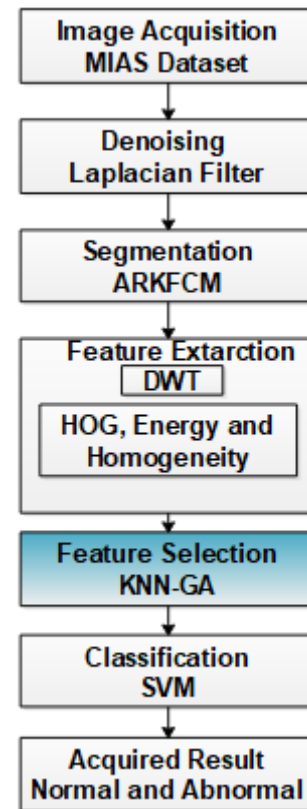


Fig. 1. Block diagram of the GA-SVM methodology

The feature selection is one of the significant step use in breast cancer detection for increasing the classification accuracy. It allows only a limited number of input features to the classifier for creating an appropriate predictive and low computational intensive model. The GA-SVM methodology has

seven steps image acquisition, image pre-processing, segmentation process, feature extraction, hybrid feature extraction, feature selection and classification, which is shown in the Fig. 1. The brief description about the GA-SVM methodology presented in the following section.

A. Image Acquisition and Pre-processing

In this research work, the MIAS database is used to compute the efficiency of the GA-SVM system. The MIAS database is a benchmark that is openly available for the research. The Laplacian filter is used as a pre-processing technique for the acquisition of the mammogram image. The Laplacian filter effectively removes the noise compared to other filters [18]. The Laplacian filter contains a smoothing operator, which is used to transform the noisy image into noiseless images for the real-mammogram images.

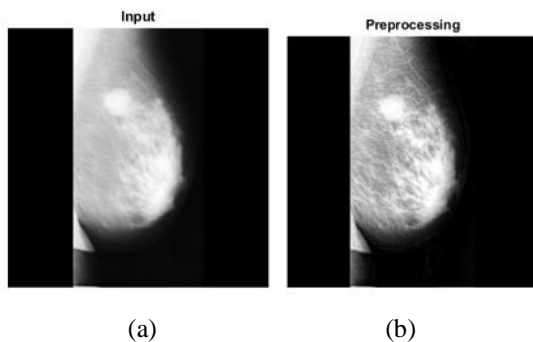


Fig. 2. a) Original mammogram image b) Noise removed mammogram image.

The Laplacian filter plays an important role in edge detection of the mammogram images. The figure 2 shows the original image and the noiseless mammogram image. The common formula of the Laplacian filter is represented in the Eq. (1).

$$G_0 = G_\sigma(u, v) \times e^{-\frac{u^2 + v^2}{2\sigma^2}} \quad (1)$$

Here, the u represents the horizontal axis and v represents the vertical axis, and σ represents the standard deviation.

B. Adaptively Regularized Kernel based on Fuzzy C-Means technique

The pre-processed mammogram images are used for the segmentation process. The ARKFCM technique is one of the efficient segmentation approach, which is undertaken for segmenting the mammogram images with and without cancer cell regions [19]. This ARKFCM technique uses the heterogeneity of the grayscales in the neighborhood for computing the Local Contextual Information (LCI). In the ARKFCM technique, the standard Euclidean Distance change with correlation function. The Fig. 3 shows the segmented image.

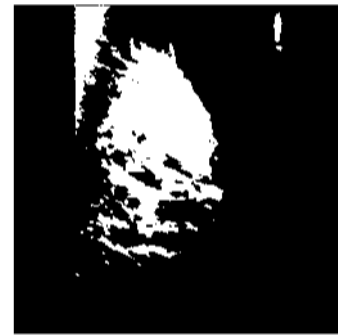


Fig. 3. Segmented breast image

C. Feature Extraction using Discrete Wavelet Transform

The segmented mammogram images are used to compute the wavelet coefficient using the DWT, which is a square shaped function. The DWT technique has properties like better compression energy and perfect reconstruction with short support filters, low-computation and no redundancy [20]. The DWT follows the fuzzy de-noising procedure, which provides shift capable sub-bands and better directional selectivity with less redundancy. In the multi-resolution process, the real texture of the mammogram image is computed by zooming in and out process. Usually, the mammogram images are decomposed into a number of sub-image at different resolution levels for preserving the low and high-frequency information. This property helps to extract the information of the texture from the mammogram images. The square integrable function $f(u)$ and wavelet transform is represented as the inner product f and original valued function $\psi(x)$. The wavelet function is given in the Eq. (2).

$$w[f(s, \tau)] = (f, \psi_{s,t}^k) = \int_{-\infty}^{\infty} f(u) \psi_{s,t}^k(x) \quad (2)$$

Where, $\varphi_{s,t}^k(u) = \left(\frac{1}{\sqrt{s}} \varphi_{s,t}^k(u-\tau)\right) / s$ is denoted as wave family, $s \in z, \tau$ and $k \in \{h, x, d\}$ are denoted as scale, translation and orientation parameters respectively. The orientation parameters h, x and d are denoted as vertical, horizontal and diagonal direction respectively. The dyadic wavelet decomposition is achieved, during $s = 2^j$ and $\tau = 2^j n, j, n \in z$. The dyadic wavelet decomposition is a scalable sample of DWT; it follows a geometric sequence of ratio two. The following wavelet decomposition employs dyadic wavelets that is implemented using perfect reconstruction filter banks. Using this wavelet function $\varphi(u)$ and the scaling function $\phi(u)$, the scaling and wavelet families are created using the Eq. (3) and (4). These are orthogonal basis of the sub-spaces and related to resolution on 2^j . The wavelet atoms are defined by scaling and three mother atoms ψ^h, ψ^x , and ψ^d . This mother atoms is computed as the tensor products of 1- dimensional $\psi(x)$ and $\phi(x)$, that is

denoted in the Eq. (5) and (6).

$$\psi_{j,n}^k(u) = \frac{1}{\sqrt{2^j}} \psi^k\left(\frac{u-2^j \cdot n}{2^j}\right) \quad (3)$$

$$\phi_{j,n}^k(u) = \frac{1}{\sqrt{2^j}} \phi^k\left(\frac{u-2^j \cdot n}{2^j}\right) \quad (4)$$

$$\phi(u) = \phi(u_1)\phi(u_2), \psi^h(u) = \psi(u)\psi(u) \quad (5)$$

$$\psi^x(u) = \phi(u_1)\psi(u_2), \psi^d(u) = \psi(u_1)\phi(u_2) \quad (6)$$

Two-dimensional DWT illustrate by the combination of down samplers and digital filter banks. The digital filter bank made up of a low-pass filter and high-pass filter. The number of bank is grouped as per the desired functions in the wavelet configuration structure. Next, the rows and columns of the mammogram images are separately undergone through the 1-dimensional wavelet transform in order to create 2-dimensional wavelet coefficient. The real mammogram images $A_{2^{j+1}}f$ at resolution 2^{j+1} decompose into 4-subband images in the frequency domain. Three sub-band images such as $D_{2^j}^h f, D_{2^j}^v f, D_{2^j}^d f$ are the mammogram images at resolution 2^{j+1} in vertical, horizontal and diagonal between the 4-subband images. The fourth image is an approximation image, $A_{2^j}f$ detected coarse resolution. So, the entire mammogram breast image $A_{2^{j+1}}f$ is presented in the Eq. (7).

$$A_{2^{j+1}}f = D_{2^j}^h f + D_{2^j}^v f + D_{2^j}^d f + A_{2^j}f \quad (7)$$

The decomposed mammogram sub-images are the 2-dimensional orthogonal wavelet. The results of the wavelet decomposition of a mammogram image is resultant into 4-orthogonal sub-bands such as Low-Low band, Low-High band, High-Low band and High-High band, which is represented as $D_{2^j}^h f, D_{2^j}^v f, D_{2^j}^d f$, and $A_{2^j}f$ respectively.

1) Hybrid Feature Extraction

In GA-SVM methodology, the hybrid feature extraction performed based on the transformed mammogram image. The feature extraction is the action of mapping the image from image space to feature space. A high-level feature named as Gray Level Co-occurrence Matrix (GLCM) is used for extracting the feature of the mammogram image in this section. The GLCM is the well-known texture analysis technique, which computes the image features related to second order statistics. This research work considered 2-effective GLCM texture features: energy and homogeneity. Additionally, the HOG descriptor is applied to the medical image processing and computer version is used for extracting the optimal feature values.

D. Feature selection using KNN based Genetic Algorithm

The feature selection selects the best features from the hybrid feature extraction [21]. The classification involves consideration of database before transferring the data to a classifier. It is recommended to consider only an important feature for selection. Hence, it is useful for choosing the significant and relevant features in this breast cancer detection.

Furthermore, the feature selection is used during the classification to find the important feature that decrease the workload of the classifier and enhances the classification accuracy. In this research work, KNN based GA significantly reduces the redundancy with-in the input voxels and also determines the maximum relevance between output and input voxels. In initial stage, generate the initial population, which is set as the subset for input voxels. Next, the fitness function is computed for input voxel subsets utilizing the KNN distance algorithm that increases the mutual-information between the voxels. Finally, crossover and mutation operators are used to find the most active voxels by decreasing the redundancy based on the fitness function. The GA selects the subset of features as the chromosomes and every chromosome is sent to the SVM for computing fitness value. The SVM classifier employs each chromosome as a mask for capturing the features. The SVM classifier defines a fitness value of each and every chromosome and GA utilizes these fitness values for the chromosome computation process. At the final stage, the GA finds an optimal subset of the feature.

E. Classification using Support Vector Machine

The suspicious portions are classified by using SVM classifier based on selected features [22]. The SVM is the type of classifier that performs well in solving the two-class problem. The generalized ability is defined by comparing the model complexity. The general formula used for the linear discriminant function is symbolized as $mn+b=0$. An optimum hyperplane is needed to separate the samples without noise and also exploit the gap between two groups, which is satisfied by implementing the below Eq. (8).

$$pi[w \cdot u + b] - 1 \geq 0, i = 1, 2, 3, \dots, N \quad (8)$$

The $\|w\|^2$, diminished in the above formula, because this optimization problem solved by the saddle point of Lagrange function with Lagrange multipliers. An ideal discriminant function is mathematically specified in Eq. (9),

$$f(u) = \text{sng}\{w \cdot u\} + b \cdot \text{sgn}\left\{\sum_{i=1}^N a_i \cdot pi(u_i \cdot u) + b\right\} \quad (9)$$

Then, interchange the interior product by a kernel function $K(u, u')$ in Eq. (10), to solve the huge computational complexity by high dimension. The linear separability of estimated samples is enhanced and discriminant function is rewritten as follows,

$$f(u) = \text{sng}\left\{\sum_{i=1}^N a_i \cdot pi \cdot k(u, u_i) + b\right\} \quad (10)$$

IV. RESULT AND DISCUSSION

The Matlab R2018a was used on Personal Computer (PC) with 3.2GHz Intel Core-5 processor. The experimental results briefly described the 3-phase such as classification evaluation, without feature selection, with feature selection of the proposed method. The performance of the GA-SVM method was implemented in terms of specificity, accuracy, sensitivity and F-measure.

A. Performance Evaluation

The performance measurement is defined as a relationship between the input and output variables of the system by using the suitable performance metrics like specificity and sensitivity. The general formula of specificity and sensitivity for breast cancer detection are given in the Eq. (11) and (12),

$$\text{Sensitivity} = \frac{TP}{TP + TN} \times 100 \quad (11)$$

$$\text{Specificity} = \frac{TN}{TN + TP} \times 100 \quad (12)$$

An accuracy and F-measure are the appropriate computation metrics for finding the effectiveness of the normality and abnormality in BCD. The performance measures are statistical variability and random errors. The general formula of accuracy and F-measure are given in the Eq. (13) and (14).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (13)$$

$$F - \text{measure} = \frac{2TP}{2TP + FP + FN} \times 100 \quad (14)$$

Here, TP- True Positive, TN- True Negative, FP-False Positive and FN-False Negative

B. Experimental results for with and without feature selection method

In this experimental research, the performance of the existing methods and the proposed method: GA-SVM evaluated by using MIAS database. The performance evaluation of the existing and GA-SVM methods are established for normal and abnormal region classification. In this phase, three prediction models such as Naïve Bayes, neural network and SVM classifier were used in the training dataset. These 3-models were constructed by utilizing all the features contained in the MIAS database without any process of the feature selection. Fig. 6 shows the accuracy, specificity, sensitivity and F-measure of the prediction models for the Naïve Bayes, neural network and SVM classifier.

TABLE I. PERFORMANCE EVALUATION OF THE EXISTING AND GA-SVM METHOD FOR NORMAL AND ABNORMAL WITHOUT FEATURE SELECTION

Methodology	Accuracy %	Specificity %	Sensitivity %	F-measure %
Naïve Bayes	98.70	99.80	97.60	98.75
Neural Network	50.60	48.20	53.00	45.089
SVM	97.50	99.60	95.40	97.60

TABLE II. PERFORMANCE EVALUATION OF THE EXISTING AND GA-SVM METHOD FOR NORMAL AND ABNORMAL WITH FEATURE SELECTION

Methodology	Accuracy %	Specificity %	Sensitivity %	F-measure %
Naïve Bayes	98.30	98.00	98.60	98.28
Neural Network	49.20	49.80	48.60	41.98
SVM	99.20	100	98.40	99.24

The Fig. 4 and 5 shows the performance of the prediction models without and with feature selection methods. The performance of the proposed method evaluated by using different performance metrics like accuracy, specificity, sensitivity and F-measure. The result from the existing and GA-SVM method are reported based on four metrics as shown

in table 1 and 2. All the classifier out-performed their counterparts without feature selection in terms of accuracy level. The proposed GA-SVM method achieved better accuracy compared to the existing method, due to an efficient feature selection approach. By using a SVM classifier in BCD, the proposed method achieved 99.20% of accuracy.

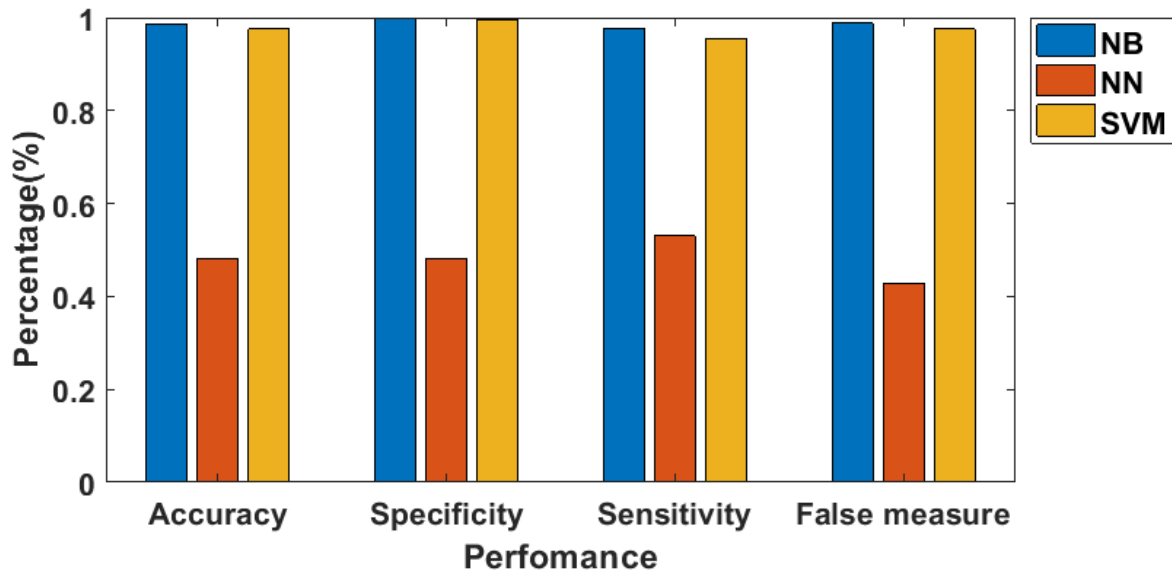


Fig. 4. Performance of the prediction models without feature selection

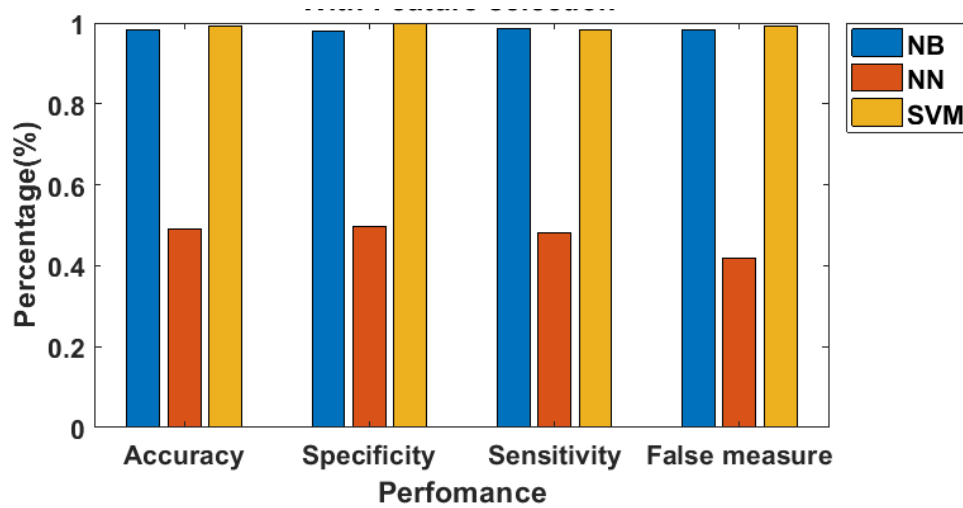


Fig. 5. Performance of the prediction models with feature selection

1) Comparative analysis

The table 3 shows the comparative analysis of existing and GA-SVM method. In the table 3, the accuracy value of proposed GA-SVM method is compared with the accuracy of the existing methods such as neural network, Radial Basis Function Neural Network and Weighted Association Rule based Classifier. The table 3 clearly shows that the GA-SVM method achieved 99.20% of the accuracy compared to the existing methods.

GA-SVM	99.20%
--------	--------

TABLE III. COMPARATIVE ANALYSIS OF THE EXISTING AND GA-SVM METHOD

Classification Techniques	Accuracy %
Neural network [23]	95.98%
Radial Basis Function Neural Network [24]	93.98%
Weighted Association Rule based Classifier [25]	89.68%

V. CONCLUSION

The breast cancer detection is one of the most significant research field in the computer-aided health monitoring systems. This paper focused on investigating the effect of integrating the feature selection algorithm with classification algorithm for BCD. The objective of the experiment was to improve an appropriate feature selection and classification techniques for classifying normal and abnormal of breast cancer cell utilizing the MIAS database. In this work, the ARKFCM segmentation technique is used to segment the cancer and non-cancer portion. The hybrid feature extractions (HoG, homogeneity and energy) extracts the optimal feature values from the segmented region. This proposed GA-SVM method improved the

classification accuracy by using GA feature selection techniques to decrease the number of features. The obtained best feature values are classified by using the SVM technique. The GA-SVM method delivered better performance by means of accuracy, sensitivity, specificity and F-measure of around 2-4% enhancement than the existing methods. The descriptor level features are combined with an effective optimal algorithm for the further improvement of the feature selection in the future work.

REFERENCES

- [1] M. M. Mehdy, P. Y. Ng, E. F. Shair, N. I. Saleh, and C. Gomes, "Artificial neural networks in image processing for early detection of breast cancer," *COMPUT MATH METHOD M.*, vol. 2017.
- [2] M. A. Jaffar, "Hybrid Texture based Classification of Breast Mammograms using Adaboost Classifier," *International Journal of Advanced Computer Science and Applications*, vol. 8, pp. 321-327, May 2017.
- [3] X. Liu, and J. Tang, "Mass classification in mammograms using selected geometry and texture features, and a new SVM-based feature selection method," *IEEE SYST J.*, vol. 8, pp. 910-920, Sep. 2014.
- [4] A. A. Onitilo, J. M. Engel, H. Liang, R. V. Stankowski, D. A. Miskowiak, M. Broton, and S. A. Doi, "Mammography utilization: patient characteristics and breast cancer stage at diagnosis," *AM J ROENTGENOL.*, vol. 201, pp. 1057-1063, Nov. 2013
- [5] D. Bardou, K. Zhang, and S. M. Ahmad, "Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks," *IEEE ACCESS.*, vol. 6, pp. 24680-24693, May, 2018.
- [6] H. Deng, W. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, "Mammogram enhancement using intuitionistic fuzzy sets," *IEEE T BIO-MED ENG.*, vol. 64, pp. 1803-1814, Aug. 2017.
- [7] J. Dheebea, N. A. Singh, and S. T. Selvi, "Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach," *Journal of biomedical informatics.*, vol. 49, pp. 45-52, Jun. 2014.
- [8] H. Shah, "Automatic Classification of Breast Masses for Diagnosis of Breast Cancer in Digital Mammograms using Neural Network," *IJSTE - International Journal of Science Technology & Engineering.*, vol. 1, pp. 47-52, May 2015.
- [9] B. W. Hong, and B. S. Sohn, "Segmentation of regions of interest in mammograms in a topographic approach," *IEEE T INF TECHNOL B.*, vol. 14, pp. 129-139, Jan. 2010.
- [10] M. A. Berbar, "Hybrid methods for feature extraction for breast masses classification," *Egyptian informatics journal.*, vol. 19, pp. 63-73, Mar. 2018.
- [11] I. Kumar, H. S. Bhadauria, J. Virmani, and S. Thakur, "A hybrid hierarchical framework for classification of breast density using digitized film screen mammograms," *MULTIMED TOOLS APPL.*, vol. 76, pp. 18789-18813, Sep. 2017.
- [12] S. Sakri, N. A. Rashid, and Z. M. Zain, "Particle Swarm Optimization Feature Selection for Breast Cancer Recurrence Prediction," *IEEE Access*, vol. 6, pp. 29637-29647, Jun. 2018.
- [13] N. F. Abubacker, A. Azman, S. Doraisamy, and M. A. A. Murad, "An integrated method of associative classification and neuro-fuzzy approach for effective mammographic classification," *NEURAL COMPUT APPL.*, vol. 28, pp. 3967-3980, Dec. 2017.
- [14] D. Zhang, L. Zou, X. Zhou, and F. He, "Integrating Feature Selection and Feature Extraction Methods with Deep Learning to Predict Clinical Outcome of Breast Cancer," *JOURNAL OF L A TEX CLASS FILES.*, vol. 14, May 2018.
- [15] D. Bardou, K. Zhang, and S. M. Ahmad, "Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks," *IEEE ACCESS.*, vol. 6, pp. 24680-24693, 2018.
- [16] K. Kourou, T. P. Exarchos, K. P. Exarchos, V. K. Michalis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and structural biotechnology journal.*, vol. 13, pp.8-17, 2015.
- [17] T. Sridevi, and A. Murugan, "A novel feature selection method for effective breast cancer diagnosis and prognosis," *International Journal of Computer Applications.*, vol. 88, Jan. 2014.
- [18] G. S. Ranu, and J. S. Sidhu, "A Comparison of Multiple Filtering Methods for Edge Detection of Breast Cancer Cells," *Indian Journal of Science and Technology.*, vol. 9, Dec. 2017.
- [19] W. Zang, W. Zhang, W. Zhang, and X. Liu, "A kernel-based intuitionistic fuzzy C-Means clustering using a DNA genetic algorithm for magnetic resonance image segmentation," *Entropy*, vol. 19, pp. 578, Oct. 2017.
- [20] S. Beura, B. Majhi, and R. Dash, "Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer," *NEUROCOMPUTING*, vol. 154, pp. 1-14. Apr. 2015.
- [21] M. A. Joadder, M. A. Nayeem, S. A. Shetu, F. R. Jamil, and A. Al Helal, "Feature selection for breast cancer detection from ultrasound images," in *Proceedings of International Conference on Informatics, Electronics & Vision (ICIEV)*, 2014, pp. 1-6.
- [22] V. Vishrutha, and M. Ravishankar, "Early detection and classification of breast cancer," in *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA)*, 2015, pp. 413-419.
- [23] M. M. Eltoukhy, I. Faye, and B. B. Samir, "A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation," *COMPUT BIOL MED.*, vol. 42, pp. 123-128, Jan. 2012.
- [24] M. Pratiwi, J. Harefa, and S. Nanda, "Mammograms classification using gray-level co-occurrence matrix and radial basis function neural network," *Procedia Computer Science*, vol. 59, pp. 83-91, Jan. 2015.
- [25] S. Dua, H. Singh, and H.W. Thompson, "Associative classification of mammograms using weighted rules," *Expert systems with applications*, vol. 36, pp. 9250-9259, Jul. 2009.



Ms. Neeti Arora is working as a Assistant Registrar, IT with Haryana Vishwakarma Skill University. She received her MCA degree from GJU, Hisar, Earlier she has worked with Advanced Institute of Technology and Management as Assistant Professor for 9 years. Presently Pursuing Ph.D from Jagannath University, NCR under the guidance of Dr. Gaurav Aggarwal , Head in Faculty Engineering & Technology domain Image Processing. Presented and published paper in National and International Conference.



Dr. Gaurav Aggarwal is Associate Professor and Head in Faculty Engineering & Tehnology with the Jagannath University, NCR, Haryana, India. He received his B.Tech in CSE from Maharishi Dayanand University, Rohtak in 2005 and M.Tech from VCE, Rohtak in 2008. He received his Ph.D. Degree from Faculty of Engineering & Technology, JNU, Jaipur in 2016. He has teaching experience of 13 years. He is

active member of R. G. Education Society, Rohtak. He has more than 35 Research papers in National and International Journals and Conferences. He has guided 20 Post graduate students and 1 doctorate student and currently 6 doctorate students are doing their work under

his supervision. His research interest includes Software Engineering, Software Reliability, Neural Networks and Data Mining.