Early Detection of Breast Cancer Detection Using SVM Classifier through Mammogram

Kaviya.S, Meena.M, Pavithra.K.M, Prasanna Mariya.L[1], A.Senthil Kumar[2],
[1]*Final Year ECE Students, Sri Shakthi Institute of Engg and Tech, Coimbatore.*
[2] *Assistant Professor (SG), Dept. of ECE, Sri Shakthi Institute of Engg and Tech, Coimbatore*
(E-mail: a.senthilkumar@siet.ac.in)

Abstract— The studies on the Cancer detection and segmentation gives more support to the doctors and radiologist for treatment and identification of cancer in early stages. One of most popular approach used for detecting breast cancer in its early stage is the mammogram screening technique. This method is found to be most cost effective and reliable one for the prediction of breast cancer. Most of hospitals today, adapting this CAD system and take their final decision from the results obtained from mammogram. The main contribution of this project work is to predict the cancer disease earlier based on tissue which is then used to categorize the data sets into multiple classes for the detection of breast cancer disease. Here classification is performed based on the characteristics of tissues such as density, texture of the breast features. It is required to choose the correct machine learning approach for obtaining accurate classification result. In this project by using a SVM classifier which has high performance compared to other classifier. The result is validated along with the normal classifiers..

Keywords— *Cancer detection CAD system; mammogram; breast cancer disease; SVM classifier; Matlab.*

I. INTRODUCTION

A survey, in India, nearly 82,000 new cases were affected by cancer of breast In India, most urban populations are affected by breast cancer. As said by the 2007 and mostly women only affected by this cancer in majority ratio. National Cancer Registry Programme(NCRP) has collected the data over the years from five urban population based cancer registries such as Delhi, Mumbai, Bhopal, Bangalore and Chennai and it has shown the statistical rising of the breast cancer incidence rate. Leading site of breast cancer in hospital-based cancer registries is Thiruvananthapuram and Mumbai. Second leading sites are Dibrugarh, Bangalore and Chennai. So, breast cancer diagnosis is very important to increase life time of people who are affected by begin malignant. Generally, the breast cancer has been caused by the growth of malignant cells in the breast. Breast cancer is the most found disease in the women in the real world environment where presence and growth of malignant cells leads to severe breast cancer. Breast cancer is commonly found in women and some men around the world. In the pathological diagnosis of breast cancer the suspect tissue biopsy or fine needle aspiration is achieved by the pathologist visualization. Mammography detects breast cancer through the identification of abnormalities such as masses and/or micro-calcifications. However, due to the complexity of breast structure, low disease prevalence, and radiologist fatigue, abnormalities are often ignored.

One of most popular approach used for detecting breast cancer in its early stage is the mammogram screening technique. This method is found to be most cost effective and reliable one for the prediction of breast cancer. The primary aim of this work is to develop an automatic breast cancer grading method in X-ray Mammogram images in order to assist pathologists in enhancing the operational efficiency as well as improving diagnostic confidence. This work presents an automatic method for breast cancer grading based on a combination of pixel, semantic-level of features. Although these features are directly computed from images reflecting explicit attributes that pathologists look for grading breast cancers, there is another category of feature generation inspired by convolutional methods. K-means clustering is a segmentation method, used to automated segment from the images. The features were extracted independently and evaluated through Fuzzy Multi-layer support vector machine (FMSVM) classification. In this research, investigation was carried out for classification of breast shape with mammogram images with the first step is the use of Weiner filter to reduce background removal and noise removal. Next, the process will continue with the image segmentation and morphological operation, then characteristic of shape extraction with invariant moment. The last step is classification with Fuzzy-Multi layer support vector machine. This research aims to develop a method that is capable of classifying various features of breast lesions from mammogram images as one of the malignancy parameters of breast cancer.

Mammogram and Breast Regions

A mammogram is an x-ray projection of the 3D structures of the breast. It is obtained by compressing the breast between two plates. Mammograms have an inherent "fuzzy" or diffuse appearance compared with other x-rays or Computed Tomography images. This is due to the superimposition of densities from differing breast tissues, and the differential x-ray attenuation characteristics associated with these various tissues. A mammogram contains two different regions: the exposed breast region and the unexposed air-background (non-breast) region. Background region in a mammogram usually appears as a black region, and it also contains high intensity parts such as bright rectangular labels, opaque markers, and artifacts (e.g. scratches). Breast regions can be partitioned into:

1. Near-skin tissue region, which contains uncompressed fatty tissue, positioned at the periphery of the breast, close to the skin-air interface where the breast is poorly compressed.

2. Fatty region, which is composed of fatty tissue that is positioned next to the uncompressed fatty tissues surrounding the denser region of fibro glandular tissue.

3. Glandular regions, which are composed of non-uniform breast density tissue with heterogeneous texture that surrounds the hyper dense region of the fibro glandular tissue.

4. Hyper dense region, which is represented by high density portions of the fibro glandular tissue, or can be a tumor.

Breast density is a measurement of the dense structure of fibro glandular tissue, which appears white on a mammogram. Fibro glandular tissues appear to have disc or cone shapes and extend through the interior of the breast from the region near the chest wall to the nipple. The breast density part contains ducts, lobular elements, and fibrous connective tissue of the breast.

II. EXISTING SYSTEM

Breast Cancer is the most common cancer in Iranian women. This study aims to demonstrate the characteristics of breast diseases- and especially breast cancer- according to pathologic records in Tehran, Iran. In this cross-sectional study, all records of pathologic specimens (biopsy or mastectomy) categorized as "breast diseases" from 1996 to 2000 in five teaching hospitals in Tehran were studied. For each patient, sex, age, breast pathology, pathological staging of malignant lesions, side and location of the tumor and the type of surgery were reviewed by a trained general practitioner. SPSS version 10 was used for statistical analysis. The mean age of women with breast cancer was 48.8. The highest frequency of malignancies was observed in the 40-49 age groups (31.8%). Twenty-three percent of breast cancers were observed in women younger than 40 years. About 83 percent of malignant lesions in women were in T2, T3 or T4 at diagnosis. Only about 4 percent of women with breast cancers had tumors in stage I or in-situ carcinomas. [1]. Today, the quality of life studies has an important role in health care especially in chronic diseases. Breast cancer has third order among women's malignancies. Now, survival rate for this cancer is long. However breast cancer has several complications that affected the patient's life. The aim of this study was to assess the quality of life in Breast cancer patients under chemotherapy a cross-sectional study conducted on 119 breast cancer patients that were admitted and treated in chemotherapy ward of Namazi hospital in Shiraz city, south of Iran, between Jan and Feb 2006.The QLQ-C30 questionnaire was used to assess quality of life in these patients we used univariate methods. [2]. usually photo texture analysis is used for clustering and classification primarily based on content of picture. This paper concentrated on Fuzzy-Multi layer SVM (FMSVM) classifier for evaluating the features extracted and to determine its effects. The proposed FMSVM version indicates promising consequences when compared with different classifiers used most generally within the literature and can a destiny for more sophisticated statistical features based most cancers prognostic models. The proposed method is evaluated on a set composed of images extracted from Mini MIAS databases. [3].

Studies based on density have been undertaken because of the relationship between breast cancer and density. Breast cancer usually occurs in the fibro glandular area of breast tissue, which appears bright on mammograms and is described as breast density. Most of the studies are focused on the classification methods for glandular tissue detection. Others highlighted on the segmentation methods for fibro glandular tissue, while few researchers performed segmentation of the breast anatomical regions based on density. There have also been works on the segmentation of other specific parts of breast regions such as either detection of nipple position, skin-air interface or pectoral muscles. [4]. A systematic method for the detection and segmentation of micro calcifications in mammograms is presented. It is important to preserve size and shape of the individual calcifications as exactly as possible. A reliable diagnosis requires both rates of false positives as well as false negatives to be extremely low. The proposed approach uses a two-stage algorithm for spot detection and shape extraction. The first stage applies a weighted difference of Gaussians filter for the noise-invariant and size specific detection of spots. A morphological filter reproduces the shape of the spots. The results of both filters are combined with a conditional thickening operation. [5]. a novel approach for segmentation of suspicious mass regions in digitized mammograms using a new adaptive density-weighted contrast enhancement (DWCE) filter in conjunction with Palladian-Gaussian (LG) edge detection. The DWCE enhances structures within the digitized mammogram so that a simple edge detection algorithm can be used to define the boundaries of the objects. Once the object boundaries are known, morphological features are extracted and used by a classification algorithm to differentiate regions within the image. [6]. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was S. Pizer.H.P. Chan and B. Sahiner are with the Department of Radiology, University of Michigan, and Ann Arbor, MI 48109-0030 USA. D. Wei is with the Department of Radiology, University of Chicago, and Chicago, IL 60637 USA. However, a well-trained computer program (which can screen a large volume of mammograms accurately and reproducibly) is needed in order for CAD to become practical in clinical settings. Such a program has yet to be developed. [7].

III. PROPOSED SYSTEM

The proposed research method attempts to extract and analyze the malignant tissues from the digital mammogram images. In the proposed research method, raw digital mammogram images are taken as input image which is analyzed in order to final abnormalities from the mammogram images. The proposed research method perform breast cancer detection in three stages namely preparation, pre-processing and statistical decision making. The processing flow of the proposed research method is illustrated in the figure 1.

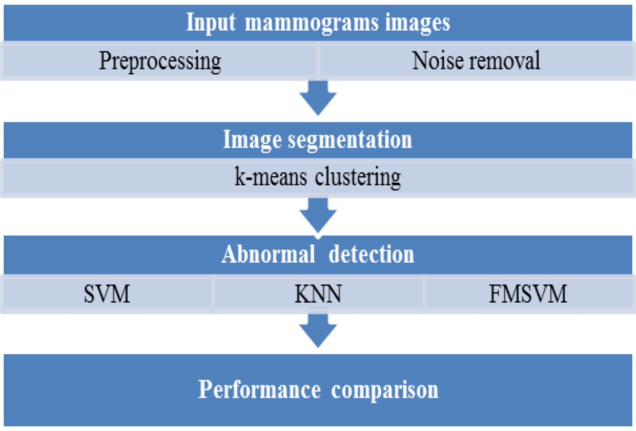

Fig 1: A Schematic Diagram of Proposed Work

Mammography database used in this work is taken from Mammographic Image Analysis Society (MIAS) which an organization of UK research groups. Films taken from the UK National Breast Screening Programme have been digitized to 50 micron pixel edge with a Joyce-Loebl scanning microdensitometer, a device linear in the optical density range 0-3.2 and representing each pixel with an 8-bit word. The database contains 322 digitized films and is available on 2.3GB 8mm (ExaByte) tape. It also includes radiologist's "truth"-markings on the locations of any abnormalities that may be present. The database has been reduced to a 200 micron pixel edge and padded/clipped so that all the images are 1024x1024. Mammographic images are available via the Pilot European Image Processing Archive (PEIPA) at the University of Essex.

A. Data Processing Techniques

This entire approach in implemented and evaluated its performance improvement over Mini Mammographic Imaging Analysis Society (Mini-MIAS) database. The steps involved in the proposed technique are given as follows: Removing the noises, tags and unwanted contents from the input breast x ray image, so that better classification can be done. Dividing and extracting the required mammogram part from the breast cancer based on region of interests which can be analyzed well for the accurate detection breast cancer detection. In this research work, k means clustering algorithm is used for the segmentation process. From the segmented breast images, features would be extracted which is then classified in this stage for the identification of the breast cancer occurrence..

B. Ultrasonic Sensors

In the proposed research method, input mammogram images are preprocessed initially in order to improve the image quality, so that the segmentation results would come more accurate. The phases involved in the post processing stage are, cropping, orientation, artifact removal and denoising. Cropping is a process of separating the non-breast back ground from the image to retrieve the breast part alone to get accurate result. And then orientation is performed to get correct projection shape. Most commonly used mammographic projections are medic-lateral (MLO) and cranio-caudel (CC). However MLO is proved to provide better result than the CC due to its horizontal orientation. It is required to introduce the standard algorithm for the converting the image due to its heterogeneity property. For example, right side of breast required to rotated 1800 horizontally to convert the image homogeneous. After orientation, artifact removal is performed which is a most complex task. Artifacts make disease detection process more complex due to its characteristics of making more intensity on images. In this research work, artefacts removal is done by converting the non-breast image part into black ground color. In this research work, wiener filter is applied to remove the sale and pepper noise before performing preprocessing.

Most important feature of the mammograms is a breast contour which is also called as skin-air interface, breast boundary and pectoral muscle. These can be performed by segmenting the breast mammogram image into non breast image part, and pectoral muscle. From this separation region of interest separation can be performed successfully. Here post processed mammogram image would be given as input to the preprocessing stage which will output region of interest of anatomical image of breast image with abnormal regions. The phases involved in the preprocessing stage are image enhancement, edge detection, and isolation, suppression of pectoral muscle, contour determination and anatomical segmentation. After preprocessing, region of interest of mammogram image would be obtained where breast cancer can be found accurately. Edge detection of mammogram images are highly depends on the mammogram image enhancement outcome due to homogeneity outcome. Both image enhancement and edge detection is done by using dynamic adaptive threshold called Maximum Distance Threshold (MDT). In order to get accurate classification results, factor affecting the accurate feature extraction such as pectoral muscle presence, non-breast region needs to be avoided. This can be done by isolating the pectoral muscles and suppressing it before extracting region of interest. This is done edge based pectoral muscle isolation and suppression algorithm.

Breast contour is the most primary feature present in the breast mammogram images. Location and orientation of breast contour in breast mammogram image make ease of detection of abnormalities present in the breast mammogram. In authors have introduced breast contour detection method based on edge detection technique to find the edge line of mammogram image. The outcome of this algorithm would provide smooth, single pixel and continuous image.

The pectoral muscle area detection and elimination is done by using the anatomical segmentation algorithm which is based on extraction ROI of breast images. The main idea of this method is to differentiate the different anatomical regions and find the ROI boundary lines. The idea behind this follows is the presence of different color variation on normal and abnormal tissues. And also remaining parts such as fat, glands, connective tissue and calcium deposits would also appear in different color shades. The proposed algorithm attempts to separate these different color regions present in the mammogram image by eliminating the unwanted edges, lines and dots from the given input processed image. For example, the preprocessed image is shown in figure 2

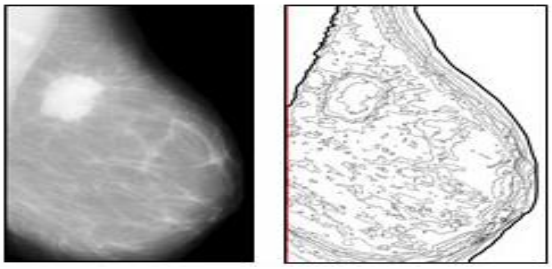

Fig 2: (a) The Prepared Mammogram and (b) The Preprocessed Mammogram.

IV. SOFTWARE REQUIREMENTS

A. *Wiener Filtering*

This filter is very effective for noise removal compared than other filtering and is used by many researchers for detection of breast cancer. It can be reduced the Mean Square Error (MSE) for salt and pepper noise removal.

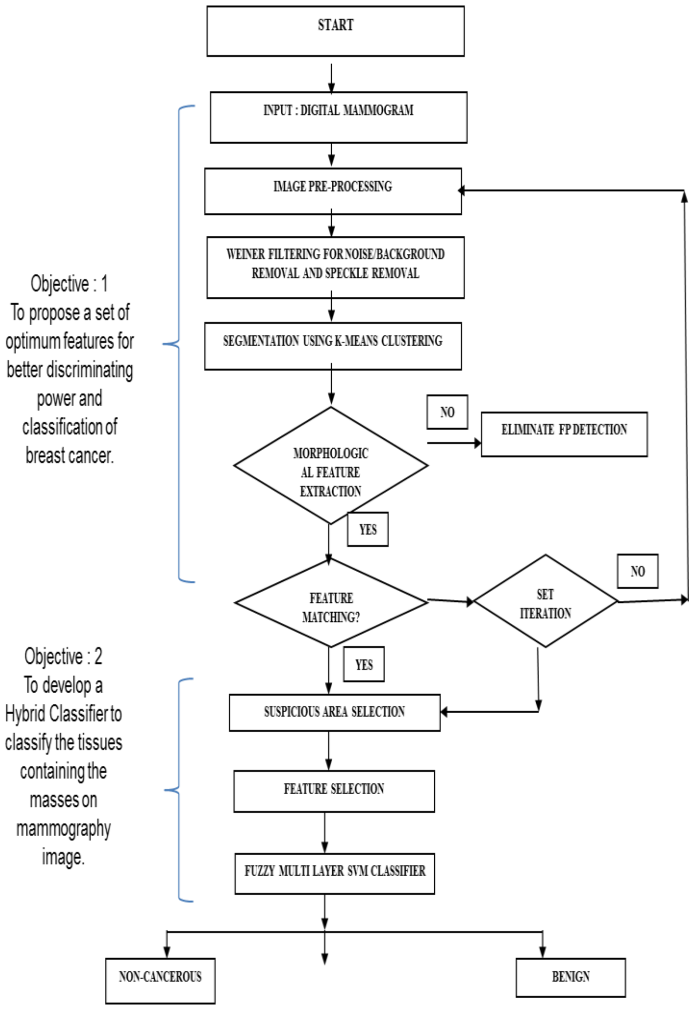

Fig 3: A flowchart of Proposed Work

B. *Segmentation By Using K-Means Clustering*

Image segmentation is the process of dividing the images into multiple segments which can be most useful in the classification task by increasing accuracy and sensitivity rate. In the proposed research work, adaptive k means clustering algorithm is used to perform segmentation of digital mammograms to split the disease affected region and non-affected regions. Adaptive k means algorithm is the improved version of k means algorithm where the certain parameter values are varied considerably. K means algorithm is an most popular iterative approach which is used to group the given input data into k number of clusters. K means algorithm is used to cluster the given number of data points into k number of clusters by assigning each data into each cluster based on its similarity. The basic flow of algorithm is given as follows: Randomly take k cluster centers. Perform clustering by assigning image pixel into each cluster which is more similar with cluster center. Update the cluster center by calculating average of pixels assigned in the corresponding cluster and repeat this process until convergence rate is attained. (e.g. no pixels change clusters. Given a set of observations (x_1 ,x_2… x_n), where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k sets ($k < n$) $S = \{S_1, S_2 \ldots S_k\}$ so as to minimize the Within-Cluster Sum of Squares (WCSS).

C. *Feature Extraction*

After segmentation, the feature extraction has been done. It is the process of representing the input image into set of features. Features are the one which is used to represent the characteristics or property of particular image. Relevant features that are extracted from the images would provide the complete information about the images. In this research work, texture features are considered, which can differentiate the normal and abnormal tissues accurately. Texture is variation present in the surface of the images in terms of color, shape, orientation and so on. It can define as spatial property of the gray scale images. Texture features are categorized into two types's namely first order and second order. In first order type, texture measure value is extracted from individual pixels.

D. *Multilayer - Support Vector Machine*

In the multi-layer SVM classifier, the architecture contains multiple sup-port vector classifiers in the output layer. To deal with multiple classes, we use a binary one vs. all classifier Mc for each class c. We use a classification dataset for each classifier Mc:{(x1, yc1), . . . , (xℓ, ycℓ)}. All classifiers Mc share the same hidden-layer of regression SVMs. Mc determines its output on an example x as follows:gc(f(x|θ))=ℓXi=1yciαciK2(f(xi|θ),f(x|θ))+bc. The target outputs for hidden-layer features are again kept between -1 and 1.The datasets for hidden-layer SVMs are made so that the sum of the dual-objective functions of the output SVMs is minimized. All SVMs are trained with the gradient ascent algorithm on their constructed datasets. For the ML-SVM

ISSN: 2393-9028 (Print) | ISSN: 2348-2281 (Online)

classifier, we use a different initialization procedure for the hidden-layer SVMs. Suppose there are d hidden-layer SVM and a total classes. The first hidden-layer SVM is first pre-trained on inputs and perturbed target outputs for class 0, the second on the perturbed target outputs for class 1, and the kth hidden-layer SVM is pre - trained on the perturbed target outputs for class k. The bias values are computed in a similar way as in the regression ML-SVM, but for the output SVMs only examples with non-bound support vector coefficients (which are not 0 or C) are used.

E. Fuzzy Logic

Fuzzy logic is all about the relative importance of precision. Fuzzy logic is a fascinating area of research because it does a good job of trading off between significance and precision something that humans have been managing for a very long time. Among various combinations of methodologies in soft computing, the one that has highest visibility at this juncture is that of fuzzy logic and neurocomputing, leading to neuron-fuzzy systems. We can get our specified needs using this. Within fuzzy logic, such systems play a particularly important role in the induction of rules from observations. Segmentation is the process of dividing the images into multiple dividends, so that specific actions carried out over on those segmented images. In this research work, segmentation is used to divide and separate the mammogram part of breast image, so that breast cancer analysis can be carried out efficiently. According to M. Mustra and M. Grgic, 2009, in order to upgrade segmented mammogram image in to process able one, following actions are carried over such as eliminating the noise, labels, markers, and other threats. The next step after segmentation of the mammogram from the breast is to remove the pectoral muscle present in the breast region, so that accurate processing can be taken over.

V. EXPERIMENTAL RESULTS

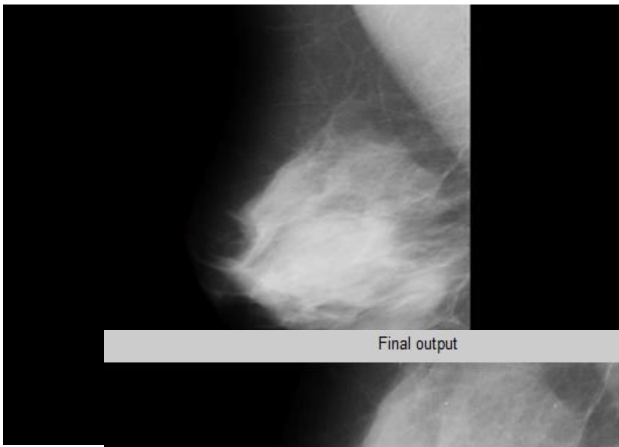

Fig 4; Input Mammography Breast Image

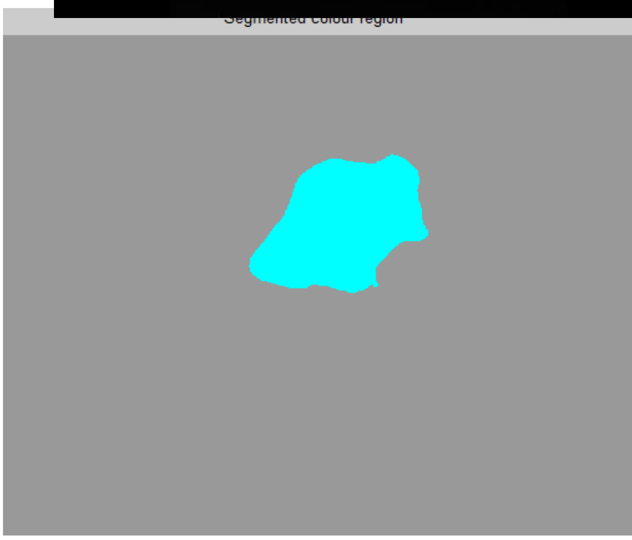

Fig 5; Noise Removed Mammography Breast Image

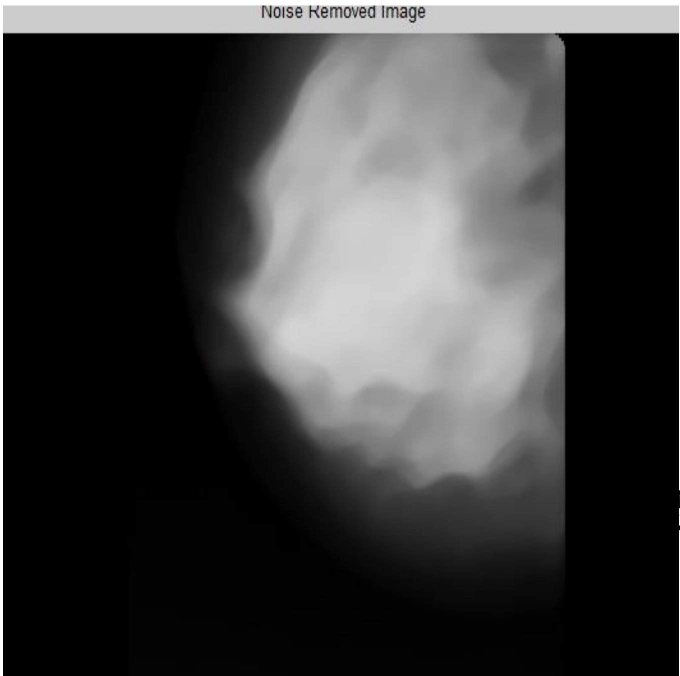

Fig 6; Segmented Cancer in Image

Fig 7: Cancer spot in mammography breast image

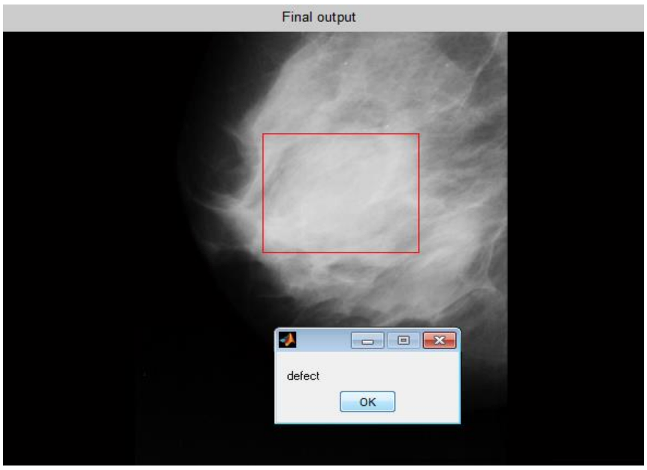

Fig 8; Classification of Cancer in Mammography Breast Image

These images are represented and focused in MLO view. The primary purpose of this comparison evaluation is finding the mass intensities of images than the existing approaches. Many research works doesn't show accuracy improvement value than the proposed research algorithms. However some researchers had shown their performance results in terms of accuracy estimation. Varying parameters that are considered for the performance evaluation of the proposed research work are accuracy, sensitivity, and specificity.

VI. CONCLUSION AND FUTURE SCOPE

Earlier we can predict the cancer region only with the help of radiologists.Inorder to predict the cancer region in the breast part without the help of radiologist; we have designed the mammography tool to view the cancer. In this project we have used SVM classifier to detect the cancer part accurately at the very earlier stage. The future scope would be more advancement in detecting the cancerous region. Here we have made calculations in terms of accuracy to get accurate result of the cancerous region. With the help of accurate result doctor can proceed further steps without the help of radiologist .This project would be more scoopful in rural areas.

ACKNOWLEDGMENT

The authors wish to thank Head of the department ECE and management of Sri Shakthi Institute of Engineering and Technology Coimbatore for his kind support throughout completion of this project.

REFERENCES

[1]. Sarfraz Ahmed, “Global cancer statistics,” CA Cancer Journal for Clinicians, vol. 61, no. 2, pp. 69–90, 2011.

[2] K.Subashini, K.Jeyanthi, ‘Masses detection and classification in ultrasound images’, IOSR Journal of Pharmacy and Biological Sciences (IOSR-JPBS),Volume 9, Issue 3 Ver. II (May -Jun. 2014), PP 48-51

[3] Mohamed Meselhy Eltoukhy, Ibrahima Faye1, Brahim Belhaouari Samir, ‘Breast cancer diagnosis in digital mammogram using multiscale curvelet transform’, Computerized Medical Imaging and Graphics 34 (2010) 269–276.

[4] Y.Ireaneus Anna Rejani, Dr.S.Thamarai Selvi Noorul, “Early Detection Of Breast Cancer Using SVM Classifier Technique”, International Journal on Computer Science and Engineering Vol.1(3), pp. 127-130, 2009.

[5] M. Mustra and M. Grgic, “Robust automatic breast and pectoral muscle segmentation from scanned mammograms,” Signal processing, vol. 93, no. 10, pp. 2817–2827, 2013.

[6] J. Ojo, T. Adepoju, E. Omdiora, O. Olabiyisi, and O. Bello, “Preprocessing method for extraction of pectoral muscle and removal of artefacts in mammogram,” IOSR Journal of Computer Engineering, vol. 16, no. 3, pp. 06–09, 2014.

[7] R. Nithya and B. Santhi, “Comparative study on feature extraction method for breast cancer classification,” Journal of Theoretical and Applied Information Technology, vol. 33, no. 2, pp. 1992–86, 2011.

[8] S.-C. Tai, Z.-S. Chen and W.-T. Tsai, “An automatic mass detection system in mammograms based on complex texture features,” IEEE journal of biomedical and health informatics, vol. 18, no. 2, pp. 618– 627, 2014.

[9] A. AlQoud and M. A. Jaffar, “Hybrid Gabor based local binary patterns texture features for classification of breast mammograms,” International Journal of Computer Science and Network Security (IJCSNS), vol. 16, no. 4, p. 16, 2016.

[10] Wenzhong Shi, Kimfung Liu, and Changqing Huang, "A FuzzyTopology-Based Area Object Extraction Method" IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 48, NO. 1, JANUARY 2010.

[11] C. K. Wong, "Fuzzy points and local properties of fuzzy topology," J. Math. Anal. Appl., vol. 46, no. 2, pp. 316-328, May 1974.

[12] C. L. Chang, "Fuzzy topological spaces," J. Math. Anal. Appl., vol. 24, no. 1, pp. 182-190, Oct. 1968.

[13] Y. M. Liu and M. K. Luo, Fuzzy Topology. Singapore: World Scientific, 1997.