



Mass Detection and Classification System for Mammography Image Preprocessing

S. Thilagavathi

Ph.D Research Scholar,
Research & development centre
Bharathiar university, Coimbatore

Dr. M. S. Irfan Ahmed

Research Supervisor,
Research & development centre
Bharathiar university, Coimbatore

ABSTRACT

Today in the entire world in numerous territories, Breast malignancy is a main source of disease especially in lady in which most normal bosom anomalies are masses and calcifications. Early location of determination is the way to bosom malignancy controls that expansion the accomplishment of treatment, spare lives and diminish cost. It is troublesome for the radiologist to perceive the majority on a mammogram since they are encompassed by muddled tissues, thus numerous frameworks have been produced to help the radiologist in distinguishing mammography injuries that may demonstrate the nearness of bosom disease. The exploratory outcomes demonstrate that the proposed CAD framework enormously enhances the five target records in correlation with mass recognition and characterizations framework for mammography picture preprocessing.

Keywords: *Preprocessing, Mammography, SVM, KNN*

1. INTRODUCTION

The term “breast cancer” refers to a malignant tumor that has developed from cells in the breast. Breast Cancer that forms in tissues of the breast, usually the ducts (tubes that carry milk to the nipple) and lobules (glands that make milk). Breast cancer is a leading cause of death among women in developed countries. The morbidity of breast cancer is increasing with a fast speed in developing countries due to the increase of life expectancy, urbanization and change in life styles. According to Breast Cancer Statistics about 40,450 women in the U.S. are expected to die in 2016 from breast cancer. As the cause of breast cancer is not

clearly known, early detection remains the corner stone in breast cancer treatment.

Breast cancer can be detected through various examinations as magnetic resonance imaging (MRI), mammography, ultrasound, CSE and BSE. Mammography is the most effective in reducing mortality rates by 30% - 70%. The reasons for the high miss rate and low specificity in mammography are: (1) the low conspicuity of mammographic lesions; (2) the noisy nature of the images; (3) the overlying and underlying structures that obscure features of the ultrasound image. Also, the biopsies are expensive and involve minor risks. In order to avoid unnecessary biopsies, the number of false positives in mammography has to be reduced. Before feature extraction and classification, the input mammogram image is pre-processed as shown in figure 1 in our method 3 steps are carried out in pre-processing. The first step is to convert the RGB image into grayscale image because RGB image takes more processing time. In second step input image is resized to standard size using resize function. Then input mammogram image filtered to remove unwanted noise. Mammograms are medical images that are difficult to interpret, thus a preprocessing phase is needed in order to improve the image quality and make the segmentation results more accurate.

2. Literature Review

U.S.Ragupathy, T.Saranya et.al proposed that the new method for improving detection of architectural distortion and mass in mammographic images using

Gabor wavelets and Adaptive Neuro-Fuzzy based classification. The suitable features are selected from the set of extracted features and given as input to ANFIS for classification. **S.N. Deepa and B. Aruna Devi et.al** proposed that the author has explained Computer-Aided Diagnosis for the medical prognosis. Artificial Neural Network forms the base of the intelligent systems. There are numerous instances wherever artificial intelligence is used for the diagnosis of the chest cancer. The intelligent computing techniques can be used for diagnostic sciences in biomedical image classification. **Faye et al** proposes a method for classification of images based on preselecting features based on their capabilities of differentiating classes using a T test. Random subsets achieving a predefined accuracy rate are then used to generate a final set of features. The method was used in this work with wavelet transform with LDA and kNN classifiers. Although the final accuracy rate obtained in the experiments is relatively low, the improvement when combining classifiers is highly encouraging. **Pereira D. C. et al** presents a set of computational tools to aid segmentation and detection of mammograms that contained mass or masses in CC and MLO views. An artifact removal algorithm is first implemented followed by an image denoising and gray-level enhancement method based on wavelet transform and Wiener filter. Finally, a method for detection and segmentation of masses using multiple thresholding, wavelet transform and genetic algorithm is employed in mammograms which were randomly selected from the Digital Database for Screening Mammography (DDSM). **Jen C. et al** proposed a high-performance CAD system for detecting abnormal mammograms by using the two-stage classifier ADC, which applied the PCA-based technique accompanied by robust feature weight adjustments. **R. Ramaniet al** has been researching on the preprocessing techniques for breast cancer detection in mammography images. They researched on median, adaptive median, mean & Wiener types of filtering used for pre-processing to improve image quality, remove noise, preserve edges within an image, enhance and smoothen the image, mainly concentrate on MSE, PSNR and AE. Finally, compared the simulated output parameters such as image quality, mean square error, Peak signal to noise ratio, structural content and normalized absolute error on 322 mammogram images (MIAS). **D. Sujitha Priya et al** research on breast cancer detection in mammogram images using region-growing and contour-based segmentation techniques by the implementation of preprocessing methods such as, mean filtering, median filtering and adaptive median filtering. Adaptive

Median Filtering technique that is implemented with a Median filter produced the best result among three with measuring MSE and PSNR value. **Jawad Nagi et al** have developed an algorithm on artifact suppression & background separation. Raw mammogram image contains wedges and labels. These may produce unnecessary disturbances during mass detection process. Hence it should be removed in preprocessing. They proposed the method of thresholding and morphological opening, closing, dilation and erosion are used to remove these artifacts. **Armen Sahakyan** has developed an algorithm on Segmentation of the Breast Region in Digital Mammograms & Detection of Masses. In mammogram images radiopaque artifacts such as wedges & labels are removed using threshold technique and morphological operations for enhancement purpose. **R. Subash Chandra Boss et al** research on automatic mammogram image breast region extraction and removal of pectoral muscle. The presence of pectoral muscle in mammograms may disturb thus he has proposed an automated method to identify the pectoral muscle in MLO view mammograms based on histogram based 8-neighborhood connected component labelling method for breast region extraction and removal of pectoral muscle. The proposed method is evaluated by using the mean values of accuracy and error. The comparative analysis shows that the proposed method identifies the breast region more accurately.

3. Existing Work

K-NN

In pattern recognition, the k-nearest neighbor algorithm (k-NN) is a method for classifying objects based on closest training examples in the feature space. k-NN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. The k-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors (k is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of its nearest neighbour. The training examples are vectors in a multidimensional feature space, each with a class label. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples. In the classification phase, k is a user-defined constant, and an unlabeled vector (a query or test point) is classified by

assigning the label which is most frequent among the k training samples nearest to that query point.

4. Proposed Work

A. Image Preprocessing

Mammogram images usually have noises due to disturbances like Gaussian noise or some little darkness and brightness noise called salt and pepper noise. In this paper, we use median filter to remove these noises. Median filter is a nonlinear method effectively used for removing noise while retaining edges. It works by moving the little window called filter that moves pixel

by pixel through the image and changes the pixel value to be the median of neighboring pixels. The median is calculated by first sorting all the pixel values from the filter into numerical order, and then picking the middle pixel value. The output of this de-noising step is the clearer image without noise. In our method Segmented Mammogram images are then filtered using three different image filters is as shown in Figure 1. These filters are intended to help compensate for both intensity variations within an image domain (such as non uniform illumination changes), as well appearance variations between image domains.

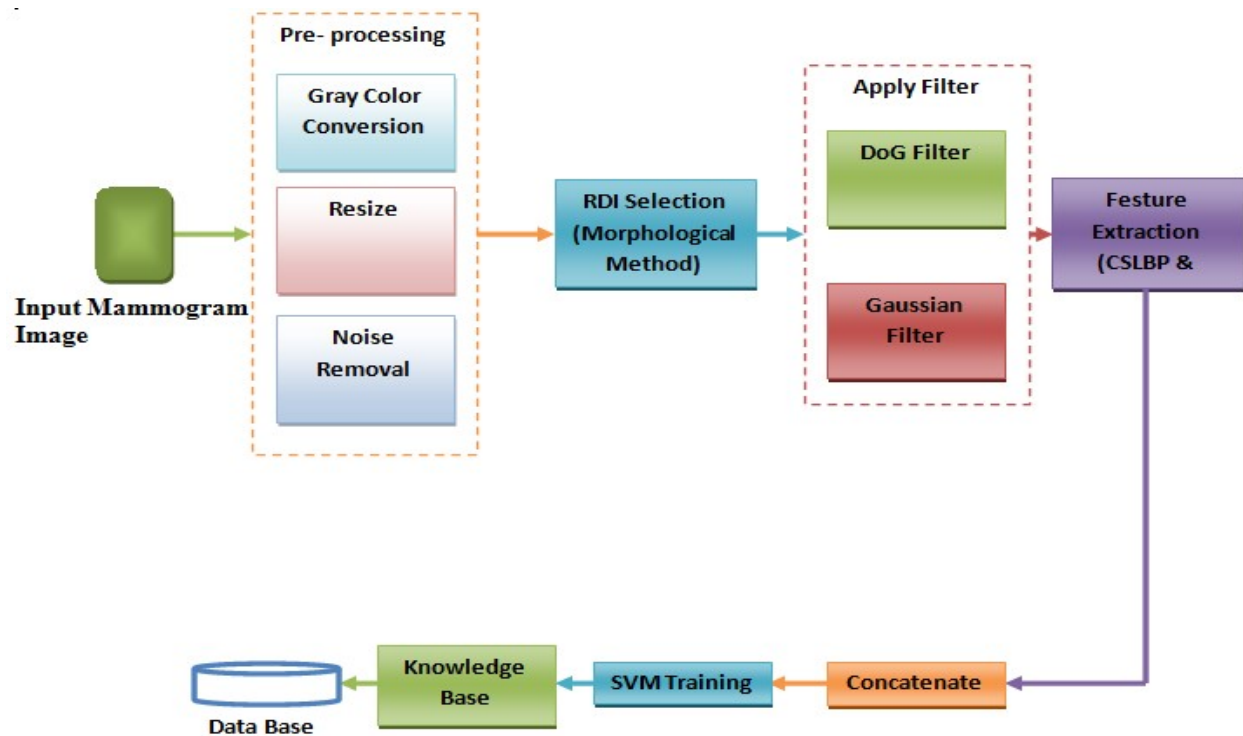


Figure 1: Block diagram of Proposed Method

A Mammogram image is taken as an input and preprocessing is carried out. Pre-processing stage is a step used to increase image quality of Mammograms as they are very difficult to interpret. An histogram equalization can be used to adjust the image contrast so that anomalies can be better emphasized.

B. SVM Classifier

SVM (Support Vector Machine) is a machine learning method that works on the principle of structural risk minimization in order to find the best hyper plane that separates two classes (normal and abnormal). The data used for this SVM is training data and testing data. In this research, testing data are divided into 3 groups. The first group, testing data were taken inside from training data. The second group, testing data were taken outside

from training data. And the third group, testing data were taken inside and outside from training data. Grouping is performed to see the accuracy from each group. The process of classification is performed to classify category of normal and abnormal from mammogram image.

The extracted features are finally combined and presented to a Support Vector Machine classifier, Consider the pattern classifier, which uses a hyper plane to separate two classes of patterns based on given examples. Where is a vector in the input space and denotes the class index taking value 1 or 0. A support vector machine is a machine learning method that classifies binary classes by finding and using a class boundary the hyper plane maximizing the margin in the

given training data. The training data samples along the hyper planes near the class boundary are called support vectors, and the margin is the distance between the support vectors and the class boundary hyper planes. The SVM are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between assets of objects having different class memberships. SVM is a useful technique for data classification. A classification task usually involves with training and testing data which consists of some data instances. Each instance in the training set contains one “target value” (class labels) and several “attributes”. In the field of medical imaging the relevant application of SVMs is in breast cancer diagnosis. The SVM is the maximum margin hyper plane that lies in some space. The original SVM is a linear classifier.

For SVMs, using the kernel trick makes the maximum margin hyper plane fit in a feature space. The feature space is a non linear map from the original input space, usually of much higher dimensionality than the original input space. In this way, non linear SVMs can be created. Support vector machines are an innovative approach to constructing learning machines that minimize the generalization error. They are constructed by locating a set of planes that separate two or more classes of data. By construction of these planes, the SVM discovers the boundaries between the input classes; the elements of the input data that define these boundaries are called support vectors.

5. Experimental Results

Table I depicts the classification precision of two classifiers in two class problem. The results show that various displacements with SVM classifier provides the best classification accuracy of 95.83%. Thus as displacements in GLDM are increased we get the best classification accuracy. In case of GLDM descriptor with the K-NN classifiers results seen are with maximum accuracy of 50% which is not at par with the SVM and combination results with 95.83%.

Displacement Verses Percentage Accuracy

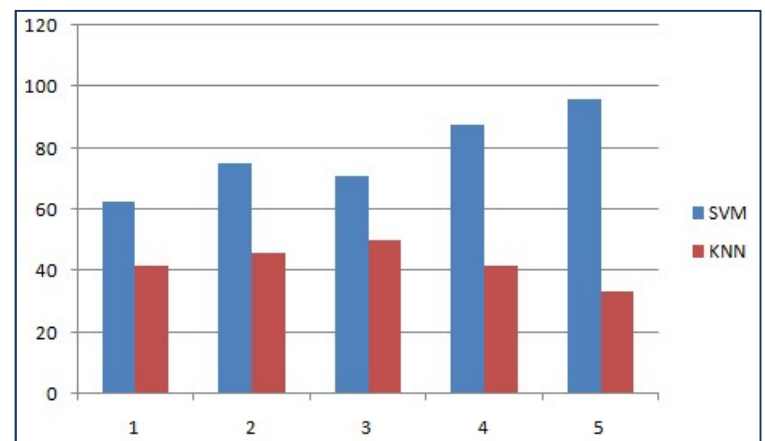


Figure 2: Classification Precision

The figure 2 graph above is plotted with x-axis showing Displacements given in GLDM descriptor and y-axis showing percentage accuracy of classifiers such as SVM classifier and K-NN classifier. Here it shows that best classification accuracy is achieved with SVM classifier whose bar graph is shown in blue than bar graph for K-nn classifier shown in red for various displacements in GLDM descriptor.

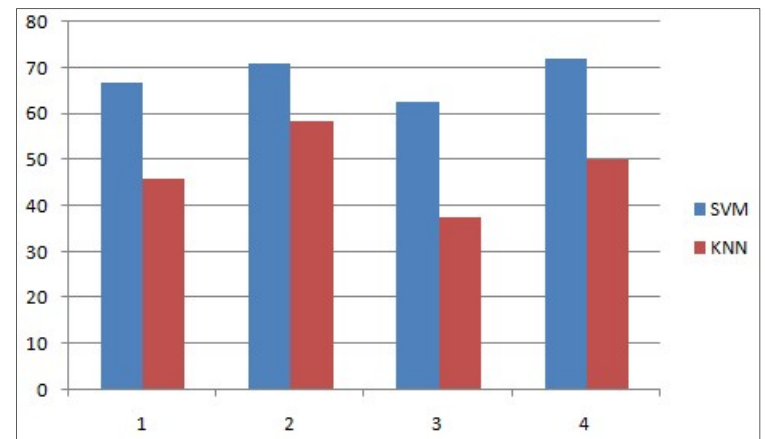


Figure 3: Percentage Accuracy

The figure 3 graph above is plotted with x-axis showing various orientations given in Gabor texture feature descriptor and y-axis showing percentage accuracy of classifiers such as SVM classifier and K-NN classifier. Here it shows that best classification accuracy of 75.5% is achieved with SVM classifier whose bar graph is shown in blue than bar graph for K-nn classifier with percentage accuracy of 54.21% shown in red for various orientations in Gabor texture feature descriptor.

Conclusion

Mammography preprocessing utilizing Support vector machine with image improvement. The proposed framework has been created for diagnosing of bosom tumor from mammogram pictures. In first stage, the preprocessing on mammogram picture is done which limit the computational cost and amplify the likelihood of precision. This exploration has demonstrated that SVM technique is exceptionally viable for the programmed recognition, preprocessing and order of variations from the norm in computerized mammogram. The assessment of the framework is completed on standard dataset.

REFERENCES

- 1) Jawad Nagi , Sameem Abdul Kareem , FarrukhNagi , Syed Khaleel Ahmed “Automated Breast Profile Segmentation for ROI Detection Using Digital Mammograms”, 2010 IEEE EMBS Conference on Biomedical Engineering & Sciences (IECBES 2010), Kuala Lumpur, Malaysia, pp87-99,2010.
- 2) R. Ramani “The Pre-Processing Techniques for Breast Cancer Detection in Mammography Images” I.J. Image, Graphics and Signal Processing, pp 47-54, 2013.
- 3) D. SujithaPriya “Breast Cancer Detection In Mammogram Images Using Region-Growing And ContourBased Segmentation Techniques” International Journal of Computer & Organization Trends ,Volume 3, Issue 8 Sep 2013,ISSN: 2249.
- 4) Jawad Nagi,1, Sameem“Automated Breast Profile Segmentation for ROI Detection Using Digital Mammograms”2010 IEEE EMBS Conference on Biomedical Engineering & Sciences (IECBES 2010), Kuala Lumpur, Malaysia, 30th November - 2nd December 2010.
- 5) Armen Sahakyan “Segmentation of the Breast Region in Digital Mammograms and Detection of Masses” International Journal of Advanced Computer Science and Applications, Vol. 3, No.2, 2012.
- 6) R. Subash Chandra Boss, K. Thangavel, D. Arul Pon Daniel “ Automatic Mammogram image Breast Region Extraction and Removal of Pectoral Muscle”.
- 7) Pragathi. J, H. T. Patil “Segmentation Method for ROI Detection in Mammographic Images using Wiener Filter and Kittler’s Method” International Journal of Computer Applications (0975 – 8887) International Conference on Recent Trends in engineering & Technology , pp27-33, 2013.
- 8) S.M.Salve, V.A.Chakkarwar et .al “Classification of Mammographic images using Gabor Wavelet and Discrete Wavelet Transform” International Journal Of advanced research in ECE ISSN:2278-909X,Vol. 2 pp.573-578, May 2013.
- 9) U.S.Ragupathy,T.Saranya et.al “Gabor Wavelet based Detection of Architectural Distortion and Mass in Mammographic Images and Classification using Adaptive Neuro Fuzzy Inference System” International Journal of Computer Applications, Vol. 46– No.22,pp. 0975 – 8887, May 2012.
- 10) S.N. Deepa ,B. Aruna Devi et.al “A survey on artificial intelligence approaches for medical image classification”, Indian Journal of Science and Technology, Vol. 4 No. 11 ,pp. 1583-1594, Nov 2011.