 ©2015 IJIRCT | ISSN: 2454-5988

Detection and Classification of Mammogram Images Using Principle Component Analysis and Lazy Classifiers

Rajkumar K. K
School of Information Science & Technology
Mangattuparamba Campus, Kannur University
Kannur, Kerala, India
rajatholy@yahoo.com

***Abstract*—Feature extraction and selection is the primary part of any mammogram classification algorithms. The choice of feature, attribute or measurements have an important influence in any classification system. Discrete Wavelet Transformation (DWT) coefficients are one of the prominent features for representing images in frequency domain. The features obtained after the decomposition of the mammogram images using wavelet transformations have higher dimension. Even though the features are higher in dimension, they were highly correlated and redundant in nature. The dimensionality reduction techniques play an important role in selecting the optimum number of features from the higher dimension data, which are highly correlated. PCA is a mathematical tool that reduces the dimensionality of the data while retaining most of the variation in the dataset. In this paper, a multilevel classification of mammogram images using reduced discrete wavelet transformation coefficients and lazy classifiers is proposed. The classification is accomplished in two different levels. In the first level, mammogram ROIs extracted from the dataset is classified as normal and abnormal types. In the second level, all the abnormal mammogram ROIs is classified into benign and malignant too. A further classification is also accomplished based on the variation in structure and intensity distribution of the images in the dataset. The Lazy classifiers called Kstar, IBL and LWL are used for classification. The classification results obtained with the reduced feature set is highly promising and the result is also compared with the performance obtained without dimension reduction.**

Keywords— PCA, Wavelet Transformation, Lazy classifiers, Kstar, IBL, LWL

I. INTRODUCTION

Breast cancer is the one of the most threatening disease found among women in all over the world. It stands second in position for the cause of deaths in women, especially in the developed and under developed countries [1]. There are no effective diagnostic methods suggested so far for this disease. The only way to decrease the mortality rate of the breast cancer is the early detection [2]. The commonly used diagnostic methods for breast cancer are biopsy, mammography, and thermograph and ultrasound imaging [3]. Out of it mammography is the best method currently available. Provide both accurate and uniform evaluation for the enormous number of mammogram generated in widespread screening. Human observations have limitations, for example

The early signs of breast tumor are very subtle and vary in appearance. However, it is very difficult for radiologist to some anomalies may be missed due to human errors as a result of fatigue. [4][5]. The most accurate detection method in the medical environment is the biopsy. But it has some discomfort for the patient and its cost is high. Biopsy also involves high percentage of negative cases. Therefore computer aided detection (CAD) will provide as a second opinion for the detection of the tumor. As a first step of the CAD, X-ray mammography is considered as a standard procedure for breast screening and diagnosis. But the performance of this X-ray mammography for the breast cancer screening is also not up to the mark, the accuracy is only about 75% [6]. Screen film mammography is the best suitable method accepted today for the diagnosis. It reduces the negative biopsy ratio and the cost to society by improving feature analysis and refining criteria for recommending biopsy.

Mammographic images are X-ray images of breast region which displays points with high intensities are suspected of being potential tumors. Early diagnosis and screening are crucial for having a successful medical treatment or cure. Typically, masses and calcium deposits are easily identifiable by visual inspection. These deposits appear much denser than other types of surrounding soft tissues. Malignant tumors are usually associated to unusual smaller and clustered calcification. Other calcification types including diffuse, regional, segmental or linear corresponds to benign tumors. Such calcifications are termed as microcalcification. Automatic tumor classification would require the segmentation of the microcalcification area from the X-ray image, followed by recognition or classification of the segmented area into one of these three classes: normal, benign or malignant tumor. 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. In addition, the presence of more or less prominent blood vessels and muscle fibers may seriously degrade the accuracy of identification of tumor recognition [7].

Pattern recognition tasks require the conversion of pattern in features describing the collected sensor data in compact form. Feature selection methods can be either classical methods or biologically oriented methods. Feature extraction and selection in pattern recognition are based on finding mathematical methods for reducing dimensionality of pattern representation. A lower dimensional representation based on pattern descriptors is a so-called feature. It plays a crucial role in determining the separating properties of pattern classes. The choice of features, attributes, or measurements has an important influence on accuracy of the classification, time needed for the classification, number of examples needed for learning and the cost of the performance of the classification [8].

Dimensionality reduction plays an important role in classification performance. A recognition system is designed using finite set of inputs. While the performance of the system increases if we add additional features, at some point a further inclusion leads to performance degradation. Thus dimensionality reduction may not always improve a classification system.

Computer aided diagnosis of breast tumor is one of the challenging task in the field of medical image processing. There are plenty of works have been already published in this area and most of them have very good results. But we cannot rely any of these method due to some artifact related to all these methods too. In most of these methods, texture information plays an important role in image analysis and detection of breast cancer. Texture is one of the important characteristics used in identifying an object or region of interest (ROI) in an image [9]. Features are extracted from these texture information and these features are analyzed using soft computing tools. Then each classification system performance is analyzed by computing sensitivity and specificity [10]. An ROI may be called cancerous (positive) or normal (negative) and a decision for detection result will be one of four possible categories: true positive (TP), true negative (TN), false positive (FP) and false negative (FN). FN and FP represents two kinds of errors. An FN error implies that true abnormality was not detected and a FP error occurs when a normal region was falsely identified as abnormal image. A TP decision is correct judgment of an existing abnormality and a TN decision means that a normal region was correctly labeled [9] [11]. Therefore the accuracy and performance of any CAD system is evaluated based on the Sensitivity, Specificity and Accuracy. They are defined as follows:

$$\text{Sensitivity} = \frac{TP}{(TP+FN)} \tag{1}$$

$$\text{Specificity} = \frac{TN}{(TN+FP)} \tag{2}$$

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+FP+TN+FN)} \tag{3}$$

In this paper we propose a new method for classifying mammogram images using lazy classifiers. The lazy classifiers namely Kstar, IBk, IB1 and LWL are use the reduced discrete wavelet transformation coefficients as feature set for the classification. The wavelet coefficients obtained after the decomposition have higher dimension. So processing these coefficients as feature set is very time consuming and introduce performance degradation due to the volume of the feature set. Therefore some dimension reduction method is used for reducing the number of features for classification. The Principle Component Analysis (PCA) technique is common and simple technique used for reducing dimensionality of the wavelet coefficients. This reduced feature set acts as a dominant feature for classifying mammogram images into different categories. The feature set of the mammogram images of ROIs of size 32x32 are extracted from the Mini-Mias database. The ROIs' are extracted based on the abnormality center of the image in the database. The classification of the mammogram images in the dataset is done in three different stages. In the first stage of the classification true normal and abnormal images are identified. In the second level, abnormal images are classified into benign and malignant types. Finally all the abnormal images in the dataset are classified into different subcategories depending upon the variation in structure and intensity distribution of the mammogram images in the abnormal set. This paper consists of eight sections. Section II discusses about the dataset used for the classification. The wavelet transformations and its characteristics are presented in Section III. In section IV principal component analysis (PCA) for dimension reduction is discussed. Section V discusses the various lazy classifiers used for the classification. The proposed method is explained in Section VI. The results obtained by the different lazy classifiers on the reduced feature sets are presented and compared with the results obtained without PCA in Section VII and finally Section VIII concludes the work with its future scope.

II. DATASET

Mammogram images are the low intensity gray scale images, which indicate the details inside the patient breast by means of its contrast. The details could be normal tissues, vessels, muscles, different types of masses and noise. Each type of masses has different properties of shape, size and brightness. These properties are normally acted as features of the dataset. These features may help the radiologist for the effective diagnosis of the breast tumors.

In this study we used a set of mammogram images provided by Mammographic Image Analysis Society (Mini-MIAS) [12]. The database contains left and right breast images of 161 patients. It consists of 322 mammogram images, which belongs to three categories namely normal, benign and malignant images. The images in the set are digitized at 50 micron pixel edge and have been reduced to 200 micron pixel edge and clipped or padded so that every image is 1024x1024 pixels. These images are investigated and labeled by an expert radiologist. From these dataset images, regions of interest (ROIs) of sizes 32 x 32 pixels are extracted for our investigation. The ROIs are extracted from the original mammogram images based on the abnormality center for the cancerous images already marked by the radiologist whereas the non-cancerous images are extracted with respect to the center of the original mammogram images. For practical evaluation of this system the entire dataset which comprises 322 ROIs of different types of lesions as shown in Table 1 are used. While extracting ROIs of cancerous images, multiple abnormal regions in an image are treated as individual ROIs for the classification purpose.

TABLE I. LESION DISTRIBUTION OF MIAS DATABASE

LESION	RISK	#
Normal		207
Architectural distortion[ARCH]	Benign	09
	Malignant	10
Asymmetry[ASYM]	Benign	06
	Malignant	06
Microcalcification[CALC]	Benign	12
	Malignant	13
Circumscribed masses[CIRC]	Benign	19
	Malignant	04
Ill-defined masses[MISC]	Benign	06
	Malignant	08
Spiculated lesions[SPIC]	Benign	11
	Malignant	08
Total		322

III. WAVELET TRANSFORMATIONS

Wavelet Transformation (WT) is a mathematical tool for analyzing signals and images in time frequency domain. It decomposes signals or images into different functions called wavelet family in which all of the basic functions are derived from scaling and translation of single function called the mother wavelet. By representing signals or images in time frequency domain has two main advantages: (a) an optimal resolution both in the time and frequency domains; and (b) the lack of stationary nature of the signal. It is defined as the convolution between the signal X (t) and the wavelet functions ψ(a, b) (t) and it is represented as:

$$\omega_{\psi} X(a,b) = < X(t) \,|\, \psi_{(a,b)}(t) > \qquad (4)$$

Where ψ(a,b)(t) are dilated or contracted and shifted versions of a unique wavelet function ψ (t)

$$\psi(a,b) = |\,\mathrm{a}\,|^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right) \qquad (5)$$

(a, b are the scale and translation parameters, respectively).The WT gives a decomposition of X (t) in different scales, tending to be maximum at those scales and time locations where the wavelet best resembles X (t). Moreover, Eq. (4) can be inverted, thus giving the reconstruction of X (t). The WT maps a signal of one independent variable t onto a function of two independent variables a, b. This procedure is redundant and not efficient for algorithmic implementations. In consequence, it is more practical to define the WT only at discrete scales a and discrete times b by choosing the set of parameters {aj = 2 –j; bj, k = 2-j k}, with integers j, k.

Contracted versions of the wavelet function match the high frequency components of the original signal and on the other hand, the dilated versions match the low frequency components. Then, by correlating the original signal with wavelet functions of different sizes we can obtain its details at different scales. These correlations with the different wavelet functions can be arranged in a hierarchical scheme called multi resolution decomposition. The multi resolution decomposition separates the signal into 'details' at different scales, the remaining part being a coarser representation of the signal called 'approximation'. The decomposed signal or image contains details and approximation. The lower levels give the details corresponding to the high frequency components and the higher levels corresponding to the low frequencies [13].

A. Discrete Wavelet Transformation (DWT)

Discrete wavelet transformation or decimated wavelet transformation is the most useful technique for frequency analysis of signals that are localized in time space. The discrete wavelet transform corresponds to multiresolution approximation expression. This method permits the analysis of the signal in many frequency bands or at many scales. In practice, mutiresolution analysis is carried out using 2 channel filter banks composed of a low-pass (G) and a high-pass (H) filter and each filter bank is then sampled at a half rate (1/2 down sampling) of the previous frequency. By repeating this procedure, it is possible to obtain wavelet transform of any order. The down sampling procedure keeps the scaling parameter constant (n=1/2) throughout successive wavelet transforms so that it benefits for simple computer implementation. In the case of an image, the filtering is implemented in a separable way by filtering the rows and columns.

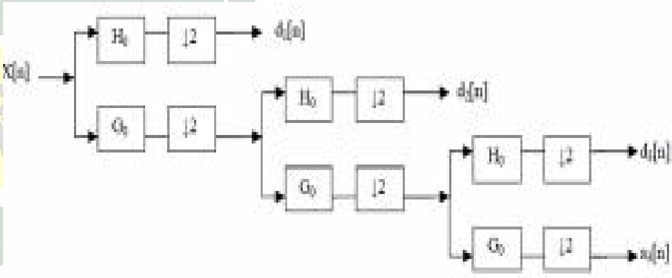

Fig. 1. The Discrete Wavelet Transform.

The discrete wavelet transform also very useful for texture analysis in the image. Its fast implementation is usually performed by using multiresolution analysis. The wavelet coefficients are sampled based on the Nyquist criteria. The transformation coefficients are non-redundant and the total number of sample in the transformation is equal to the total number of the image pixels. It also reduces the computation time because of the down sampling of the coefficients [14][15].

IV. PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal Component Analysis (PCA) is a mathematical algorithm that reduces the dimensionality of the data while retaining most of the variation in the dataset. It accomplishes the reduction by identifying directions called principal components along which the variation in the data is maximal. By using PCA, each sample can be represented by relatively few numbers instead of by values for thousands of variables. Samples can then be plotted, making it possible to visually assess similarities and differences between samples and determine whether samples can be grouped or not [16].

PCA identifies new variables, the principal components, which are linear combinations of the original variables. It is easy to see that the first principal component is the direction along which the samples show the largest variation. The second principal component is the direction uncorrelated to the first component along which samples show the largest variation. If dataset are standardized such that each element in the dataset is centered to zero average expression level, the principal components are normalized eigenvectors of the covariance matrix of the element in the dataset and ordered according to how much of the variation present in the dataset contain. Each component can then be interpreted as the direction, uncorrelated to previous components, which maximizes the variance of the samples when projected onto the component [17].

V. LAZY CLASSIFIERS

A classification problem occurs when an object needs to be assigned to a predefined group or class based on a number of observed attributes related to that object [18]. Different types of classification algorithms are available today for the classification in which eager learning and instance based learning algorithms are most prominent. Lazy learning classifiers are instance based or memory based classification algorithm proposed against the common eager learning algorithms. They are the important category of classifiers that can be implemented and tested easily with minimum cost. This learning algorithm utilizes a kind of distance measure between test instances and training instances for the classification. Entropy and distance measures are the two common methods adopted by the Lazy classifiers.

The common eager learning methods eagerly compile the training data into some concept descriptors such as rule sets, decision trees, artificial neural network and graphical models. After constructing such type of models, common eager learning methods attempt to seek a particular general hypothesis, which covers the entire instance space. But the lazy learning models do not conduct any processing of developing a model for classification before they encounter the unseen instance to be classified. The lazy classifier constructs model only when they are directed to classify the unseen instance and discard all the customized models and all the intermediate results after the learning process for the unseen instance completes. Therefore lazy learning algorithms need much less training costs but more storage and computational resources than the eager algorithms. Lazy learning algorithms can make use of the characteristics of the unseen instance to explore a richer hypothesis space during the classification. Due to this richer hypothesis space, lazy learning methods outperform significantly than some of the common eager learning methods.

Lazy learning exhibits many advantages in learning scenarios. Common eager learning methods need to learn a new global classifier every time the training data is updated. When the training data is large and complex, it is not economical for the service provider to conduct eager learning frequently. Lazy learning methods have no such problems. Generally, the updating of training data is the only operation required by lazy learning methods. Another learning scenario for which lazy learning is competitive is that the learning target class is not fixed and the attribute set is large. Lazy learning handles each classification as an independent learning process, and hence it can be customized to the unseen instance and focuses only on the local data patterns [20]. In this paper we use three different instance based classifiers K*, IBL and LWL algorithms.

A. *K* Classifier*

K * is an instance based classifier that classifies the test instance based on the classes of those training instances similar to the test instance determined by some similarity functions. The similarity function is determined by using entropy as a distance measure. The result obtained by this method is comparatively better than the several other machine learning algorithms.

The similarity function computes the similarity between a test instances against the instances in the concept descriptor computed using the training instances in the samples. Let xi , yi denotes test instance and concept descriptor respectively, then the similarity function between these two instances are computed by using the following equation.

$$Similarity(x_i, y_i) = -\sqrt{\sum_{i=1}^{n} f(x_i, y_i)} \qquad (6)$$

Where the instances are described by n attributes. We define $f(x_i, y_i) = (x_i - y_i)^2$ for numeric valued attributes and $f(x_i, y_i) = (x_i \# y_i)$ for Boolean and symbolic-valued attributes. Missing attributes values are assumed to be maximally different from the value present. If the both instances are missing, then $f(x_i, y_i)$ yields 1. The function $f(x, y)$ is the entropy computed using the concept descriptors of the training samples using the equation $E(S) = \sum_i p_i \log p_i$, where $p_i = \frac{s_i}{|s|}$, S_i denotes the number of training instances with class C_i , and $|S| = \sum_i S_i$ be the total number of training instances [21].

B. *IBL Classifier*

Storing and using specific instances improves the performance of several supervised learning algorithms. Instance-based Learning algorithm generates classification prediction using only specific instances. It does not maintain a set of abstractions derived from specific instances. This approach extends the nearest neighbor algorithm which requires large storage requirements similarity function is used for categorizing the matches between testing samples against specific instances. Using these specific instances, Instance based learning algorithm reduces the cost incurred for updating concept descriptors and increases the learning rates. Instance based learning algorithm is derived from the nearest neighbor pattern classifier, which uses only selected instances to generate classification prediction. Therefore instance-based learning algorithm reduces storage requirements and at the same time there is small degradation in classification accuracy [22]

Each instance in IBL classifier is represented by a set of attribute-value pairs. This set of attributes defines an n-dimensional instance space. Exactly one of these attributes corresponds to the category attribute; the other attributes are predictor attributes. A category is the set of all instances in an instance space that have the same value for their category attribute. However, IBL algorithms can learn multiple overlapping concept descriptions simultaneously. The concept description is a function that maps instances to categories that yields the classification. An instance-based concept description includes a set of stored instances and possibly some information concerning their past performance during the classification. This set of instances can change after each training instance is processed. However, IBL algorithms do not construct extensional concept descriptions. Instead, concept descriptions are determined by how the IBL algorithms selected similarity and the classification functions uses the current set of saved instances. The classification function determines how the set of saved instances in the concept descriptions are effectively used to predict the values for the category attribute.

The IBL classification function used for defining concept description have the following components:

1) Similarity Function: This computes the similarity between a testing instances i and the instances in the concept description. 2. Classification Function: This receives the similarity function's results and the classification performance records of the instances in the concept description. It yields a classification for the instance i.

2) Classification Function: This receives the similarity function's results and the classification performance records of the instances in the concept description. It yields a classification for the instance i.

3) Concept Description Updater: This maintains records on classification performance and decides which instances to include in the concept description. Inputs include i, the similarity results, the classification results, and a current concept description. It yields the modified concept description.

The similarity and classification functions determine how the set of saved instances in the concept description are used to predict values for the category attribute. Therefore, IBL concept descriptions not only contain a set of instances, but also include these two functions.

IBL algorithms differ from most other supervised learning methods: they do not construct explicit abstractions such as decision trees or rules. Most learning algorithms derive generalizations from instances when they are presented and used for simple matching procedures to classify subsequently presented instances. This incorporates the purpose of the generalizations at the presentation time. IBL algorithms perform comparatively little work at the presentation time since they do not store explicit generalizations. However its work load is higher when presented with subsequent instances for classification, at which time they compute the similarities of their saved instances with the newly presented instance. This obviates the need for IBL algorithms to store rigid generalizations in concept descriptions, which can require large updating costs to account for prediction errors. [23]

C. LWL Classifier

Lazy learning methods defer processing of training data until a query needs to be answered. This usually involves storing the training data in memory, and finding relevant data in the database to answer a particular query. Relevance is often measured using a distance function, with nearby points having high relevance. One form of lazy learning finds a set of nearest neighbors and selects or votes on the predictions made by each of the stored points.[20]

Locally Weighted Learning (LWL) is lazy classifier that uses statistical learning techniques for training and classifying complex tasks. It provides an approach to learning models of complex phenomena, dealing with large amounts of data, training quickly, and avoiding interference between multiple tasks during control of complex systems. LWL methods can even deal successfully with high dimensional input data that have redundant and irrelevant inputs while keeping the computational complexity of the algorithms linear in the number of inputs.[20][22]

LWL methods come in two different strategies. Memory-based LWL is a “lazy learning” method that simply stores all training data in memory and uses efficient lookup and interpolation techniques when a prediction for a new input has to be generated [20] [22]. This kind of LWL is useful when data needs to be interpreted in flexible ways, for instance, as forward or inverse transformation. Memory-based LWL is therefore a “least commitment” approach and very data efficient. Non-memory-based LWL has essentially the same statistical properties as memory based LWL, but it avoids storing data in memory by using recursive system identification techniques [24]. In this way, non-memory-based LWL caches the information about training data in compact representations, at the cost that a flexible re-evaluation of data becomes impossible, but lookup times for new data become significantly faster.

VI. Proposed Method

In this paper we proposed a multi-stage and multi-level classification of mammogram images using the Wavelet Transformations and lazy learning classifiers. The classification is accomplished in two different levels. In the first level mammogram ROIs extracted from the Mini-Mias dataset are decomposed into discrete wavelet transformation coefficients. Then only fractional parts of the highest wavelet coefficients are taken as sample for representing the feature of the mammogram for classification. Wavelet transformation coefficients of an image have high dimension and redundant in nature. Therefore the processing of high dimension data degrades the performance of the algorithm. Fortunately Wavelet transformation coefficients are highly correlated and therefore most of the coefficients which posses the basic characteristics are retained and the remaining coefficients can be ignored. The Principle Component Analysis (PCA) is effective mathematical tool for identifying such patterns in data and expressing the data in such a way as to highlight their similarities and differences. The main advantage of PCA is that once a pattern in the data is found, it can reduce the number of dimension without much loss of information. By reducing the wavelet transformation coefficients using PCA original mammogram ROIs are initially classified into normal and abnormal images respectively. In the next level all the abnormal images that are classified in the initial level further categorized into benign and malignant types using the same features. Finally all the abnormal images are identified and labeled into different sub categories such as calcification(CALC), asymmetry(ASYM), Architectural distortions(ARCH), Circumscribed masses(CIRC), ill-defined masses(MISC) and speculated masses(SPIC) depending upon the texture features of the mammogram images. The popular lazy learning classifiers K*, IBL and LWL classifiers are used for classifying the mammograms. The architecture of the proposed system is shown in below.

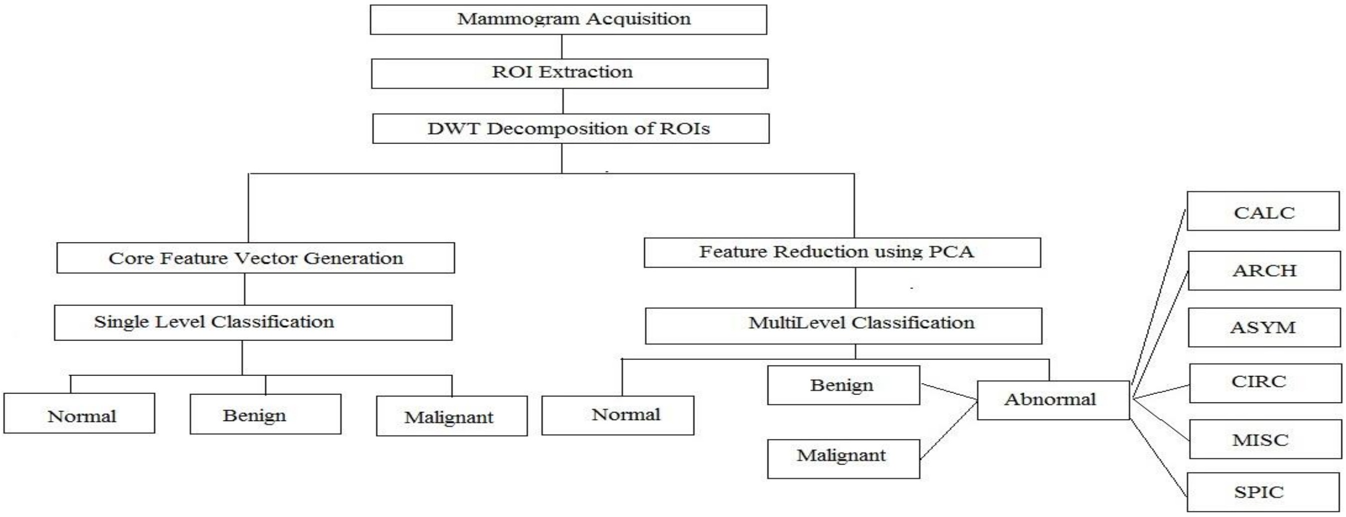

Fig. 2. Block Diagram of the proposed system.

In the first stage of the classification, the different texture characteristics of mammogram images are analyzed using multi-level decomposition of mammogram images using WT. The decomposed images have both approximation as well as detailed coefficients. Approximation coefficients reveal the most of the significant characteristics of the textures of the mammogram. While decomposing the images using WT the coefficients obtained have higher dimension. The processing of all these coefficients is unnecessary and time consuming. Only the top level coefficients are much significant for the classification purpose. Therefore some dimension reduction method is useful for avoiding the insignificant coefficients of the transformations. In the first stage of the classification, we take only the fractional part of the highest wavelet approximation coefficients. The ROIs of size 32 x 32 pixels are extracted from the set of 322 mammogram images from the Mini-Mias Database for the classification. The ROIs are extracted manually by identifying the center location of the abnormality of the mammogram images. Thus we extracted ROIs for both benign and malignant types of mammogram images. But for normal mammogram images, ROIs are extracted of the same size around the centre of the each mammogram images. The classification part consists of two parts viz training and testing. In the training part, we created class core vectors for each class of the mammogram images. The classes are normal, benign and malignant. The class core vectors are computed by taking 10 percent of the ROIs are randomly selected from each class of the image. The class core vectors are created on all four levels of wavelet decomposition using the following equation:

$$C_m^j = \frac{1}{N}\sum_{n=1}^{N} A_m^j(i)$$

Where C_m^j the mth class core vector at j level decomposition, N is the number of ROI's selected to produce the class core vector and A_m^j is the fraction of biggest wavelet coefficients of the ROI's selected from the mammogram image for the class *m* at decomposition level *j*.

For the testing purpose, ROIs of the same size as specified in the training part are extracted from the test group of the mammogram and created a feature vector of the all test images and then each test ROIs are classified into appropriate group of the classification based on calculating the distance between the feature vector and the class core vector of the training set on all four different levels of the wavelet decomposition. Then this new system automatically classifies the test image in the dataset by finding the minimum Euclidean distance between feature vectors of the test image to the each of the class core vector by using Euclidean distance formulae.

$$\mathrm{Dist}(A, C_m^j) = \frac{1}{J}\sum_{1}^{J}\sum_{1}^{m}\sqrt{(A^j(i) - C_m^j(i))^2} \qquad (8)$$

Where A^m is the coefficient vector of the *j*th decomposition level for the test image, C_m^j is the class core vector for class m at decomposition level *j* and *m* is the number of classification classes. Here m is 3 (Normal, Benign and Malignant)

There is no specific dimension reduction method is applied on above classification algorithm. Here we reduce the dimension of the core and feature vector by considering only the fractional part of the highest wavelet approximation coefficients for the training and testing part of the classification. This fractional part of the approximation coefficients also contains insignificant data which are greatly affect the performance of the above algorithm. So in the second stage of the proposed classification algorithm we use a mathematical model named Principle Component Analysis for the dimension reduction of the approximation coefficients of DWT obtained in different levels of the decomposition. The transformation coefficients received on different levels of the decomposition after the wavelet decomposition is further reduced by applying the principal component analysis (PCA). The PCA analysis retrieves transformation coefficients, which are reduced in different plane directions show the highest variation is treated as feature vector. This feature vector is used for further classification. Based on this reduced feature set, we prepared a training data of all the mammogram ROIs of size 32 x 32 extracted from the Mini-mias database. The training dataset prepared for the training purpose is then divided into ten different folds of equal size and each fold is trained using different machine learning algorithms and the remaining nine folds of the dataset are used for the testing. The most common instance based machine learning algorithms namely K*, IB1 and LWL classification algorithms are used for training and testing of the reduced wavelet feature set of the dataset.

VII. RESULTS

The classification algorithms discussed above are implemented using Matlab7.8 and Weka3.6 software. Using the above algorithms, we classified all the mammogram images in the Mini-Mias database. Out of 322 different mammogram images, we extracted 329 ROIs of size 32 x 32 pixels based on the abnormality center of the abnormal images. There are multiple ROIs of many abnormal images in the dataset, which is also considered for the classification. The different categories of the mammogram images available in the dataset are shown in Table 1.

A multi-stage classification is performed on the above dataset using these algorithms. In the first stage of the classification algorithm, the entire dataset in Mini-Mias are classified into three different classes namely normal, benign and malignant using Stationary Wavelet Transformations (SWT) and Discrete Wavelet Transformations (DWT). For this classification, we created a class core vector for each class using ten percent of images randomly selected from the dataset. This class core vector is acted as a training set and remaining dataset is used for testing. The confusion matrix obtained after the classification using SWT and DWT transformations are shown in Table 1. The overall accuracy of the above two classification algorithm is also shown in Table 2.

In the second stage of the algorithm, we classified all the mammogram images in the Mini-Mias database based on the feature reduction method using PCA on approximation coefficients of Wavelet Transformations using the lazy classifiers namely K* IB1 and LWL. The classification is done into two different levels. Initially we performed a binary classification, which means the whole images in the dataset are classified into normal and abnormal images. Then all the abnormal images are then further classified into benign and malignant types too. The performance as well as accuracy of the above classification could be assessed using the parameters specificity and sensitivity. Table 3 shows the confusion matrix generated by the K*, IB1 and LWL classifiers during the classification of normal and abnormal images. Table 4 reveals the specificity, sensitivity and accuracy obtained by K*, IB1 and LWL classifier. 100 % specificity, sensitivity and accuracy were obtained for K* and IB1 whereas in LWL classifier it is 64%, 100% and 65.35% only.

The table 5 shows the confusion matrix generated by the K*, IB1 and LWL classifier for the classification of all the abnormal mammogram images classified in the previous stage into benign and malignant categories respectively. Confusion matrix shown in table 5 indicates that out of 122 abnormal images all the 69 benign and 53 malignant images are correctly identified and labeled by K* and IB1. For LWL classifier, out of 69 benign images 41 benign images are correctly classified and out of 53 malignant images 42 malignant images are also correctly identified and labeled from the dataset. The performance parameter sensitivity, specificity as well as accuracy obtained by the K* and IB1 are 100% each. But LWL classifier, sensitivity, specificity and accuracy measured are 78.89%, 60% and 68 respectively. Table 6 shows the performance parameters obtained for the above classification.

Finally, all the abnormal mammogram images are classified into respective sub categories depending on the texture feature distribution of the ROIs in the image. Table 7 shows the confusion matrix obtained by the different lazy classifiers such as K* and IBL and LWL. This table reveals that both the K* and IBL classifier exactly classified all the abnormal images into six different subcategories. This table also reveals that LWL classifier classifies all the abnormal images in a different way rather than K* and IBL. From the table we can make a conclusion that the LWL classifier classifies most of the abnormal images into circumscribed masses. The accuracy obtained by the multilevel classification is 100% for K* and IB1 classifier whereas for LWL classifier, it is only 31.71 %. The overall accuracy of the multi-level classification is shown in table 8.

TABLE II. CONFUSION MATRIX GENERATED BY EUCLIDEAN DISTANCE MEASURE CLASSIFICATION USING SWT AND DWT COEFFICIENT AS FEATURE VECTOR

SWT				**DWT**		
	Normal	**Benign**	**Malignant**	**Normal**	**Benign**	**Malignant**
Normal	164	19	24	150	22	35
Benign	14	50	05	02	67	0
Malignant	15	06	32	05	03	45
Total	193	75	61	157	92	80

TABLE III. CLASSIFICATION ACCURACY OBTAINED BY EUCLIDEAN CLASSIFICATION USING DIFFERENT WAVELET TRANSFORMATIONS

Wavelet	**Accuracy**
SWT	74.77%
DWT	79.64%

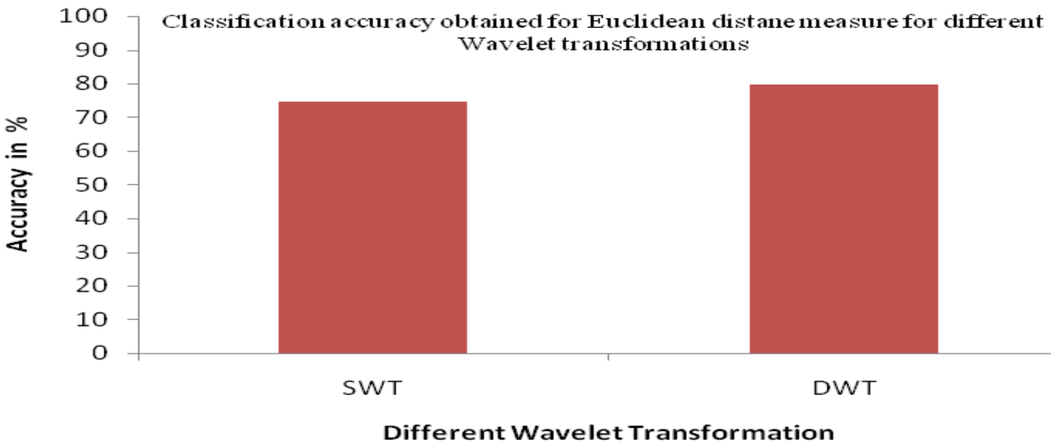

Fig. 3. Classification accuracy obtained by Euclidean distance measure for different wavelet transformations.

TABLE IV. CONFUSION MATRIX OBTAINED FOR CLASSIFYING MAMMOGRAM IMAGES INTO NORMAL AND ABNORMAL USING DIFFERENT LAZY CLASSIFIERS

	K*		**IBL**		**LWL**	
	Normal	**Abnormal**	**Normal**	**Abnormal**	**Normal**	**Abnormal**
Normal	207	0	207	0	207	0
Abnormal	0	122	0	122	114	8
Total	207	122	207	122	321	8

TABLE V. CLASSIFICATION ACCURACY IN NORMAL AND ABNORMAL CLASSIFICATION OF MAMMOGRAMS USING DIFFERENT LAZY CLASSIFIERS

Classifier	**Sensitivity**	**Specificity**	**Accuracy**
K*	100 %	100%	100%
IB1	100 %	100%	100%
LWL	64 %	100%	65.35%

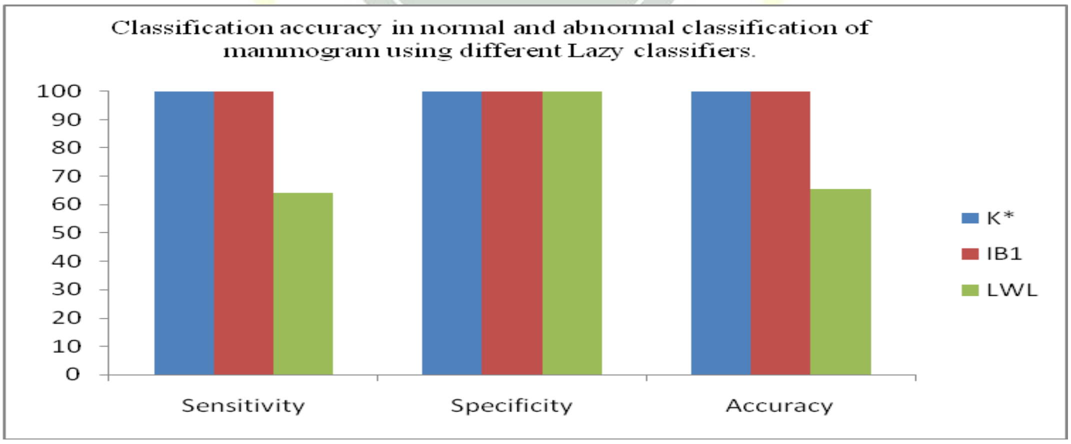

Fig. 4. Classification accuracy obtained in normal and abnormal classification of mammogram images using different Lazy Classifiers.

TABLE VI. CONFUSION MATRIX OBTAINED FOR CLASSIFYING MAMMOGRAM IMAGES INTO BENIGN AND MALIGNANT USING DIFFERENT LAZY CLASSIFIERS

	K*		**IBL**		**LWL**	
	Benign	**Malignant**	**Benign**	**Malignant**	**Benign**	**Malignant**
Benign	69	0	69	0	41	28
Malignant	0	53	0	53	11	42

Total	69	53	69	53	52	70

TABLE VII. CLASSIFICATION ACCURACY IN BENIGN AND MALIGNANT CLASSIFICATION OF MAMMOGRAMS USING DIFFERENT LAZY CLASSIFIERS

Classifier	**Sensitivity**	**Specificity**	**Accuracy**
K*	100 %	100%	100%
IB1	100 %	100%	100%
LWL	78.84%	60%	68%

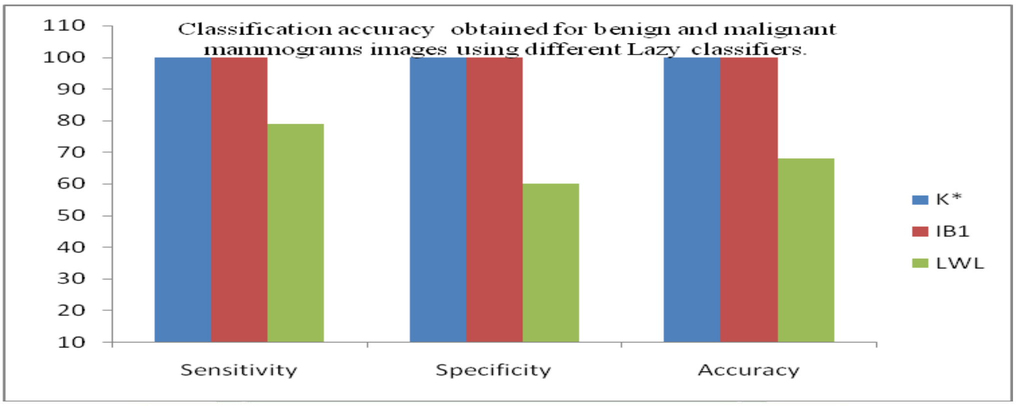

Fig. 5. classification accuracy(in %) obtained for benign and malignant mammogram images using different Lazy classifiers.

TABLE VIII. : CONFUSION MATRIX OBTAINED FOR CLASSIFYING ABNORMAL MAMMOGRAM IMAGES INTO DIFFERENT SUB CATEGORIES OF ABNORMALITIES USING DIFFERENT LAZY CLASSIFIERS

	K*						**IBL**						**LWL**					
	C	**A**	**M**	**R**	**M**	**S**	**C**	**A**	**M**	**R**	**S**	**C**	**C**	**A**	**M**	**R**	**S**	**C**
CALC (C)	30	0	0	0	0	0	30	0	0	0	0	0	8	0	0	22	0	0
ARCH (A)	0	19	0	0	0	0	0	19	0	0	0	0	1	5	0	13	0	0
ASYM(M)	0	0	15	0	0	0	0	0	15	0	0	0	4	0	0	11	0	0
CIRC (R)	0	0	0	25	0	0	0	0	0	25	0	0	0	0	0	25	0	0
MISC (M)	0	0	0	0	15	0	0	0	0	0	15	0	0	0	0	15	0	0
SPIC (S)	0	0	0	0	0	19	0	0	0	0	0	19	3	0	0	15	0	1
TOTAL	30	19	15	25	15	19	30	19	15	25	15	19	16	5	0	101	0	1

TABLE IX. CLASSIFICATION ACCURACY IN SUB CATEGORIES OF MAMMOGRAMS

Classifier	**Accuracy**
K*	100%
IB1	100%
LWL	31.71%

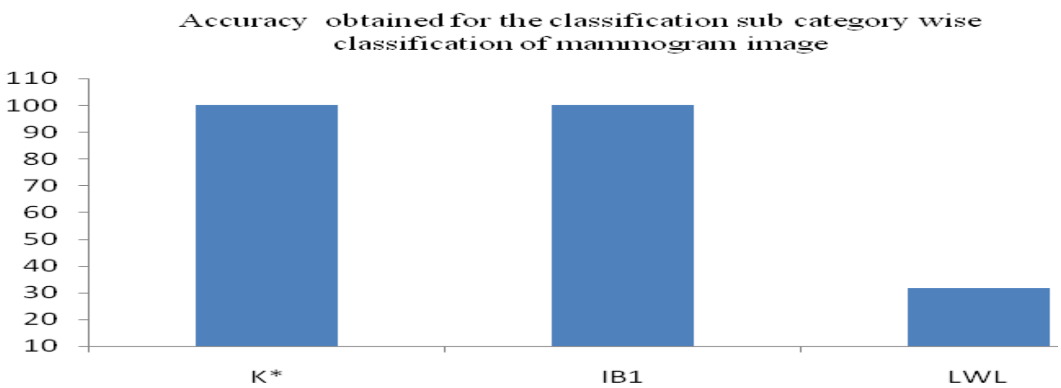

Fig. 6. Classification accuracy (in %) obtained for the classification of mammogram images into various sub categories.

Conclusion

Designing a computer aided diagnosis system which isolate cent percent breast tumor region from the original mammogram images is the one of the challenging task. In this paper we suggested such a method, which classifies all the mammogram images exactly in the Mini-Mias database. Here we started the classification using stationary wavelet approximation coefficients to discrete wavelet coefficients and then used Principle Component Analysis for reducing the wavelet coefficients dimension. After reducing the wavelet coefficient dimension, classification is done using the different lazy classifiers called K*, IB1 and LWL. We achieved 100 % accuracy on the Mini-Mias dataset using K* and IB1 lazy learning classifier. But the result obtained using LWL classifier is not up to the satisfactory level. This is due to the fact that LWL classifiers use statistical parameters for the classification. So some more characteristics other than statistical features are computed from the wavelet feature coefficients are taken into consideration for getting a better classification rates.

References

[1] G. Eason, B. Noble, and I.N. Sneddon, "On certain integrals of K. Thangavel, M. Karnan, R. Sivakumar, A. Kaja Mohideen, "Automatic Detection of Mircocalcification in Mammograms – A Reviw", ICGST_GVIP Journal, Volume (5), Issue (5), May 2005, 31-61.

[2] Essam A Rashed, Ismail A Ismail, Sherif I Zaki, "Multiresolution mammogram analysis in multilevel decomposition", Pattern Recognition Letters 28 (2007), 286-292.

[3] Tomasz Arodez, Marcin Kurdziel, Tadeusz J. Popiela, Erik O.D. Sevre, David A. Yuen, "Detection of clustered microcalcifications in small field digital mammography", Computer Methods and Programs in Biomedicine 81 (2006) 56-65.

[4] F. Moayedi, Z. Azimifar, R. Bootsani,S.Katebi, "Contourlet-based mammography mass classification", ICIAR 2007, Lecture Notes in Computer Science 4633,pp. 923-934(2007).

[5] A. Hassanien," Fuzzy rough sets hybrid scheme for breast cancer detection", Image and Vision Computing 25, 2007, pp.172-183

[6] Rajkumar K.K, "Classification of Mammogram images using stationary wavelet decomposition", NCC2011

[7] Ioan Buciu, Aleandru Gacsadi, "Directional features for automatic tumor classification of mammogram images", Biomedical Signal Processing and Control 6, 370-378, 2011, Elsevier

[8] Anke Meyer-Base, "Pattern Recognition for Medical Imaging", Elseiver Academic press,2004, ISBN 0-12-493290-8.

[9] Jong Kool Kim,Hyun Wook Park: Statistical texture features for detetction of microcalcifications in Digitizes Mammograms, IEEE transactions on Medical Imaging, Vol 18, No. 3, march 1999.

[10] E. Sakka, A Prentza, D Koutsouris, "Classification algorithms for microcalcifications in mammograms(Review)", Oncology Reports 15: 1049-1055 (2005).

[11] A.Mohd. Khuzi, R Besar, Wan Zaki, NN Ahmad, "Identification of masses in digital mammogram using gray level co-occurrences matrices", Biomedical Imaging and Intervention Journal(biij), Vol.5 No.3, 2009

[12] http://www.wiau.man.ac.uk/services/MIAS/MIAS/mini.htm: The Mammographic Image Analysis Society: Mini Mammography Database, 2008.

[13] Nason G.P. , Silverman B.W. " The Stationary Wavelet Transform and Some Statistical Applications." Tech. Rep. BS8 1Tw, University of Bristol, 1995.

[14] Beylkin G. "On the Representation of operators in Bases of compactly Supported Wavelets." SIAM Journal Numerical analysis, 29 (1992) 1716 - 1740.

[15] Shensa M.J. "The Discrete Wavelet Transform: Wedding the a Trous and Mallat Algorithm." IEEE Trans. on Information Theory, 40 (2002) 2464 - 2482.

[16] Jolliffe, I.T. "Principal Component Analysis", Springer, New York, 2002

[17] Markus Ringner ."What is principal component analysis?", Computational Biology - Primer, Nature Publishing Group(NPG), 2008

[18] Yang Song, Jian Huang, Ding Zhou, Hongyuan Zha, and C. Lee Giles, "IKNN: Informative K-Nearest Neighbor Pattern Classification", LNAI 4702, pp 248-264, springer-VerlG Berlin Heidelberg, 2007

[19] C. G. Atkeson, A. W. Moore, and S. Schaal, "Locally weighted learning," Artificial Intelligence Review, vol. 11, pp. 11-73, 1997.

[20] John G Leary, Leonard E. Trigg, K*:An instance-based Learner Using an Entropic Distance Measure, Proceedings of the 12th international conference on machine learning.

[21] C. G. Atkeson, A. W. Moore, and S. Schaal, "Locally weighted learning for control," Artificial Intelligence Review, vol. 11, pp. 75-113

[22] David W, Aha, Dennis Kibler, Marc K Albert, "Instance based learning algorithms", Kluwer Academic Publishers, Boston, Machine Learning 6, 37-66, 1991.

[23] R. Beichel and M. Sonka. Computer vision approaches to medical image analysis. Lecture Notes in Computer Science, Springer, 4241, 2006.