

A Research on Computer Aided Detection System for Women Breast Cancer Diagnosis from Digital Mammographic Images

P. Malathi, A. Kalaivani

Abstract—The Women breast cancer is the most critical cancer that are found in women. Its the second important cause of death in the world. Breast cancer has been ranked number one cancer in Indian females with rates occurrence of 25.8 per 1,00,000 females and death rate 12.7 among 1,00,000. Generally breast cancer is a malignant tumor that begins in the cells of the breast and eventually it spreads to the surrounding tissues. Early detection and diagnosis can reduce the mortality rate. Radiologist misdiagnosis the disease due to technical issues such as imaging quality and human error. Radiologists can improve the performance of Computer Aided Detection/Diagnosis (CAD) systems to finding and discriminating between the normal and abnormal tissues. Breast cancer diagnosis can applied are applied recent CAD systems on imaging modalities such as mammogram, ultrasound, MRI and biopsy histopathological images. CAD system have four stages for diagnosis which are pre-processing, segmentation, Feature Extraction and Classification. CAD system are developed to reduce the time taken to diagnose the breast cancer and reduce the death rate. This paper focus on the survey of CAD system to detect women breast cancer disease from the digital mammographic images to achieve high accuracy and low computational cost.

Index Terms— Benign Tumor, Computer Aided Diagnosis (CAD), Digital Mammograms, Malignant Tumor.

I. INTRODUCTION

Cancer is the most harmful disease which is a collection of related diseases. Cancer is the uncontrolled growth of the cells, which may spread into other tissues [1]. This disease is usually evident in women but men also get affected from it. Breast cancer types are Ductal cancer, Lobular cancer, Paget disease of the nipple, Phyllodes tumour, Angiosarcoma. Recent advancement in medical field and more precisely the involvement of information technology in the medical field introduces a new diagnosis mechanism called Medical Image Processing. Image processing techniques easy to detect tumor from an infected breast and diagnose breast cancer. Several methods can be utilized to detect the breast cancer such as Magnetic Resonance Imaging (MRI), biopsy, and X-ray mammogram. X-ray mammography is one of the most reliable methods for early detection of breast cancer [2]. Early detection of breast cancer is need to reduce the death

rate. So automated computer aided detection is inaveitable. It may be genetic or environmental. Genetic factors include family history, personal health history, menstrual and reproductive history, dense breast tissue, certain genome changes, age, gender etc. Environmental factors that can include obesity, poor diet, alcohol consumption, radiation, lack of physical activity. Computer Aided Detection (CAD) systems developed to help radiologists in the evaluation of mammographic images. The CAD system can detect the region of suspicion, identify whether tissue is normal or abnormal, and differentiate between cases to assess whether they are benign or malignant. This can enhance the prediction accuracy of tissue abnormalities (e.g. microcalcifications, masses, and distortions).

In this paper, we present an extensive survey on the research work carried to develop computer aided diagnosis tools that can help the radiologists in making accurate interpretation of the digital mammograms. The approaches which applied to design different stages of CAD system are summarised.

The rest of the paper is organized as follows: Section II gives the methodology of women breast cancer diagnosis. In Section III the research work on breast cancer is discussed. Section IV gives the performance analysis of breast cancer diagnosis. Section V describes the inference from the survey. Section VI gives the future research direction and concludes the paper.

II METHDOLOGY

The Computer Aided Diagnosis system of Women Breast Cancer Diagnosis consists of five major tasks such as preprocessing, segmentation, feature extraction and classification.

Revised Version Manuscript Received on 16 September, 2019.

P. Malathi, Assistant Professor, Department of CSE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamilnadu, India.

(Email: malathip.sse@saveetha.com)

A. Kalaivani, Associate Professor, Department of CSE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamilnadu, India.

(Email: kalaivania.sse@saveetha.com)

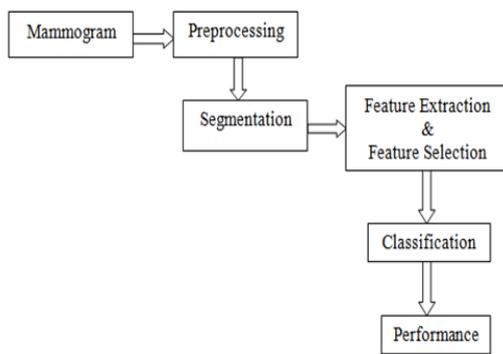


Fig. 1. CAD System for Women Breast Cancer Diagnosis[6]

A. Pre-processing:

Preprocessing is a method to convert an image into digital form. Removing the noise and any kind of irregularities present in an image using the digital computer. The median filter reduces the noise sources, while preserving the texture of mammographic images, so that it has better performance than other types of filters such as linear filters. The original mammograms images can be taken from the Mammographic Image Analysis Society (MIAS) database. After removing noise and artifacts sources, the pectoral muscle is removed. Image enhancement is the process of adjusting digital images, remove noise, sharpen, or brighten an image, making it easier to identify key features. After the preprocessing stage, mammogram images are enhanced using the un-sharp mask technique because of its ability to provide high contrast improved index and high signal to noise ratio.

B. Segmentation:

The main purpose of segmentation is the separation of the region of interest (ROI). Image segmentation is the vital step to partitioning a digital image into multiple segments and efficient step towards the development of CAD systems. The goal of segmentation is to simplify the representation of an image into something that is more meaningful and easier to analyze. The main purpose of segmentation is the separation of the region of interest (ROI) to identify breast lesions. The appropriate segmentation method also depends on the image modality and nature of the disease. The similarity-based method is categorised into region-based, thresholding-based, and clustering-based methods. Edge-based segmentation methods are a structural technique to detect edges or pixels among different regions that have abrupt intensity change. Thresholding-based automatic selection of threshold value requires the knowledge on the intensity characteristics of the objects, sizes of the objects, and the number of various types of objects existing in the image.

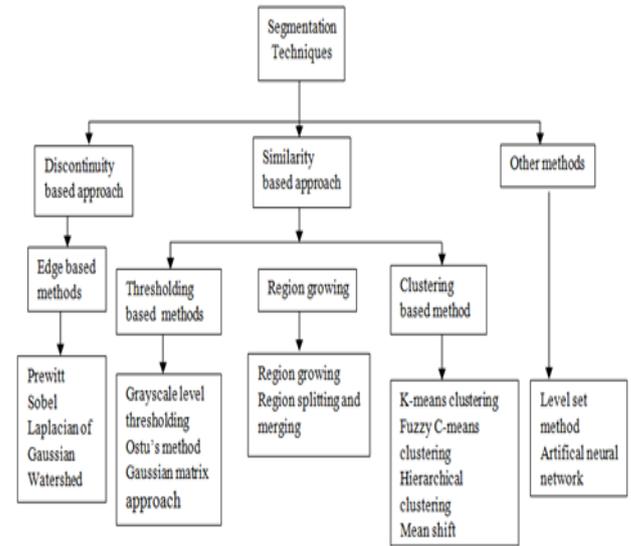


Fig.2. Overview of Segmentation Techniques[6]

SEGMENTATION TECHNIQUES	DESCRIPTION	ADVANTAGES	DISADVANTAGES
Region-based approaches	Grouping pixels in homogeneous regions based on seed points	Simple concept, It performs well with respect noise. We can choose multiple criteria at the same time	Dependence to selected seed points. In terms of computation time and memory are expensive
Thresholding based approaches	Achieve threshold value based on peaks and valleys in the histogram of images which corresponds to regions	Works well with low computation complexity in most of the images. Simplicity of threshold calculation.	Does not work well in images with close color spectrum
Edge-based approaches	It works based on identify sharp discontinuities in the image	Easy to perceive for human. Works well for image with good contrast.	Very sensitive to noise. Does not work well on images with low contrast and smooth change
Clustering based approaches	Categorize objects into the specific groups based on their similarity	Easy task to implementation	How to define number of clusters

Table 1: Comparison of segmentation techniques[6]

C. Feature Extraction and Feature Selection:

Different features are extracted based on the characteristics of lesions from the breast image. These features selected should distinguish benign or malignant lesions. The feature set is usually very large and the selection of the most effective features is important for the accuracy of the system. Feature selection is also called variable selection or attribute selection. It is the automatic selection of attributes in your data. Features are reduced by eliminating closely related data without affecting the accuracy of the classification.

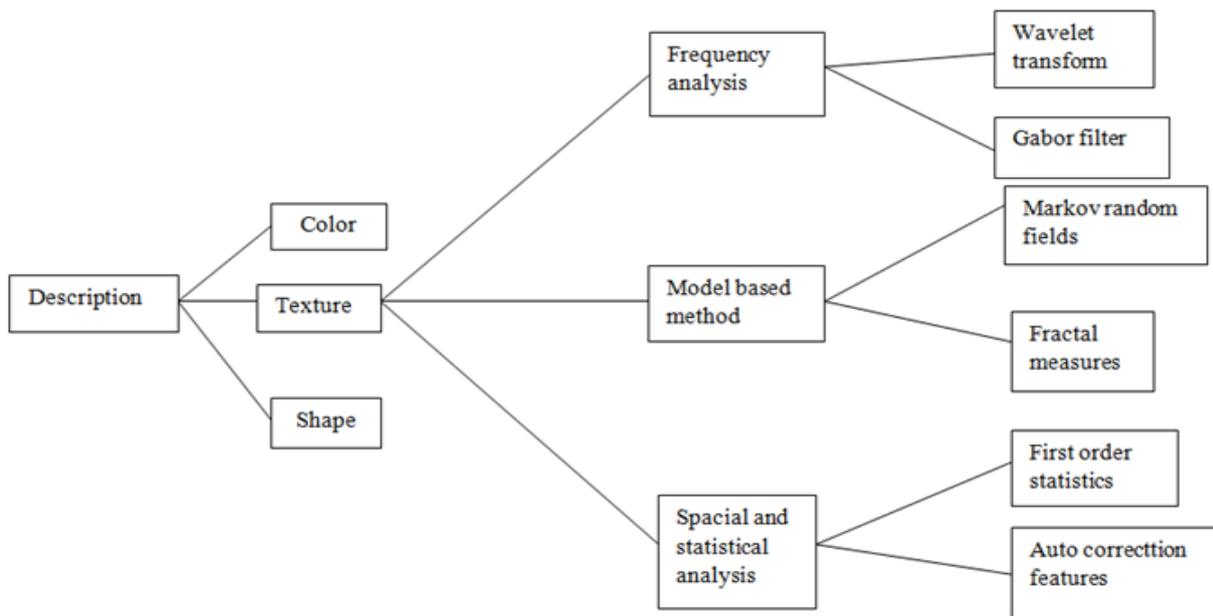


Fig 3. Overview of Feature Extraction[6]

D. Classification:

Classification methods play an important role in the diagnosis and educational purposes in medicine. According Feature extraction starts from an initial set of measured data and builds derived values. Dividing the ROI into sub-images and then extracting these features would result a higher cost

in the overall algorithm implemented in this paper. Feature extraction involves reducing the amount of resources required to describe a large set of data. When performing analysis of complex data one of the major problems stems from the number of variables involved. The description is provided in images using color, texture and shape.

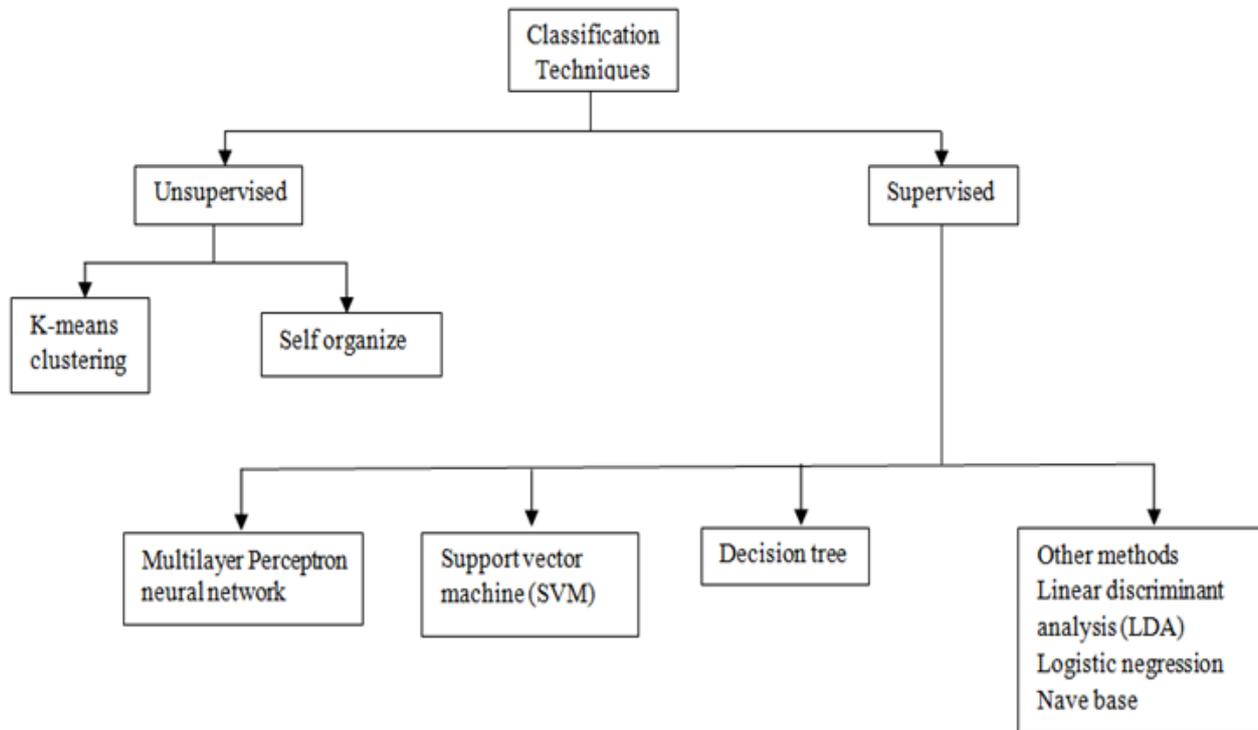


Fig.4. Overview of classification techniques[6]

Method	Description	Advantages	Disadvantages
Multilayer Perceptron Neural Networks (MLPNNs)	MLP is a feed forward artificial neural network model that is used to design recognition systems in order to distinguish specific patterns. An ANN is a sequence of several layers, each layer is composed of a set of neurons. All neurons in a layer are connected by certain weights the previous layer and the next layer neurons. Effectiveness and precision depend upon the number of inputs and number of layers.	It is non-parametric learning and enforce able on noisy input. Able to modeling complex nonlinear and high dimensional problems. Different kernel function can be selected.	Dependence on the number of hidden layer and number of neurons in each layer. Select the type of network is very effective on classification result.
Support Vector Machines (SVMs)	A discriminative classifier to identify hyper planes for binary classes or higher dimensions. Optimal results achieve by finding the maximum distance to the nearest data point of the every class. Lower generalization error of classification can be obtained by a larger margin.	It is a non-parametric learning technique. Effective in the high dimensional classifier. Different kernel function can be specified.	Performance and accuracy highly depend on the selection of kernel parameter. Poor performances with a greater number of features are much rather than the number of samples.
Decision Tree(DT)	DT is a hierarchical decision support tools to partitioning a data set to uniform subsets. It is a recursive function run to select the best attribute to split the data and expand to the leaf nodes.	It is a non-parametric learning method used for classification and regression. Simple to understand and to interpret. Able to handle multi-output problems.	Extremely unstable and dependent the data. Small change in data leads to generate the completely different tree.
Linear Discriminant Analysis (LDA)	LDA is a dimensionality reduction algorithm for classification.	It can deal with the nonlinearity of the distribution of samples. Learns low dimensional representations for classification properly	Unable to find non-linear structure hidden in the high dimensional data. Inherent matrix singularity problem

Table: Classification methods[6]

III. RELATED WORK

Breast cancer screening performed professionally medical examination to check women's breasts for abnormalities such as tumour and cysts while biopsy is the process to identify malignancies. X-rays are used as diagnostic tool in mammography for the examination of human breast. Soft computing techniques are widely used for classification of disease in medical field. Experienced radiologists can do errors while detecting breast cancer in radiographic findings. Thus the computer aided diagnosis demands an enhancement in mammographic images there by increasing the efficiency in detection based on advanced image processing and pattern recognition techniques.

P. Hamsagayathri, P. Sampath [5],proposed a work to analyzes the different decision tree classifier algorithms for Wisconsin original, diagnostic and prognostic dataset using WEKA software. The performance of the classifiers are evaluated against the parameters like accuracy, Kappa statistic, Entropy, RMSE, TP Rate, FP Rate, Precision, Recall, F-Measure, ROC, Specificity, Sensitivity.

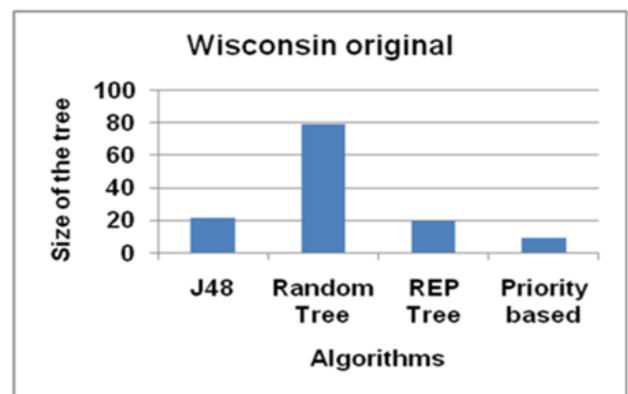


Fig. 5 Decision tree algorithm[5]

Afsaneh Jalalian1,Syamsiah Mashohor[6], Image segmentation is a crucial component in computer vision and pattern recognition. This paper reviewed about different types of segmentation techniques. They are Edge-based segmentation methods, Thresholding-based segmentation methods, Region growing segmentation methods, Clustering-based segmentation methods. Texture features also are another expression of visual characteristics that is convenient for various domain of computer vision and pattern recognition.



Indra kanta maitra and Samir Kumar bandyopadhyay[4], Detection requires Mammography followed by biopsy of the tumour or lesions present in the breast tissue. This paper provides both effective and efficient improvements over existing algorithms and introduces some innovative ideas.

There is a high correlation between high breast parenchymal density and high risk of breast cancer. The proposed system use the suggested density estimation algorithm and classifies the mammograms according to BI-RAD system of classification.

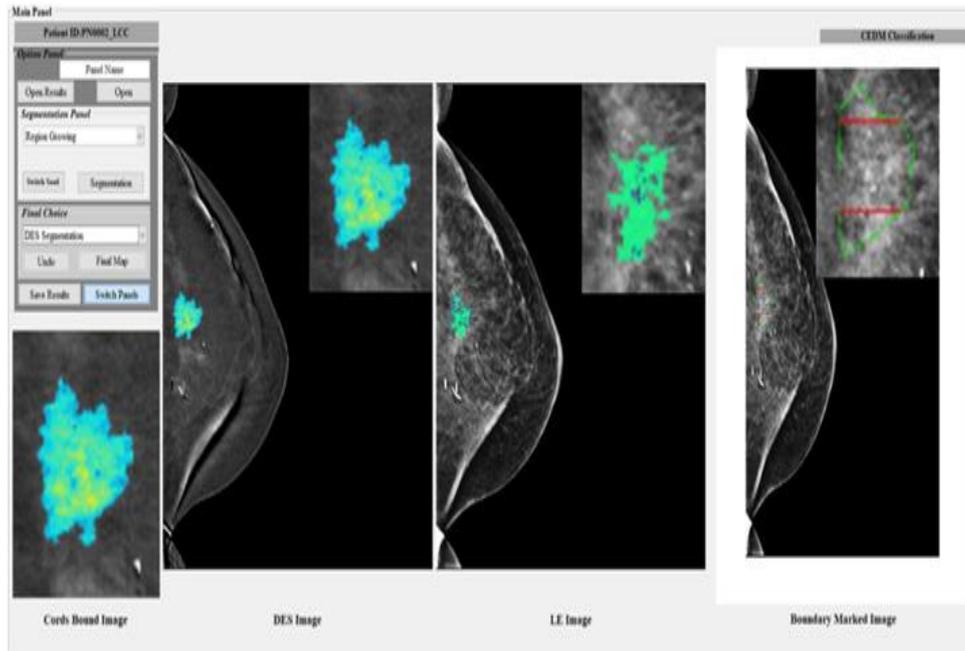


Fig.6. Graphical user interface (GUI) of the CAD scheme[1]

Nadia El Atlas, Mohammed El Aroussi [16], This paper reviewed about different types of segmentation and classification techniques. Segmentation are supervised and unsupervised techniques. This methods of segmentation include a training stage to learn specific areas or lesions to be detected (masses or microcalcifications) in order to construct a model. Discussed about different feature extraction color, texture, shape.

Gopichandh danala, Bhavika patel[1], CEDM imaging examinations were performed using the imaging acquisition protocol. Unique characteristics of using CEDM is that it can overcome effect of tissue overlapping in FFDM and enable detection of tumor's neovascularity related functional information. developing computer-aided diagnosis (CAD) schemes aiming to assist radiologists in their decision-making for better assessing risk of lesion malignancy has been attracting extensive research interest.

Ms.S.Pushpalatha, S.Kayalvizhi[9], imaging techniques that play a vital role in detecting breast cancer. breast cancer is a malignant tumour that begins in the cells of the breast and eventually it spreads to the surrounding tissues. Cancer begins in the cells which are the basic building blocks of the tissues. 3D mammogram and it is an advanced form of imaging where many number of images are captured from various angles and they together form an image set in Breast Tomosynthesis.

Mohamed S. Salamaet al., Radiologists can have a second opinion in the diagnosis. Vigorosity of this disease and the death rate is increasing. So there is a need of early detection of breast cancer to reduce the death rate. So automated computer aided detection is inaveitable. The initial symptom of a breast cancer is the formation of a lump. This is due to

tiny deposits of calcium called microcalcifications and tumors called circumscribed masses. These tumors are generally Benign and malignant. The benign tumors are generally non aggressive and non cancerous. They will not spread to other body parts. The WBC algorithm overcomes the drawbacks of Wavelet Transform (WT) and provides higher accuracy than WT.

S. Boughorbelet al., The mammogram images will be prepared and pre-processed before starting of actual process. The tumoral masses present as thickenings, which appear on images as lesions. These lesions can vary considerably in optical density, shape, position, size and characteristics at the edge. Mammogram are significant for future risk determination of breast cancer development. Focus on highlighting different techniques on enhancement, detection and classification of breast cancer along with its accuracy.

A. Bhardwaj et al., Benign and malignant tumors requires precise and reliable diagnosis to ensure that doctors. Mammography widely used as an effective imaging modality for tumor classification. Localized tumors can be treated successfully before the cancer spreads. Effective prevention and accurate diagnosis of breast tumor is a crucial and important problem in medical science community. A hybrid automatic system for detection and classification of mammographic mass using local seed region.

Gopichandh danala [1] et al., Promising imaging modality in breast cancer diagnosis to use Contrast-enhanced digital mammography (CEDM). Computer-aided diagnosis (CAD) scheme of CEDM images to classify breast masses to investigate how to develop optimally. Two types of

images in CEDM namely, low energy (LE) and dual-energy subtracted (DES) images. A CAD scheme was applied to segment mass regions depicting on LE and DES describe in table separately.

Table.3. Two confusion matrices generated when applying MLP classifier to the originally segmented breast mass regions depicting on LE and DES images.[1]

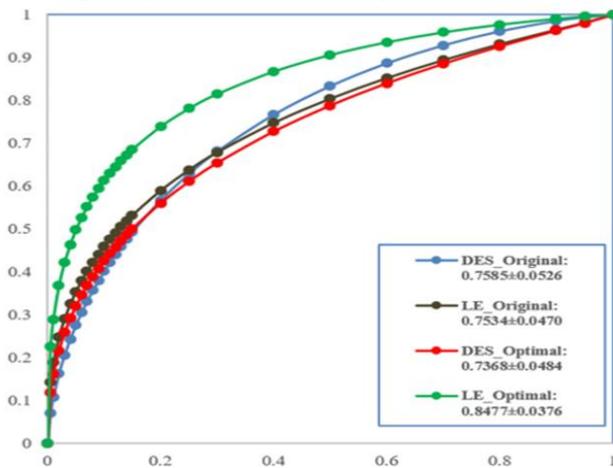


Fig.7. Comparison of four ROC curves[1]

F.Paulinet al., The diagnosis requires precise and reliable diagnosis to ensure that doctors can distinguish between benign and malignant tumors. Detect very small tumors, even before they are tangible or they manifest other symptoms. Increasing sensitivity to rates comparable to those obtained by double reading, in a cost-effective manner to improve performance of radiologists. This paper able to extract useful and important features which can be used in the discrimination between benign and Malignant, implement the feature extraction. Features are Geometricfeature and Texture feature.

IV. PERFORMANCE ANALYSIS & RESULTS

The performance of the classifiers in detecting the breast cancer can be evaluated from the analysis of confusion matrix and below parameters are calculated. Accuracy is the percentage measure of correctly classified instances for all instances. It can be obtained as below

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Where TP - True Positive, TN - True Negative, FP - False Positive, FN - False Negative and these are given by the following equations.

Precision is of correctly classified instances for those instances that are classified as positive, and it is calculated using the equation.

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

Recall is the measure of the positive instance that are correctly classified, and it can be calculated with below

equation

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

F-Measure is the combined metric of precision and recall, i.e., it is harmonic mean of both. It shows how precise the classifier is and also how well the classifier is robust.

F-measure use below equation for calculation

$$F - Measure = \frac{2*Recall*Precision}{Precision+Recall} \tag{4}$$

Sensitivity is the measure of correctly classified positive instances to a total number of positive instances.

$$Sensitivity = \frac{TP}{TP+FN} \tag{5}$$

Specificity is the measure of correctly classified negative instances to a total number of negative instances

$$Specificity = \frac{TN}{TN+FP} \tag{6}$$

Receiver operating curve (ROC) is graphical representation of sensitivity against specificity.

The precision-recall curve is the graphical representation of recall against precision.

V. INFERENCE FROM THE SURVEY

The inference on the survey women's breast cancer is as follows

- Chandresh Arya proposed soft computing approaches are gaining importance in medical disease diagnosis. The data set used for diagnosis purpose is digital mammogram and WBCD (Wisconsin breast cancer data set) is obtained from UCI machine learning repository. The diagnosis is done using different methodology such as statistical, fundamental, probabilistic, neural network, support vector machine, fuzzy systems and hybrid of these techniques. Early detect of breast cancer based on detection of micro calcification and computer based decision system. The paper developed an adaptive k-means clustering algorithm[7].

- Bhardwaj [14], proposed to identify abnormal masses in mammography image. filtered image is then forwarded to median filter in order to further remove the noise and high frequency components without disturbing the edges. The infected region or a region with micro calcifications is segmented region by Markov random field (MRF) hybrid with HPACO algorithm. The diagnostic part of the tool can be used for confirmatory test using histopathological slide of surgical biopsy[4].

- Mohamed S. Salama, to improved technique for feature extraction based on Wavelet Based Contourlet Transform (WBCT) is investigated to obtain the features of the Region of Interest (ROI), allowing for accuracy improvement over other standard approaches. Proposed a hybrid feature selection approach in which the Genetic Algorithm (GA) and the Support Vector Machine (SVM) are combined along with



the Mutual Information (MI) in order to select the best combination of tumor indicators. Enhancement is achieved by intensifying the contrast and edges and by suppressing the noise and background areas[9].

• Gopichandh danala and Bhavika patel[1], computer-aided diagnosis (CAD) scheme of CEDM images to classify breast masses. After computing image features, multilayer perceptron based machine learning classifiers that integrate with a correlation-based feature. Full-field digital mammography (FFDM) and dynamic contrast-enhanced breast magnetic resonance imaging, to enhance the accuracy of extracted feature and to reduce number of attribute by eliminating unnecessary sample so that classification accuracy can be improved[6].

VI. CONCLUSION

In recent years, CAD systems are used to automate breast cancer detections and classification of benign and malignant lesions from image modality digital mammograms. The CAD systems improve radiologists' performance in finding and discriminating between normal and abnormal tissues. The stages of the CAD system and techniques adopted for each stage found in the literature are highlighted. The region-based segmentation and clustering-based algorithms are used to identify breast cancer lesion. Feature extraction and selection is based on the nature of image and image modality. The most important classification techniques used are artificial intelligence and support vector machines in recent years. The imbalance data set is a crucial issue which can be overcome in data and functional level. The future work can be focused on developing an efficient CAD system to diagnose women breast cancer.

REFERENCES

1. Gopichandh danala, Bhavika patel, Faranak aghaei, "Classification of Breast Masses Using a Computer-Aided Diagnosis Scheme of Contrast Enhanced Digital Mammograms", BMES-2018.
2. Asim Ali Khan, Arora AS, "Computer aided diagnosis of breast cancer based on level set segmentation of masses and classification using ensemble classifiers", Biomed Res 2018 Volume 29 Issue 19-2018.
3. Mohamed S. Salama, Ahmed S. Eltrass, and Hassan M. Elkamchouchi, "An Improved Approach for Computer-Aided Diagnosis of Breast Cancer in Digital Mammography", Vol. 8, No. 9, IEEE-2018.
4. Indra kanta maitra and Samir Kumar bandyopadhyay, "CAD Based Method for Detection of Breast Cancer", OJCST, Vol. 11, No. (3)2018.
5. P. Hamsagayathri, P. Sampath, Performance analysis of breast cancer classification using decision tree classifiers, IJCP, Vol 9, Issue 2, 2017.
6. Afsaneh Jalalian, Syamsiah Mashohor, Foundation and methodologies in computer-aided diagnosis systems for breast cancer detection, EXCLI-2017.
7. Vidya Kattapura, Dr. Kurian M Z, "Effectiveness of Existing CAD-Based Research Work towards Screening Breast Cancer", (IJACSA), Vol. 8, No. 9, 2017.
8. Sangeetha R, Dr. Srikanta Murthy K, A Novel Approach for Detection of Breast Cancer at an early stage using Digital Image Processing techniques, (ICISC-2017).
9. [9] Ms.S.Pushpalatha, S.Kayalvizhi, A Survey on Breast Cancer Detection Techniques, IJCSIT, Vol. 7 (6) – 2016.
10. S. Boughorbel, R. Al-Ali, and N. Elkum, "Model Comparison for Breast Cancer Prognosis Based on

- Clinical Data", PLoS One, vol. 11, no. 1, pp.1–15, 2016.
11. Chandresh Arya, Ritu Tiwari, "Expert System for Breast Cancer Diagnosis: A Survey", (ICCCI -2016), Jan. 07 – 09, 2016.
12. S. Boughorbel, R. Al-Ali, and N. Elkum, "Model Comparison for Breast Cancer Prognosis Based on Clinical Data", PLoS One, vol. 11, no. 1, pp.1–15, 2016.
13. Anu Appukkuttan, Sindhu L., "Breast Cancer-Early Detection and Classification Techniques: A Survey", IJCA, Volume 132 – No.11, December 2015.
14. [14]A. Bhardwaj and A. Tiwari, "Breast cancer diagnosis using genetically optimized neural network model", Expert Systems with Applications, vol. 42, no. 10, pp. 4611–4620, 2015.
15. Neeta Jog, Arvind Pandey, "Implementation of Segmentation and Classification Techniques for Mammogram Images", IOSR Journal of Engineering (IOSRJEN)-2015.
16. Nadia El Atlas, Mohammed El Aroussi, Computer-Aided Breast Cancer Detection Using Mammograms: A Review, IEEE-2014.
17. V. Vishrutha and M. Ravishankar, Early Detection and Classification of Breast Cancer, 413– Vol. 1, (FICTA) - 2014.
18. Mussarat Yasmin, Muhammad Sharif and Sajjad Mohsin, "Survey Paper on Diagnosis of Breast Cancer Using Image Processing Techniques", RJRS Vol. 2(10), 88-98, October (2013).
19. Boulehmi Hela, Mahersia Hela, Hamrouni Kamel, "Breast Cancer Detection: A Review On Mammograms Analysis Techniques", March 18-21, 2013.
20. F.Paulin, A.Santhakumaran, "Classification of Breast cancer by comparing Back propagation training algorithms", International Journal on Computer Science and Engineering (IJCSE) – 2011.