IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-6, Issue-8, Aug 2018 E-ISSN: 2347-2693

Breast Cancer Prediction Using Soft Computing Techniques – A Survey

M. Durairaj [1*], K. Priya [2]

[1]School of Computer Science and Engineering, Bharathidasan University, Tiruchirappalli, India.
[2] School of Computer Science and Engineering, Bharathidasan University, Tiruchirappalli, India.

Corresponding Author: durairaj.bdu@gmail.com, Tel.: +91- 94875-42202

Available online at: www.ijcseonline.org

Accepted: 15/Aug/2018, Published: 31/Aug/2018

Abstract— Breast cancer is a disease where there is excessive growth or uncontrolled growth of cells of the breast tissue. Breast cancer is a type of cancer that is often found as lump, bloody nipple, pain or sore and change in size in the most of the cases. Breast cancer occurs when the cell tissues of the breast become abnormal and uncontrollably divided. These abnormal cells form a large lump of tissues, which consequently becomes a tumor. Digital imaging techniques like Scintimammography are used to analyse metabolic activities and vascular circulation for pre-cancerous analysis based on breast tissue. Finally biopsy result is used to ascertain cancer when other physical exam and mammograms show breast change. Soft computing techniques are interestingly gaining popularity in medical disease diagnosis and decision making. There are different soft computing techniques for medical data processing. These techniques can be used individually or hybridizing more to process medical data yields near accurate results in decision making by medical practioners. This paper reviews different breast cancer diagnosis methodologies which use information obtained from imaging techniques such as Magnetic Resonance Imaging (MRI), Mammogram, Ultrasound and Biopsy. Aim of this paper is to propose or identify methodologies that process cancerous images obtained from different imaging techniques and predict breast cancer with relatively better accuracy. In this paper, we reviewed research articles published in the recent years on breast cancer prediction using soft computing techniques. Comparative analysis of different methods in terms of accuracy, sensitivity, specificity and computational time is presented.

Keywords—Soft Computing, Machine Learning, Breast Cancer, Disease Diagnosis, Mammograms, MRI, Classification

I. INTRODUCTION

Breast Cancer is the most common cancers among women. The breast cancer in India is rapidly becoming the number one cancer among females. Breast mammography is one of the strategies that utilized for early identification of the Breast disease. Breast mammography is a radiation apparatus for identifying a conceivably destructive process much sooner than the presence of a tumour. As indicated by the American Academy of mammography, it is the therapeutic science that gets symptomatic signs from profoundly nifty gritty and delicate infrared pictures of the human body. Mammography is totally non-contact and includes no type of vitality conferred onto or into the body. Mammography has perceived applications in chiropractic, dentistry, neurology, orthopaedics, word related drug, torment administration and vascular medicine.

Mammography is a kind of imaging techniques that utilizes low-measurement x-beams to diagnose disease in earlier stage. Since, this mammography is useful to diagnose breast cancer in very earlier stage before women actually experiencing symptoms, makes the cancer most treatable. Such an imaging technique provides physiological information. Mammography shows mass patterns which may indicate cancer infection, inflammation, surface lesions and more. The CAD (Computer Aided Detection) is used to see digitized mammographic images, it's abnormalities based on density, mass, calcification. The CAD framework features these territories on the pictures, alarming the radiologist to painstakingly evaluate this zone. Mammogram uses infrared cameras to detect radiation with the range of electromagnetic spectrum which is the size of 0.9 to 14 and produces the images. Since infrared radiation emitted by all objects depends on their temperatures, according to the black body radiation law. Computerized mammography is called as Full-Field Digital Mammography (FFDM). This is a mammography system in which the x-ray film is replaced by electronics that convert x-rays into mammographic pictures of the breast. These systems are similar to those found in digital cameras and their efficiency enables better pictures with a lower radiation dose. The radiologists' store these pictures in computer for later use.

Breast Mammography is a diagnostic process that takes images of the breasts for early detection of breast cancer. It is an important tool in Breast cancer Screening. The technique utilizes the rule that synthetic and vein movement in both precancerous tissue and the area surrounding the tissue in breast cancer is almost always higher than in the normal breast. Anyway precancerous and malignant masses are very metabolic tissues and they require a vast sum supply of supplements to keep up their development. So to do this they increment course of their cells by conveying synthetic substances to continue existing veins open and make new ones (neo-angiogenesis).

The precancerous and cancerous masses are highly metabolic tissues. These tissues need a large amount supply of supplements to maintain their development in order to do this they increase cells circulation by sending out chemicals. This process keeps existing blood vessels open and create new ones called neo-angiogenesis. This process results in an increase in surface temperatures of the breast. Mammography uses an infrared camera and computer to detect analyses and produce high resolution images of these temperature changes in the breast.

A. Imaging Instruction

Skin surface temperature is greatly affected by various conditions. In order to reduce the errors due to thermal artefacts, images are taken using a recommended set of instructions to ensure the usefulness and consistency of thermal images. For Thermal Breast Scan, certain protocols must be followed in order to ensure that the images convey accurate information is mentioned below.

- Prolonged sun exposure to the chest and breast should be avoided at least 5 days before the scan.
- 2It is better to avoid creams, lotions, powders, makeup on the breasts on the day of the exam.
- Hair removal and shaving in the areas around chest, breasts, under arms for one day before the scan.
- Exercise six hours prior to the scan should be avoided.
- There should not be any physical stimulation of the breasts before one day of the scan

To improve the accuracy of diagnosis of breast cancer, Computer-aided diagnosis/detection (CAD) techniques are used. The major method used by the physician to detect the breast abnormalities present in the mammogram image is by visual inspection of asymmetries in the left and right breast mammogram images. It is impossible to have a breast tumour growing symmetrically in the left and right breast. Due to the limitations of human visual system, it is not possible to detect all kinds of abnormalities present in the breast mammogram images. Appropriate segmentation technique to differentiate left and right breast for asymmetry analysis is needed.

The accuracy of detection of breast cancer using a computer aided detection system for breast thermograph depends on the quality of the collected data and several image processing techniques such as Pre-processing, Segmentation, Feature extraction and Classification (final decision). Gathering a good informative and an adequate data is very important to get accurate results from the CAD systems. Pre-processing and appropriate segmentation are the essential requirements for the extraction of relevant features from the mammogram images. Development of a fully automated segmentation is a very difficult task due to their amorphous nature and lack of clear limits. It also depends on the distance from which the image is captured, height of the image, the size of the breast, image background, presence of noise etc. The mammographic image is shown in Figure 1.

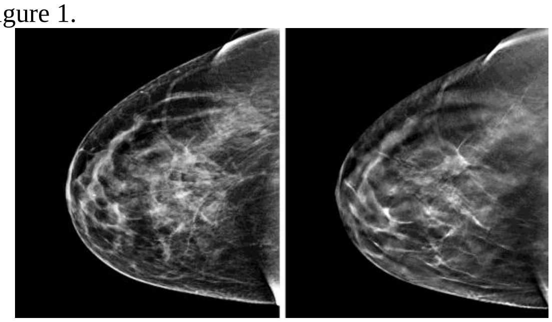

Figure 1. Mammographic Image

B. Soft Computing Techniques

Soft Computing is combination of different methodologies that can be used to develop model of real world problems to enable solutions, which are too difficult to address mathematically. Different Soft computing techniques are works synergistically and provides a form with capable of process the information flexibly in handling of real-life situations. This is to exploit the imprecision, uncertainty, approximate reasoning and partial truth in order to achieve tractability, robustness and low-cost solutions. The guiding principle is to devise methods of computation that lead to an acceptable solution at low cost, by seeking for an approximate solution to an imprecisely or precisely formulated problem. Major part of Soft Computing techniques uses Neural Networks. There are millions of very simple processing elements or neurons in the brain, linked together in a massively parallel manner. This is believed to be responsible for the human intelligence and discriminating power. Neural Networks are developed to try to achieve biological system type performance using a dense interconnection of simple processing elements analogous to biological neurons. Neural Networks are information driven rather than data driven. Typically, there are at least two layers, an input layer and an output layer. One of the most common networks is the Back Propagation Network (BPN) which consists of an input layer, and an output layer with one or more intermediate hidden layers.

C. Soft computing application for cancer image diagnosis

Soft Computing based medical image analysis is a promising field and researchers around the world shown their interest. It includes image enhancement, segmentation, classification-based soft computing, and their application in diagnostic imaging, as well as an extensive background for the development of intelligent systems based on soft computing used in medical image analysis and processing. In Medical imaging system, the images which are used in an efficient way to acquire, process, display and interpret information about complex systems. These images reveal certain characteristics of the object under consideration and how these characteristics change with time. A general purpose imaging system includes a physical device that is sensitive to the energy radiated by the object we wish to image. There is a device called digitizer, which converts the output of the physical sensing device into digital form, which is understandable, by computer system. Finally, this computer system processes the image and result in some diagnosis. So image processing is one of the basic components of CAD system. So its function is to enhance and extract the features of parts under consideration. Recent developments, which are the base of medical imaging, are X-Ray, Computer tomography, magnetic resonance, ultrasonography, and nuclear medicine. These developments are contributing a lot in improving the health and well-being of people worldwide. This paper aims to review the research articles published during last five years on the breast cancer diagnosis and applications of soft computing techniques for the prediction of cancer cells through image analysis. In this work, recently published 50 research articles are reviewed and summarized. Based on the review, this paper aims to propose a methodology which effectively diagnosis and predicts cancer cells in the early stage of disease.

This paper is organized as, the Section I consists of a brief introduction of cancer disease, impacts, diagnosis and soft computing techniques. Section II elaborates the related work. Section III provides the details about existing methodologies. Section IV explains the proposed work with provide comparative analysis in Section V and lists the various techniques in experimental result in Section VI. In Section VII, the overall survey paper is presented with a conclusion.

II. RELATED WORK

Han, Fei, et al. [1] proposed the system for quantitative analysis in breast mammography by taking into account embedded equivalent point heat source and the irrelevant temperature distribution of raw data, which was scarcely observed in previous work. A semi-analytical modelling technique was developed to investigate the relationship between breast mammography and the information on the internal tumors. The calculation and estimation of the properties of the internal heat source in this paper served as a quantitative analysis of tumors. The current study enables quantitative estimation of breast cancer detection and is promising in providing the satisfactory approaches to distinguish between benign and malignant tumors. The clinical cases verified that the methodology could be efficiently applied to clinical practice, regardless of the shape or size of breasts. This work promotes breast cancer detection using IR (Infra-Red) imaging in tumor classification.

Boquete, Luciano, et al. [2] conducted a preliminary studyof the use of Independent Component Analysis (ICA) in the early detection of breast cancerbased on mammographic images. ICA is a subspaceprojection technique that projects data from a high dimensional space to a lower-dimensional space.This work evaluated the application of the ICA algorithm to extract areas with high relative body temperature associated with tumor regions of high risk. The first results are encouraging and show the potential appeal of ICA applied to early detection of breast cancer showing a high accuracy, discrimination of the relative highest temperature zone associated with tumor areas, pinpointing minimum defects of 4×4 pixels, representing 1% of the overall size of the thermogram. In this method, the appearance of a heat anomaly indicates a potentially cancerous zone which reflected as an independent source.

Ali, Mona AS, et al. [3] proposed an automatic breast segmentation approach based on the distance between body and camera is 1 meter (dynamic protocol). In order to evaluate the accuracy of the proposed segmentation approach, different selective features were extracted from the segmented regions and then the SVM classifier was used to detect the breast abnormalities. This work proposes an automatic segmentation method for thermograms. The results prove its reliability in extracting the ROI (Region of Interest)for different cases. This method uses statistical and texture features from the segmented ROI and the SVM (Support Vector Machine) with its kernel function to detect the normal and abnormal breasts.

W. Peng, et. al. [4] proposed an Automated Confirmatory System (ACS) to process digitalized mammogram online. This method generates a high quality filtered segmentation of an image for biological interpretation and texture-feature diagnosis. The authors use different image processing and segmentation techniques include 2D median filtering, seeded region growing (SRG) algorithm, image contract enhancement, to remove noise, delete radiopaque artefacts. The elimination of the projection pectoral muscle from a digitalized mammogram was observed. They used Rough Set for extracting features and ANN for classification as normal, benign and malignant tumor.

H. G. Zadeh, et.al. [5] proposed a combinatorial model that consists of back propagation Neural Network and Genetic Algorithm. In their work, the best diagnostic factors are separated from other factors. The diagnostic specifications were inserted into the combinatorial model and the system selects and extracts the diagnostic parameters. One of the reasons for using the Neural Networks is its ability to simulate the non-linear functions. But to simulate the effect of the dependent variables on independent variables the genetic algorithm is used. They had 8 diagnostic parameters in this system. The best diagnosed results of breast cancer obtained from intelligent system without the help of physicians. This work suggests that the information processing volume must be as low as possible to reduce the process's time and errors, and prevent the system from diagnostic errors. Only eight parameters were used which are very low and can be processed easily, but the intelligent systems need to be trained using large data.

NehaTuteja, et.al. [6] proposed an energy minimization technique to estimate and correct the bias field and segment the mammogram images simultaneously. It optimizes two multiplicative intrinsic components of images in the mammogram, the bias field and the true image. Bias field deals with intensity in homogeneities present in the image space and the true image defines a physical property of the tissue. Segmentation of images is one amongst the primitive and most vital stages of processing in images and plays a very crucial role in analysing medical mammogram images. The images of mammogram have moderate level of distinction and are disrupted with sturdy speckle noise. Due to high noise, low distinction, and alternative imaging artefacts, region boundaries in mammogram images often do not adjust to the assumptions of many image processing algorithms. This paper addresses the potencies and weaknesses of the existing techniques of carcinoma detection in mammograms

Anuj Kumar Singh, et.al. [7] developed an approach which can segment malignant regions properly. This work addresses the problem of breast cancer detection with mammogram images. As consumption of time in execution is also important to provide good results in real time, they developed an important a method which first detects the cancerous region and then segment the area covered by malignant tissues. The malignant tissues were detected based on their higher intensity values compared to background information and other regions of the breast. However, in case of some normal dense tissues having similar intensities to tumor region, it is necessary to detect tumor region excluding those regions successfully. A method including detection followed by segmentation of mammogram images based on simple image processing techniques was proposed which as they claim provides good results in real time. In their work, a simple and easy approach was introduced for detection of cancerous tissues in mammogram. Detection phase is followed by segmentation of the tumor region in a mammogram image. Their approach uses simple image processing techniques such as averaging and thresholding. A Max-Mean and Least-Variance technique was used for tumor detection.

In [8], a multi resolution analysis system for interpreting digital mammograms is proposed. This system is based on using fractional amount of biggest wavelet coefficients in multilevel decomposition. A set of real labelled database is used in evaluating the proposed system. The evaluation results show that the system has a remarkably high efficiency compared with other systems known till present, especially in the area of distinguishing between benign and malignant tumors. Designing an effective diagnosis system for digital mammograms is still a challenging problem that needs more investigation. Two main functions included in this system as the first is to distinguish between normal and abnormal tissues, speculated lesions, circumscribed masses, and ill-defined lesions. The second function is to differentiate between benign and malignant tumors.

Prathibha, et.al. [9] proposed a methodology to analyze mammogram images and classify normal, benign and malign using Stationary Wavelet Transform (SWT) and Discrete Cosine Transform (DCT) with Support Vector Machine (SVM) classifier. The DCT gives a subset of the transform coefficients that is sufficient to preserve the most important features. The SWT gives a better approximation since it is redundant, linear and shift invariant. This system extracts SWT features in DCT domain and giving a more precise classification. Through the experiments, it was found that the SVM classifier with RBF kernel works excellently in the hybrid transform domain comparing to K-Nearest Neighbor (KNN) and Kernel Discriminant Analysis (KDA) classification techniques. It suggests that the computerized mammograms bolstered by Computer Aided Diagnostic (CAD) frameworks help the radiologists in taking solid choices. The wavelet and spectral features and spectral features were extracted in this system for classification of mammograms.

NebiGed, et.al. [10] implemented a CAD system and investigated the performance of Computed Tomography (CT) method in the problem of recognizing breast cancer in the ROI of digital mammograms. A different classification scheme is applied in the classification of breast cancer. The originality of the study is that the different features obtained by CT are used in the hybrid classification of breast cancer with the SVM, k-NN, PCA (Principal Component Analysis), and LDA (Linear Discriminant Analysis) algorithms. The presented results demonstrate that CT is a useful tool to discriminate malignant, benign, and normal tissues.

M. Durairaj, et. al. [11] compared and analyzed different classifier algorithms. The classifiers namely Naïve Baise, Decision tree and Lazzy classifiers were applied on the data set obtained from UCI repository. These three algorithms give accuracy of 82% to 91 %. In [12], Wisconsin diagnostic breast cancer (WDBC) dataset was preprocessed using independent component analysis (ICA). The accuracy of the classifiers such k-NN, ANN, RBFNN (Radial Basis Function Neural Network), and SVM were tested on the data set for detecting benign and malignin cancer cells.

III. METHODOLOGY

Mammographic images are used to diagnose breast cancer after a lump or other symptoms appeared. Signs of breast cancer include pain, skin thickening, nipple discharge, or a change in breast size or shape. Strong correlation exists between breast cancer and abnormalities of tissues in mammograms. The CAD system should classify the breast tissues in mammogram images into regions of interest (ROIs) such as calcifications, macro-calcifications, cysts and fibro adenomas. Thus, radiologists could diagnosis using CAD systems with the abilities of automated classification of breast tissues. This section explains mammograms and some of the computing methodologies applied to interpret mammographic images for understanding and classification.

A. Mammograms for Diagnosis

Studying mammograms specific features are sought in routine examinations as common indicators of malignancy. The abnormalities or some specific features appear in the mammogram images can be classified into calcification, circumscribed masses, speculated masses, ill-defined masses, architectural distortion and asymmetry.

1) *Calcifications:* Calcifications are calcium deposits present in the breast, calcifications are further classified into Macro-calcifications and Micro-calcifications. Macro-calcifications are the large calcium deposits which are not associated with cancer. Micro Calcifications (MC) appear as small white spots similar to grains of sand with a diameter of less than 0.5 mm and are grouped closely together to form clusters.
2) *Circumscribed masses:* Uniform and smooth masses appeared in the shape of irregular circles are called as circumscribed masses.
3) *Spiculated lesions:* Spiculated lesions appear as a region with segments distributed in many directions as a multi-arms star.
4) *Ill-defined masses:* Ill-defined masses are the masses that do not have a fixed pattern.
5) *Architectural distortion:* An architectural distortion on a mammogram is basically a distortion of the regular random shape of curvilinear and fine linear radiopaque structures which is normally realized on the mammogram. There will not be any visible masses but the distortion often visible as a 'stellate' shape or with radiating speculation.
6) *Asymmetrical distortion:* Asymmetrical distortion refers to asymmetry between the breasts (left and right breast of the same patient), it can be in shape or in certain area in the mammograms.

B. Mammographic Image Analysis

In the process of mammographic image analysis for cancer diagnosis, different domain of classification procedures are exists such as Texture Feature Classification, Statistical Modelling and Machine Learning approaches. This section briefly discusses some of these cancer image analysis approaches.

1) *Texture Feature Analysis:* Texture Feature Analysis typically analyze the characteristics of texture features extracted from Regions of Interest (ROIs) to classify into well-known knowledge categories. Measures of the skewness of the image brightness histogram and measures of image texture characterized by the fractal dimension are strongly correlated with radiologists' subjective classifications of mammographic parenchyma. Spearman Correlation Coefficients are used for skewness and fractal dimension measurements [1].

A discrimination of breast density is based on the underlying texture of breast tissue deceptive on a digital mammogram. The testing data set is split into four categories for analysis. They are listed as:

a) the breast area which is predominantly fat
b) the breast area which is fat with fibro glandular issue
c) the breast are which is heterogeneously dense
d) the breast area which is extremely dense.

Discriminating lesions from normal tissues are based on intensity, contrast, isodensity, location and texture. The combination of different classifiers can be used to segment the breast area into fatty versus dense mammographic tissue. The extracted morphological and texture features from the segmented breast areas can be processed using appropriate classifiers. Linear Discriminant Classifier (LDC) using the multi-resolution texture features can be used for effective classification of masses from normal tissues on mammograms. Texture features such as wavelet coefficients, variable distances, and the average area and under the ROC curve are useful for performance analysis. The computerized mass detection program was used for multi-resolution global and local texture features to reduce false positive detection [3].

In [13], digital mammography images were obtained from the Picture Archiving and Communication System (PACS) for analysis. Texture features of mammographic images were calculated. In order to differentiate benign and malignant group, Mann-Whitney U test was used and Receiver Operating Characteristic (ROC) curve analysis were used which assess performance of texture features.

2) *Statistical Modeling:* Statistical Modeling of ROI's features are important to identify different types of breast tissues. A commonly employed statistical model is Linear Discriminant Analysis (LDA). Stepwise feature selection and LDA are employed to identify features that differentiate low-risk women and gene-mutation carriers.The computer-extracted features may be useful for identifying women at high risk for breast cancer and for monitoring the treatment of breast cancer patients. As an alternative to LDA, an approach based on Generalized Additive Models (GAMs) to deal with a broad variety of variables and to reduce the number of false detections

[14]. The results showed the GAMs approach had better performance than LDAs.

Texture models are used to capture the mammographic appearance within the breast area [15]. The parenchymal density patterns are modelled as a statistical distribution of clustered, rotationally invariant filter responses in a low dimensional space. A physical model for image acquisition determines the dense tissue and maps it to a pixel for estimation of volume of dense breast tissue. The density function of the model is represented by a mixture of up to four weighted Gaussians, each one corresponding to a specific density class in the breast. A devised mammogram modelling system which segmented the five major components of a mammogram as background, uncompressed-fat, fat, dense, and muscle. Automated algorithms can consider the components independently or adapt their parameters based on component-specific statistics. A Finite Generalized Gaussian Mixture (FGGM) model is a heuristic optimization approach [16] to estimate the model parameter set more accurately by Particle Swarm Optimization (PSO) and evolutionary programming (EP) techniques [15].

The statistical methods like Binary Logistic Regression (BLR) and Receiver Operating Characteristic (ROC) were used to assess the performance of texture analysis for diagnosis of benign and malignant breast tumors [13].

3) *Machine Learning Technique:* Huge data from different mammograms attracts data science experts to device technologies based on Machine Learning to process and extract required information from the data.Since a machine learner can take advantage of examples to capture unknown underlying characteristics, classification techniques based on machine learning are very popular currently Machine learning classifiers aim to automatically learn to recognize complex patterns and classify data intelligently. However, the performance of different machine learning methods may vary.

Kernel based methods such as SVM (Support Vector Machine), KFD (Kernel Fisher Discriminant) [17] and RVM (Relevance Vector Machine) are effective tools for classification [18]. The breast mass classifications as benign and malignant can be improved with the usage of suitable machine learning tools with reduced features of data set. The Artificial Neural Network (ANN) [19] with back propagation algorithm and Radial Basis function Neural Network (RBFNN) [19] has showed excellent accuracy in micro calcification detection task.

Probabilistic Neural Network (PNN) is a network formulation of probability density estimation. It is a model based on competitive learning with a concept of “winner takes all attitudes” and the core concept is based on multivariate probability estimation. The PNN architecture consists of an input layer, a pattern layer, a summation layer, and an output layer. Figure 2 shows the basic structure of PNN.

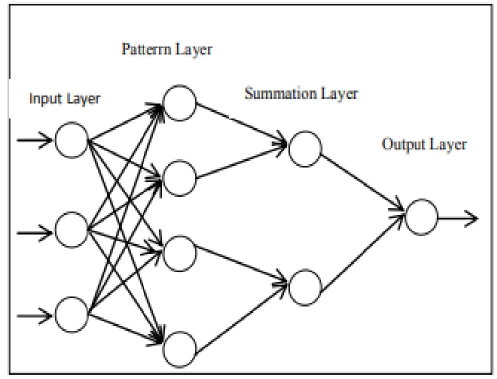

Figure. 2. Probabilistic Neural Network

IV. EXPERIMENTAL RESULTSAND DISCUSSION

Comparative analysis of various research articles and their proposed work are presented in this section. The comparison restuls of various techniques applied for breast cancer image analysis are depicted in Table 1. The algorithm proposed for quantitative analysis in breast thermography is used to classify malignant and benign based on embedded equivalent point heat source and irrelevant temperature distribution of raw data [1]. This work observed significant difference in diagnosis of the malignancy of breast cancer. Statistical analysis shows the negative predictive value is very high ensuring that no cancer case was being overlooked [2]. The positive predictive value indicates that this method detects abnormal tissue in earlier stages. The automated breast segmentation approach was used ROI and SVM for segmentation and classification of breast abnormalities [3]. An Automated Confirmatory System (ACS) was proposed to process a digitalized mammogram online [4]. This ACS generates high quality filtered segmentation of a mammogram image for biological interpretation and texture-feature based diagnosis. Infrared imaging was used to create a database and a combinatorial model of GA and ANN was used to analyze the effect of the independent variables on the dependent variables [4]. This work highlights the parameters which indicate breast cancer.

Computer-aided detection systems and intensity-based methods were introduced for breast cancer segmentation in mammogram images, but not the best promising [6]. No techniques satisfy the detection criteria of finding cancerous region successfully. The malignant tissues were founded by creating a rectangular window around the outputted region area and applied Max-Mean and Least-Variance techniques [7]. In this work, tumor patch was found using morphological closing operations and region boundary was found using the image gradient technique. Abnormality indicators such as micro calcifications, circumscribed masses, speculated lesions and ill-defined masses are classified using supervised classifier for mammograms [8]. SVM classifier with RBF kernel produces optimal smoothing parameters for classification [9]. Support Vector Machine (SVM) and K-Nearest Neighbor (K-NN) algorithm are applied for classification to build the diagnostic model and Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) applied for further feature selection from mammograms [10].

A. Performance measure

Performance of the classifiers can be evaluated in several methods. Researchers use confusion matrix to keeps their results correct and incorrect classification to measure the quality of classifier. The sensitivity and specificity analyses are performed in many of the research results to measure performance of their proposed works. The expression used for calculations are shown in "equation (1)", "(2)" and "(3)":

$$\text{Sensitivity} = {}^{\text{TP}}/_{(\text{TP}+\text{FN})}\,(\%) \tag{1}$$

$$\text{Specificity} = {}^{\text{TN}}/_{(\text{FP}+\text{TN})}\,(\%) \tag{2}$$

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{3}$$

where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative counts, respectively.

B. Comparative Analysis

In Table 1, the comparative analysis of various mammographic image classifiers for breast cancer diagnosis is illustrated. Some of the methodologies applied for image classification were used computation time as a performance analysis metric rather than accuracy, sensitivity and accuracy [6,7]. In majority of works used accuracy as a standard metric to evaluate performance of their methodology [4], other methods were used sensitivity and specificity metrics to analyse their performance [2,5,9,10]. The performance evaluations of different methods reviewed in this paper is presented in Figure 3.

Table 1. Comparative Analysis of Various Classifiers on Breast Cancer Interpretations

Author names	**Methodology**	**Procedure / Performance**	**Advantages**	**Disadvantages**	**Accuracy, Sensitivity and Specificity**
Han, Fei, et al. 2015 [1]	• Analytical-based steady-state algorithm • Nonlinear heat conduction model	• Uses the breast thermography to classify malignant and benign based on embedded equivalent point heat sources	• Extracts shape and size of breast • Served as quantitative analysis of cancer	• Need to be tested with larger data set of patients	• Specificity 98%
Boquete, Luciano, et. al. 2012 [2]	• Independent Component Analysis (ICA)	• Uses thermographic image analysis for automated detection of high tumor risk areas	• Observed high accuracy discrimination of relative highest temperature zone associated with tumors	• Broader study is needed to validate.	• Sensitivity 100% • Specificity 94.75%
Ali, Mona AS, et. al. 2015[3]	• ROI and SVM techniques	• Statistical and texture features from ROI were extracted and SVM used to detect the normal and abnormal breast.	• Extracted ROI of breast thermograms and classified with SVM shows high accuracy	• Conversion of data needed multiple methods like Linear, polynomial, RBF and quadratic	• Accuracy 100%
W. Peng, et. al. 2016 [4]	• Automated Confirmatory System (ACS) • Rough Set & ANN	• Processes digitalized mammogram online. • RS for pre-process. • ANN for classification.	• Generates high quality filtered segmentation of an image for biological interpretation.	• ANN needed to be trained offline	• Accuracy 92%
H. G. Zadeh, et.al. 2012 [5]	• Genetic Algorithm and ANN	• Breast cancer image processed by quantitative and qualitative information of medical infrared	• Shows good precision in diagnosis	• High level computational complexity due to ANN	• Sensitivity 50%, • Specificity 70%

		imaging			
NehaTuteja, et.al. 2018 [6]	• Energy Minimization Technique	• Studies the weakness of existing techniques by comparing different techniques	• Suggests good quality Mammograms can find cancers with more accuracy • Finds no screen tool is 100% effective	• Study shows segmentation of breast cancer images are challenging task • Computational time • Inaccuracy in size of cancer	• Performance is not measured
Anuj Kumar Singh et.al. 2015 [7]	• Max-Mean and Least-Variance	• Simple image processing techniques, averaging and thresholding applied for segmentation.	• Highlights resultant region boundary • Detected malignant • Gives results in real time	• Comparison made with the only other method • Accuracy is not measured	• Lesser time • More accurately
Essam A. Rashed, et.al. 2007. [8]	• Supervised classifier • Multilevel wavelets decomposition analysis	• Supervised classifier applied for mammograms using discrete wavelet transform decomposition	**Classifies** • cancerous versus cancerous-free patterns • abnormality indicator • risk level	• Results and accuracy depends on the change of coefficients lesion	• Accuracy 100%
Prathibha, et.al. 2011. [9]	• CAD with SVM	• Gives a subset of the transform coefficients that is sufficient to preserve the most important features	• Achieves the goals of optimality by reducing the cost and increasing accuracy of diagnostics	• Dimensionality problems can exist	• Accuracy for discrimination of malign is 93.14% • Benign is 87.25%
NebiGed, et.al. 2013. [10]	• Curvelet Transform (CT) algorithm • SVM and K-NN as classifiers • PCA and LDA for reduction	• SVM and K-NN algorithm applied for classification to build diagnostic model. • PCA and LDA for further dimensionality reduction and feature selection.	• Classifies normal and abnormal mammograms • Effectively discriminates malignant, benign and normal tissues	• Vector construction become difficult	• Success rate 87% • CT Sensitivity 80% • Specificity 94%
Durairaj, M, et.al. 2015. [11]	• Naïve Baise, Decision Tree, Lazzy Classifier	• Irvine UCI dataset was used and three different classifiers were applied and compared	• Effectively classifies mammography images.	• Data set is not cleaned or pre-processed	• NB accuracy 91% • DT accuracy 91% • LC accuracy 82%
Ahmet Mart, et. al. 2015 [12]	• ICA for feature reduction. • k-NN, ANN, RBFNN, SVM for classification.	• WDBC data set is reduced using ICA. • Classifications done using k-NN, ANN, RBFNN and SVM.	• Feature reduced effectively. • Classification done with increased accuracy.	• Dimensionality problem exists in feature reduction	• **Accuracy:** ○ k-NN – 91.03 % ○ ANN – 90.5 % ○ RBFNN – 90.49 % ○ SVM – 90.86 % • **Specificity:** ○ k-NN – 84.9 % ○ ANN, RBFNN, SVM – 79.71% • **Sensitivity:** ○ k-NN – 94.6 % ○ ANN – 96.9 %

					○ RBFNN,SVM– 97 %
Zhiming Li, MD, et. al. (2017) [13]	• Mann-Whitney U Test, • Histogram, GLCM, RLM	• Calculates Texture features of mammographic images. • Mann-Whitney U test conducted to differentiate benign and malignant.	• Discriminate benign and malignant with higher accuracy level of 95% achieved.	• Diverse component of breast tumors affects the result.	• CI of Histogram–95%, • CI of GLCM-95%, • CI of RLM – 95%
Woo Kyung Moon, et. al. (2018) [20]	• Backward feature selection and Linear Logistic regression for feature selction, • ALN for prediction	• Texture and Morphological features of breast cancer image analysed.	• Computer-aided prediction (CAP) proposed. • Higher accuracy achieved	• Limited data supporting correlation between ALN status and US breast cancer data.	• Accuracy 75.1%, • Sensitivity 79 %, Specificity 71.5%

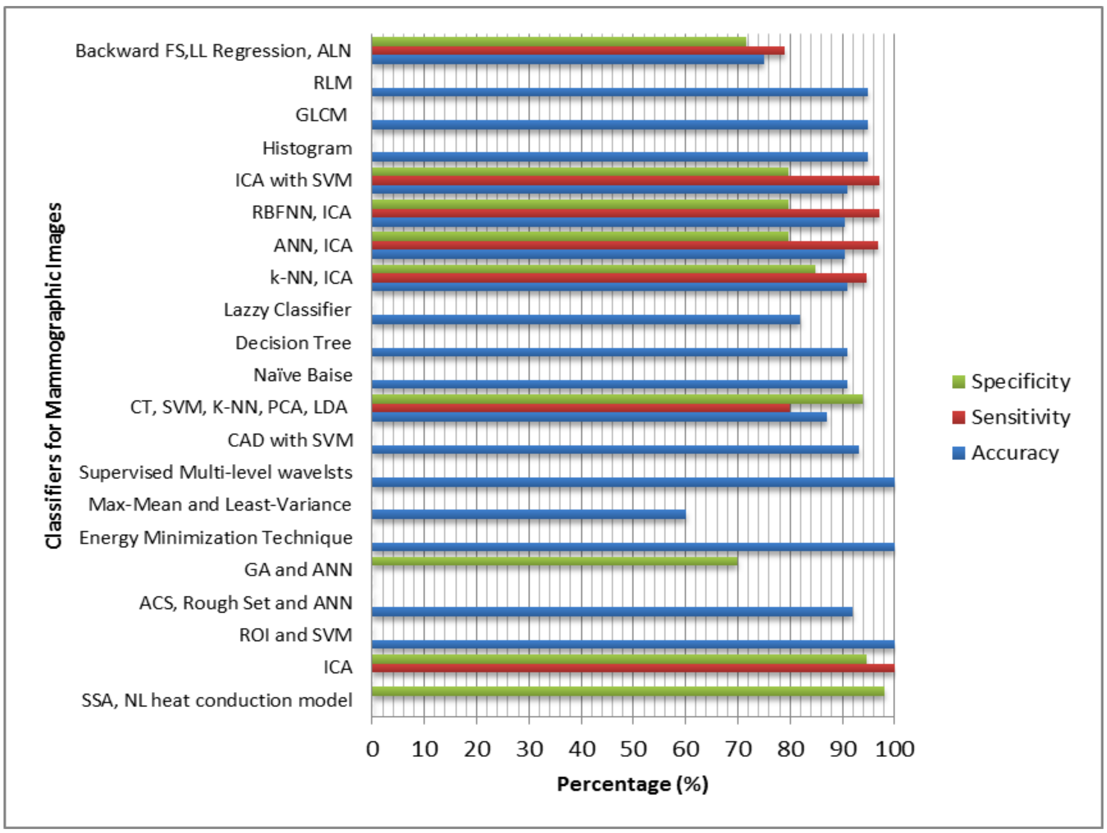

Figure 3. Different Breast Cancer Diagnosis Methods with Performance (Accuracy, Sensitivity, Specificity)

V. Conclusion

Different categories of classification techniques and soft computing methodologies applied for breast cancer diagnosis using mammographic images are presented in this paper. The categorization of these techniques with their advantages and disadvantages are discussed and tabulated.

This review is in agree with other researchers' findings such that of texture feature analysis are sensitive to types of mammogram machines, and statistics modeling could be inaccurate in some specific situations. Each and every methods own some sort of advantages and disadvantages. Diagnostic results of breast cancer using soft computing methods are depends on type of images and imaging techniques. The images play foremost roles in determining the results of breast cancer diagnosis. In this work, we reviewed various statistical and machine learning methodologies that perform analysis with texture feature of images and different data pre-processing techniques. We presented comparative analysis of various techniques based on their performance. This paper conclude based on our review results that suitable selection of single or combination of machine learning / soft computing algorithms, depends up on the data set, capable to yield results with the accuracy of more or equal to 95% on the earlier detection of breast cancer.

REFERENCES

[1] FeiHan, Guilian Shi, Chengwen Liang, Lin Wang, Kaiyang Li, "*A simple and efficient method for breast cancer diagnosis based on infrared thermal imaging*", Cell biochemistry and biophysics, Vol. 71, pp. 491-498, 2015.

[2] Luciano Boquete, Sergio Ortego, Juan Manuel Miguel-Jimenez, Jose Manuel, Roman Blanco, "*Automated detection of breast cancer in thermal infrared images, based on independent component analysis*"' Journal of medical systems, Vol. 36, pp. 103-111, 2012.

[3] Ali, Mona AS, Gehad Ismail Sayed, Tarek Gaber, Aboul Ella Hassanien, Vclav Snasel, Lincoln F. Silva, "*Detection of breast abnormalities of thermograms based on a new segmentation method*", In Proceeding of Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on IEEE, Vol. 5, pp. 255-261, 2015.

[4] W. Peng, R.V. Mayorga, E.M.A. Hussein, "*An Automated Confirmatory System for Analysis of Mammograms*", Computer Methods and Programs in Biomedicine, Vol. , pp. 1-22, 2016.

[5] Hossein Ghayoumi Zadeh, Javad Haddadnia, Maryam Hashemian, Kazem Hassanpour, "*Diagnosis of breast cancer using a combination of genetic algorithm and artificial neural network in medical infrared thermal imaging*", Iranian Journal of Medical Physics, Vol. 9, No. 4, pp. 265-274, 2012.

[6] Neha Tuteja, Parvinder Singh, MohitBansal, "*Comparative analysis of different techniques for breast cancer detection in Mammograms*", International Journal of Advnce Research, Ideas and Innovtions in Technology, Vol. 4, No. 2, pp. 1361-1365, 2018.

[7] Anuj Kumar Singh, Bhupendra Gupta, "*A novel approach for breast cancer detection and segmentation in a mammogram*", Procedia Computer Science, Vol. 54, pp. 676-682, 2015.

[8] Essam A. Rashed, Ismail A. Ismail, Sherif I. Zaki, "*Multiresolution mammogram analysis in multilevel decomposition*", Pattern Recognition Letters, Vol. 28, pp. 286-292, 2007.

[9] Prathibha, B. N., and V. Sadasivam, "*Mammogram analysis using the SVM classifier in combined transforms domain*", ICTACT Journal on Image and Video Processing, Vol. 1, No. 03, pp. 172-177, 2011.

[10] Nebi Gedik, AytenAtasoy, "*A computer-aided diagnosis system for breast cancer detection by using a curvelet transform*", Turkish Journal of Electrical Engineering & Computer Sciences, Vol. 21, pp. 1002-1014, 2013.

[11] M. Duraraj and R. Deepika, "*Comparative Analysis of Classification Algorithms for the Prediction of Leukemia Cancer*", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, No. 6, pp. 787–791, Aug. 2015.

[12] Ahmet Mert, NiyaziKilic, ErdemBilgili, Aydin Akan, "*Breast Cancer Detection with Reduced Feature Set*", Computational and Mathematical Methods in Medicine, Vol. 2015, article Id 265138, pp. 1-12, 2015.

[13] Zhiming Li, MD, Lan Yu, MD, Xin Wang, MD, Haiyang Yu, MD, Yuanxiang Gao, MD,Yande Ren, MD, Gang Wang, MD, Xiaoming Zhou, MD, "*Diagnostic performance of mammographic texture analysis in the differentialdiagnosis of benign and malignant breast tumors*", Clinical Breast Cancer, doi: 10.1016/j.clbc.2017.11.004, 2017.

[14] M. J. Lado, C. Cadarso-Suarez, J. Roca-Pardinas and P. G. Tahoces, "*Using generalized additive models for construction of nonlinear classifiers in computer-aided diagnosis systems*", In proceedings of the IEEE Transactions on Information Technology in Biomedicine, Vol. 10, No. 2, pp. 246-253, April 2006.

[15] Petroudi S., Constantinou I., Pattichis M., Tziakouri C., Marias K., Pattichis C, "*Evaluation of Spatial Dependence Matrices on Multiscale Instantaneous Amplitude for Mammogram Classification*", In: Lacković I., Vasic D. (eds) 6th European Conference of the International Federation for Medical and Biological Engineering. IFMBE Proceedings, Vol. 45, 2015.

[16] Selvan, S, Cecil Xavier, C,Karssemeijer, Nico,Sequeira, Jean, A. Cherian, Rekha, Y. Dhala, Bharathi, "*Parameter Estimation in Stochastic Mammogram Model by Heuristic Optimization Techniques*", Information Technology in Biomedicine," IEEE Transactions on. Vol. 10, pp. 685 – 695, 2006.

[17] Li Zhang, Wei-Da Zhou, "*Fisher-regularized support vector machine*", Information Sciences, Elsevier, Vol. 343, pp. 79-93, 2016.

[18] Cong Y.C., Brady M., Petroudi S, "*Texture Based Mammogram Classification and Segmentation*", In: Astley S.M., Brady M., Rose C., Zwiggelaar R. (eds) Digital Mammography. IWDM 2006. Lecture Notes in Computer Science, Vol. 4046, Springer, Berlin, Heidelberg, 2006.

[19] M. Durairaj, R. Nandhakumar, "*Feature Diminution by Hybrid Algorithm for improving the Success Rate for IVF Treatment*," Pakistan Journal of Biotechnology, Vol. 14, No. Sp-II, pp. 1100–1104, 2017.

[20] Woo Kyung Moona, I-Ling Chenb, Ann Yi, Min Sun Baea, Sung Ui Shina, Ruey-Feng Chang, "*Computer-aided prediction model for axillary lymph node metastasisin breast cancer using tumor morphological and textural features on Ultrasound*", Computer Methods and Programs in Biomedicine, Elsevier, Vol. 162, pp. 129-137, 2018.

Authors Profile

M. Durairaj completed his Master of Computer Applications in 1997 and Ph.D. in Computer Science in 2011 from Bharathidasan University. He is working as Assistant Professor in Computer Science at School of Computer Science and Engineering, Bharathidasan University. Prior to this he was working at National Research Centre on Rapeseed-Mustard (Indian Council of Agricultural Research), Rajastan, India and National Institute of Animal Nutrition and Physiology (Indian Council of Agricultural Research), Bangalore, India for 12 years. He has 20 years of teaching and research experience in his credit. He has more than 90 publications in his credit. His area of research includes Data Mining, Data Science, Soft Computing, Cloud Computing and Big Data Analytics.

K. Pria pursed Bachelor of Science and Master of Science from Bharathidasan University, Tiruchirappalli, India. She is currently pursuing M.Phil. in Computer Science from Bharathidasan University. Her main research work focuses on Soft Computing Algorithms and Data Mining.