Identification of Micro-calcification in Mammogram for Breast Cancer Analysis using SVM Classifier

Gaurav Makwana¹,², Lalita Gupta²

¹ Department of Electronics Engineering, SVITS, SVVV, Indore, India
²Department of Electronics & Communication Engineering, MANIT, Bhopal, India
*E-mail gauravmakwana@gmail.com:

Abstract

Breast cancer is most common disease in women of all ages. To identify & confirm the state of tumor in breast cancer diagnosis, patients are undergo biopsy number of times to identify malignancy. Early detection of cancer can save the patient. In this paper a novel approach for automatic segmentation & classification of breast calcification is proposed. The diagnostic test technique for detection of breast condition is very costly & requires human expertise whereas proposed method can help in automatically identifying the disease by comparing the data with the standard database. In proposed method a database has been created to define various stage of breast calcification & testing images are pre-processed to resize, enhance & filtered to remove background noise. Clustering is performed by using k-means clustering algorithm. GLCM is used to extract out statistical feature like area, mean, variance, standard deviation, homogeneity, skewness etc. to classify the state of tumor. SVM classifier is used for the classification using extracted feature.

Keywords: Breast mammogram; micro-calcification; benign; malignant; K-means; GLCM; SVM

1. Introduction

Medical imaging technology helps doctors to see the entire portions of the body for easy diagnosis. Various Imaging techniques like CT scan ultrasound MRI, X-ray etc. helps in 3-D diagnosis of various body parts such that sever surgeries can be performed with ease without opening too much of the body. Image processing technique can help in analyzing the data of various medical imaging systems. By various image processing techniques data can be modified to analyze symptoms of the patient with ease.

Approximately more than 1300 persons detected cancer every day. Cancer has become major cause of death in India due to unhealthy lifestyle, food, environmental condition and family past medical history. During 2012 to 2014 Mortality Rate due to cancer is approximately six percent in India as per the record of National Cancer Registry Program of ICMR. As per data of ICMR 28,20,179 cases detected in year 2014 in which approximately 4,91,998 people died, while in 2013 29,34,314 cases detected and 4,78,180 died and in 2012 30,16,628 cases detected and approximately 4,65,169 people died due to cancer.

Breast cancer is the commonest cancer of urban Indian woman and the second commonest in the rural woman. There is no cancer screening program available in India and due to lack of high quality infrastructure, skill and lack of awareness of this disease major-ity of breast cancer are diagnosed at a relatively advance stage. For breast cancer care in India early diagnosis of cancer and more facilities for cancer treatment in low cost is required. According to International Agency of Research on Cancer and WHO study says that one woman dies by cervical cancer every 8 minute and in India approximately 50 percent women with newly diagnose breast cancer dies. [1,2]. In year 2012 approximately 70218 wom-en died out of 144,937 cases whereas in year 2015 approximately 75,957 women died out of 1,55,863 newly diagnose cases of breast cancer. [1,2].

2. Literature Review

Mammography is used to examine the human breast. It is used for early detection of breast cancer. In detection and diagnosis of breast cancer, first identify the suspicious area in mammogram. Experience of radiologist helps him to know what information to look for and how to interpret that information on the basis of the knowledge from the previous experience with same type of image but it is not always true radiologist face problem in detection of abnormalities like calcification, masses etc in mammogram. Computer aided diagnosis can help the radiologist to identify these problem.

There are two type of mammography: Screening and diagnostic mammography. A patient with already found abnormality like breast lump is recommended diagnostic mammography. Diagnostic mammography is performed as follow-up to abnormal screening mammography. Based on the diagnostic mammography result, additional diagnosis like breast imaging or biopsy [17] is recommended.

Positron Emission Mammography (PEM), Digital Breast Tomo-synthesis (DBT), Ultrasound and Magnetic resonance Imaging (MRI) are used for breast cancer diagnosis. PEM scan is mostly used to detect and localize breast cancer and determine the extent of spread of cancer and it’s response to therapy. DBT is an emerging technique used in detection and characterization of breast lesion especially in women with non-fatty breasts, Digital Tomosynthesis uses X-ray to create three- dimensional picture of breast.[12] Ultrasound is high frequency sound waves which can...
help doctor to get images of the inside of the breasts. Only drawback of Ultrasound imaging is lacks the resolution and contrast of mammography. For woman under age 30, ultrasound is sufficient to determine whether particular region required biopsy or not but for woman older than 30 year mammogram and ultrasound used together for the diagnosis. MRI of the breast offers valuable information about breast condition that can’t be obtained by other image modalities, such as mammography or ultrasound [13]. Dynamic Contrast Enhanced MRI (DCE-MRI) can be used for women with dense breast. DCE-MRI techniques are based on injection of an MR contrast agent and acquisition of T1 weighed image over time which provides information on the rate of passage of the agent between the blood and tissue.

MRI and ultrasound imaging techniques are used to get particular analysis but above two procedure don’t guarantee the identification of malignant or benign so biopsy required for exact identification. [5]. Cluster of mammogram of micro-calcification in mammographic image indicates the possibility of breast cancer. Different shape, size and distribution of micro-calcification are used by radiologist for identification of malignant or benign. Malignant are generally small dot-like, variable in size, shape and density whereas benign are larger, more rounded, smaller and homogeneous in shape [6].

As the structure of breast is very complex therefore very reliable image processing algorithm is required for detection and diagnosis of breast characteristics. Computer aided tools are very helpful which can compare the data from radiologist reading [7, 8]. Contrast enhancement method is used to increase the contrast of micro-calcification over a threshold [9] but some regions may be under enhanced or some may be over enhanced in both condition information will lost [10]. Taurani et al. [11] used template matching technique where each region of interest of database served as a template and mutual information was used as a metric of similarity to decide it’s region of interest contained a mass, moreover many cancerous lesion occur with different length, width and density etc. which can be find using CAD.

Another important image processing tool is image segmentation. Image segmentation changes the representation of the image in more meaningful way. There are many algorithm suggested for image segmentation. Local raw pixel information and Edge based method are some thresholding technique used for decision making is centered on contour. To identify basic shape of an image thresholding and edge detection techniques are very helpful. For analyzing images these methods comes prior to feature extraction and image recognition system [14]. Another method known as histogram thresholding [15] in which first find the histograms of two equally divided halves of the image and then histograms are compared to detect the tumor and after detecting the tumor site it is important to find the appropriate physical dimension, image cropping method can be used to get the region of interest. Region based technique is another method in which similar type of pixels are partitioned and organized to find the similarity with in image. Spatial gray level dependence method extract optimal texture features from normal and tumor regions [16].

As Mammography have low contrast which makes it difficult for radiologist to interpret the result [18], to avoid this situation second opinion (radiologist) is recommended but this is not cost effective method so CAD can help radiologist in interpretation of mammogram. Mortao et al. [19] and Brem et al. [20] proposed CAD for interpretation of mammograms with increased sensitivity of 7.62 % and 21.2 % respectively.

Machine learning process is used to determine the presence of micro-calcification at pixel location. Peng et al. [21] proposed evolutionary-genetic algorithm to identify the bright spot in mammogram. SVM is another machine learning approach which achieves high accuracy in MC detection. Sampat and Bovik [22] use filter algorithm to enhance breast image and speculation filter to detect location. Companini et al. [23] proposed SVM based featureless approach. Author used wavelet representation to identify image and two SVM classifier: first SVM classifier was used to find the mass and second to reduce the false positive. DDSM database was used to verify the algorithm which achieved approximately 80% positive result. Eltousy et al. [24] proposed algorithm in which ROI were pre-processed by segmentation and granulation technique. The knowledge based learning is used to detect the suspicious area. Same half DDSM database was used for learning and other half was used for testing final test result was 92%, 88% and 81%. Bellotti et al. [25] implemented edge based segmentation algorithm to select the suspicious region and then GLCM matrix used to identify the feature of ROI, finally classification of masses is achieve by neural network which was based on gradient descent learning rule. A database of 3369 mammogram are used with 2307 and 1062 are negative and positive cases respectively. This approach achieves 80% accuracy of mass detection.

This paper proposes a system to identify the presence of micro-calcification in digital mammogram and state of tumor. Section 2 presents the review of proposed technique for mammographic feature analysis and classification. Section 3 describes the structure of the proposed system. Section 4 represent the result associated with proposed method. Section 5 includes concluding remarks.

### 3. System Description

![Fig. 1 Block Diagram Representation of Proposed method](image)

As can be seen from above discussion there are many different method proposed to identify the micro-calcification but accuracy is big concern. In proposed method our primarily focus is on the finding method for identifying the state of tumors in the breast tissues. The main aim of the pre-processing is to resize, contrast enhancement & remove noise from the image. The Otsu thresholding method applied to find the foreground pixels of concern. It minimizes the interclass variance of the thresholded black and white pixels. Figure 2 shows the mammogram image and image after thresholding pectoral muscles appears brighter. There are possibility of false positive due to image noise, blood vessels and tissues which can be reduced by using by median filtering, whereas segmentation is used to get multiple segment of an image based on the property of image pixel. Image clustering is used to identify natural groupings of data from a large data set to produce a concise representation of image behaviour. The exposed breast region is divided in Region of interest (ROI). K-means clustering is a partitioning method which is used to get clusters of image. It partitions data into k mutually exclusive clusters, and returns the index of the cluster to which it has assigned each observation. For the image analysis of mammogram the measure of region is based on a set of statistical texture descriptor. GLCM based texture descriptor is used to get pair-wise spatial co-occurrence of

---

**Input Image** → **Pre-processing** → **Segmentation Method** → **K-means Algorithm** → **Feature Extraction (GLCM)** → **SVM Classifier**
pixels separated by the relative distance \(d\) between pixel pair and their relative orientation. Variation in the pixel intensity has calculated by the texture feature obtained by GLCM texture descriptor. Support vector machine (SVM) is a supervised learning method. This method constructs a set of hyperplanes in a high dimensional space, which is use for classification. SVM classifier can apply for object based image analysis. GLCM based texture descriptor used as interface between mammogram image and the classifier. This interface provides training data to the classifier. Micro-calcification usually appears brighter than their surroundings. Major problem with many algorithms is that identification of micro-calcification is difficult because contrast of dense breast is very low.

The aim of this work is to create and implement an algorithm that accepts a raw mammographic image and mark any possible mass regions that may appear in the image based on intensity, shape, size and other texture measures. This pixel information can be feed into the machine so ROI related with that particular pixel can be finding out. Computer aided system include steps for lesion segmentation feature extraction, feature selection and classification.

4. Dataset Description

Real mammogram data base is required to have exact analysis because computer aided diagnostic depend upon the feature of the real database, that is why in proposed method of extraction have been applied to the ROI in mammogram from the mammogram image analysis society's (MIAS) database and the Digital Database for Screening Mammography (DDSM) for screening mammography. This database contain all type of abnormalities in breast like ASYM, CALC, CIRC, MISC, SPIC (categorize in dense benign, dense malignant, fatty benign, fatty glandular benign, fatty glandular malignant, fatty malignant) and Dense glandular normal, fatty glandular normal, fatty normal this database. In DDSM database each abnormality is outlined by radiologist whom we used for detection of lesion. A Set of 329 images are used to create training dataset and 70 images are used to check the proposed algorithm.

5. Result

Calcification that are irregular in shape, size fall closer to malignant their boundary profile is not self-similar. The proposed method of classification has been applied to region of interest. The mammogram data has been taken in which abnormalities were outlined by the radiologist. These images are used for detection of spatial location of lesion. By using GLCM method we have found the statistical parameter of the training data set which is used to compare the test image. SVM classifier is used to classify the test sample by using feature extracted data.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Category</th>
<th>Number of images used for training</th>
<th>Number of Images used for testing</th>
<th>Detection accuracy K-means + SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Benign</td>
<td>68</td>
<td>20</td>
<td>90%</td>
</tr>
<tr>
<td>2</td>
<td>Malignant</td>
<td>52</td>
<td>20</td>
<td>85%</td>
</tr>
<tr>
<td>3</td>
<td>Normal Breast</td>
<td>209</td>
<td>30</td>
<td>90%</td>
</tr>
</tbody>
</table>

6. Conclusion

The major challenge in Medical images analysis is that we cannot identify which types of tumors and masses are located inside the organ of the body. So with the help of different kind of processing techniques like: segmentation, enhancement, and classification, we can easily find out the tumor type for instance. In the breast cancer the major challenge is to identify the tumor either it’s malignant or benign. Image processing provides access to a wide variety of function for reading, writing, and filtering images of various kinds in different ways. The proposed algorithm showed great success in identifying
the region of interest and correctly segmenting the input test image. We demonstrated the support vector machine approach for image classification which offers significant flexibility in the design of CAD system for breast cancer diagnosis. Proposed method provides a very high level of robustness. The results gave better accuracy for each breast cancer stage. The resulting diagnosis showed great promise for being an invaluable and dependable tool for the diagnosis of breast cancer.

References

[1] WHO Summary report on HPV & cervical cancer statistics in India (18/03/2008)


[15] Yu Guang Zhang, Wen Lu, Fu Yun Cheng, Li Song “A Edge Detection Method for microcalcification Clusters in Mammograms”, Science and Technology Development Project of Taishan Medical University (No.1240)


[26] JawadNagi, “The Application Of Image Processing And Machine Learning Techniques For Detection And Classification Of Cancerous Tissues In Digital Mammograms” Faculty Of Computer Science and Information Technology University Of Malaya Kuala Lumpur,


