GABOR WAVELET BASED FEATURES FOR MAMMOGRAM ANALYSIS AND CLASSIFICATION OF BREAST CANCER

1SNEHAL A. MANE, 2K. V. KULHALLI

1P.G.Student, Department Of Electronics & Telecommunication Engg, D.Y.Patil College of Engineering & Technology, Kolhapur
2Vice Principal, H.O.D (I.T), Department of Information Technology, D.Y.Patil College of Engineering & Technology, Kolhapur
E-mail: 1snehalmane.17@gmail.com, 2kvkulhalli@gmail.com

Abstract—Breast cancer is the most commonly occurring cancer in women. Early detection of breast cancer is important step in diagnosis of the abnormalities which may reduce the mortality rate. It can be achieved using digital mammography. Digital mammography has become the most effective techniques for the early detection of breast cancer. Mammography is most reliable and widespread method for early detection of breast cancer. This system includes Preprocessing on mammogram image and uses wavelet feature extraction to improve sensitivity. It involves three major steps-Preprocessing, Feature Extraction and Classification. Gabor wavelets based features are extracted from medical mammogram images representing normal tissues or benign and malign tumors. After completing preprocessing, Principal Component Analysis (PCA) is used to extract feature vectors of all images in the database. PCA is a useful statistical technique that has found application in fields such as face recognition and image compression, and is a common technique for finding patterns in data of high dimension. These extracted features are used for classification of mammogram into malignant and benign. The SVM classifier is used for classification.

Index Terms—Gabor Wavelet, Mammogram images, Principal Component Analysis, Support Vector Machine.

I. INTRODUCTION

Image manipulation is commonly used in image processing and pattern recognition field. Image manipulation may include quality enhancement, filtering, segmentation, feature selection or extraction and dimensionality reduction. When it comes to image classification, it is desirable to keep discriminate features and discard non-relevant features that may negatively affects the classification performances. Mammographic images are X-ray images of breast region displaying points with high intensities density that are suspected of being potential tumors.

Automatic tumor detection is extremely challenging as the suspicious calcification or masses appear as free shape and irregular texture, so that no precise patterns can be associated to them. Breast cancer is one of the frequent and leading causes of mortality among women, especially in developed countries. Mammography is a special type of x-ray imaging used to create detailed images of the breast. Mammography uses low dose x-ray, high contrast, high resolution film specifically for imaging the breasts. Mammography plays a major role in early detection of breast cancer. This paper proposes a technique for detecting architectural distortion and mass in mammographic images using Gabor wavelets. The detection of tumours in mammogram is divided into three main stages. [2] The first step involves an enhancement procedure, image enhancement techniques are used to improve an image. Then the Gabor Wavelet based features are extracted from mammogram. Then the next stage involves the classification using SVM classifier.

II. PROPOSED SYSTEM

A simplified Flow chart of proposed system is depicted in following fig.1

Fig -1: Flowchart of Proposed System

1. Digitization: Digitization produces the digital image, which is fed to the pre-processing phase.

2. Preprocessing: After digitization image may carry some unwanted noise. The preprocessing stage reduces
noise and distortion, removes skewness and performs skeltonizing of the image. After preprocessing phase, a cleaned image goes to the segmentation phase.

3. Segmentation: The segmentation stage takes in the image and separates the different logical parts. Segmentation is the process of partitioning an image into semantically interpretable regions.

4. Feature Extraction: After segmentation, set of features are required for each image. In feature extraction stage every image is assigned a feature vector to identify it. This vector is used to distinguish the image. Gabor Wavelet is used for Feature Extraction.

5. Classification: Classification is the main decision making stage of image recognition and finding whether the image is cancerous or not. It uses the features extracted in the previous stage to identify the image according to preset rules. Advanced classifiers such as Support Vector Machines (SVM) applied to image patches extracted around image.

III. IMAGE PREPROCESSING

The preprocessing plays an important role in any CAD system. A preprocessing phase of the images is necessary to improve the quality of the images and make the feature extraction phase more reliable. Collected mammogram images used as database are as follows:

![Mammogram images](image)

Above images are collected from Mammographic Image Analysis Society and consider as a database in this proposed system. These are normal and abnormal images.

To reduce the noise and normalize the staining intensity variation, median filter and histogram based contrast enhancement have been used. Figure 3 represents results of pre-processing.

IV. SEGMENTATION

In segmentation methods divide the image into number of small segments. The goal of segmentation is to
identify the correct areas and to analyse the diagnosis. In this proposed system active counter is used. Automatic Seeded Region increasing using haralick texture features are as follows Divide improved ROI into RxR nonoverlapping blocks. If block is too tiny, difference of mass textures starting typical textures cannot be fine characterized. If it is too huge, result may be too common compute Haralick texture features starting Spatial Gray Level Dependence 23 Matrix (SGLD) of every block. Select significant features that can easily distinguish mass and non mass region. Seed point is automatically selected that correspond to maximum Sum Average feature value. Segmented image is smoothed using some morphological operators like dilation, erosion, imclose. Extracted mass is approximated to a circle and compare its radius to the original radius. Extracted mass is used as an input to the classification stage.

The segmentation stage takes the image and separates the different logical parts. Segmentation is process of partitioning a picture into semantically interpretable regions. Simulation Results of ROI.

V. FEATURE EXTRACTION

When preprocessing and desired stage of segmentation has been accomplished, some feature extraction procedure is applied to segments to gain features, and it is followed by application of classification in addition to post processing methods. It is necessary to focus on feature extraction stage as it has a noticeable impact on effectiveness of recognition scheme. Feature selection of a feature extraction technique is single most significant factor in achieving high recognition presentation. The feature extraction has been given as extracting from raw information that is most appropriate for classification reasons, while minimizing within class example variability in addition to enhancing between class pattern variability. So, selection is done to fit feature extraction technique according to input to be applied desires to be done with highest care.

Figure 5. shows simple block diagram of Feature Extraction. After segmentation phase, collections of features are required for each image. In the feature extraction phase each picture is assigned feature vector to recognize it. First stage of normal recognition is to map mammogram into a collection of features. The features can be a collection of real numbers and it is characterize a normal mammogram. The question is how to locate characteristic features which can be used to distinguish class of normal mammograms from class of unusual mammograms. A few simple statistical features could be mean with standard deviation of pixel intensities of region.

In mammography image the spatial resolution of x-ray which is in the order of few microns permits to visualize masses. But the conventional mammograms are highly textured and complex. This makes the interpretation difficult. For this reason, it is necessary to extract features from the mammogram to improve performances of the diagnosis in terms of precision and reliability. The feature extraction and selection from an image plays a critical role in the performance of any classifier. Higher accuracy of the classifier can be achieved by the selection of optimum feature set.

2D Gabor wavelets have been widely used in computer vision applications and modelling biological vision, since recent studies have shown that Gabor elementary functions are suitable for modelling simple cells in visual cortex. Other nice property is provided by their optimal joint resolution in both space and frequency, suggesting simultaneously analysis in both domains [6].

A complex Gabor wavelet (filter) is defined as the product of a Gaussian kernel with a complex sinusoid. A 2D Gabor wavelet transform is defined as the convolution of the image $I(z)$:

$$J_k(z) = \int \int I(z') \psi_k(z - z') dz'$$

(4)

With a family of Gabor filters (functions):

$$\psi_k(z) = \frac{k^T k}{\sigma^2} \exp\left(-\frac{k^T k}{2\sigma^2} z^2\right) \left(\exp(k^T z) - \exp(-\sigma^2/2)\right),$$

(5)
Where \( \mathbf{z} = (x, y) \) and \( \mathbf{k} \) is the characteristic wave vector:

\[
\mathbf{k} = (k_x \cos \varphi \mu, k_y \sin \varphi \mu)^T
\]

\[\ldots \ldots (6)\]

With

\[
k_x = 2^{-1/2} \pi, \quad \varphi \mu = \mu \frac{\pi}{8}, \quad \nu = 0, 1, 2, 3, 4, \quad \mu = 0, 4, \frac{\pi}{2}, \frac{3\pi}{4}.
\]

\[\ldots \ldots (7)\]

The parameters \( \nu \) and \( \mu \) define the frequency and orientation of the filter. PCA is the simplest of the true eigenvector-based multivariate analyses. Often, its operation can be thought of as revealing the internal structure of the data in a way that best explains the variance in the data. If a multivariate dataset is visualized as a set of coordinates in a high-dimensional data space, PCA can supply the user with a lower-dimensional picture, a "shadow" of this object when viewed from its most informative viewpoint. This is done by using only the first few principal components so that the dimensionality of the transformed data is reduced.[3,4]

The parameters \( \nu \) and \( \mu \) define the frequency and orientation of the filter. PCA is the simplest of the true eigenvector-based multivariate analyses. Often, its operation can be thought of as revealing the internal structure of the data in a way that best explains the variance in the data. If a multivariate dataset is visualized as a set of coordinates in a high-dimensional data space, PCA can supply the user with a lower-dimensional picture, a "shadow" of this object when viewed from its most informative viewpoint. This is done by using only the first few principal components so that the dimensionality of the transformed data is reduced.[3,4]

After extraction, select create feature vector by clicking on features menu and selecting create feature vector option.

**VI. CLASSIFICATION**

Classification is main decision making step of image recognition and also finding whether image is cancerous or not. It uses features extracted in preceding stage to recognize image segment according to preset rules. The advanced classifiers such as Support Vector Machines (SVM) applied to picture or image patches extracted around image. Support vector machines (SVM) are based on the Structural Risk Minimization principle [14] from statistical learning theory. SVM is also applied on different real world problems such as face recognition, cancer diagnosis and text categorization. The idea of structural risk minimization is to find a hypothesis \( h \) with the lowest true error. Vapnik connects the bounds on the true error with the margin of separating hyper planes. In their basic form, support vector machines find the hyper plane that separates the training data with maximum margin. SVM is a useful technique for data classification. A classification task usually involves with training and testing data which consist of some data instances.

Each instance in the training set contains one “target value” (class labels) and several “attributes” (features). The standard SVM (figure 4) takes a set of input data, and predicts, for each given input, which of two possible classes the input is a member of which makes the SVM a non-probabilistic binary linear classifier.

![Support Vector Machine with a hyper plane](image)

Since an SVM is a classifier, then given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that predicts whether a new example falls into one category or the other. More formally, a support vector machine constructs a hyper plane or set of hyper planes in a high or infinite dimensional space, which can be used for classification, regression or other tasks. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training data points of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

![Classification](image)
In this system, SVM classifier is used to detect whether the image is cancerous or not. Following fig 8 represents how the SVM classifier works in this system. SVM classifier is used to detect whether the image is cancerous or not.

VII. CLASSIFIER RESULTS

![Classifier results for database](image)

VIII. EXPERIMENTAL RESULTS

The images are taken from the MIAS database [16] which consists of 41 normal images and 85 abnormal images. The abnormal images are further classified into two classes i.e. benign and malign. There are total 64 benign images and 20 malign images. To discard irrelevant (background) information like breast contour, patches of 140 × 140 pixels surrounding the abnormality region were extracted from the original 1024 × 1024 pixels images. The patch size assures that, for most abnormal cases not only the abnormality region is captured but also the surrounding area, providing us information about the abnormality shape.

![Results for same types of images in database](image)

Figure 10 Provides Experimental Results Of Gabor Filter and PCA.

CONCLUSION

The proposed system has been developed for diagnosing of breast cancer from mammogram pictures. In first stage, the preprocessing on mammogram picture is done which minimize the computational cost and maximize the probability of accuracy. To summarize the developed method, the initial step, based on gray level information of image enrichment and segments the chest cancer. In second stage Gabor Wavelet features are extracted. These extracted features are used for classification of mammogram into malignant and benign. The SVM classifier is used for classification.

The images are taken from MIAS database which includes of 41 normal images and 85 abnormal images. The abnormal images are more classified into two classes which are benign and malign. There are total 64 benign images and 21 malign images. To remove irrelevant background information like breast contour, patches of 140 into 140 pixels surrounding the oddity region were extracted from original 1024 into 1024 pixels images. Patch size assures that, for most odd cases not only the oddity region is captured but also the surrounding area, providing us information about the oddity shape.

REFERENCES


