Abstract: Screening and early detection of breast cancer needs an automated system that identifies the breast cancer in the mammograms as early as possible. The mammogram images are preprocessed by using median filter. The median filter removes the noisy pixels in the images and improves the performance. The breast masses are segmented using Otsu segmentation process. The preprocessed images are segmented based on the intensity values of pixels. The features are extracted from the segmented image using canny edge detection algorithm and Hough transformation algorithm. Canny edge detector detects the edges in the images and features are extracted using Hough Transform. The selected features were then classified...
using Linear Discriminant Analysis. The classifiers performance is measured by measuring accuracy, sensitivity and specificity. Receiver Optimistic Curve (ROC) curve were plotted to show the performance of the classifiers. Keywords: mammogram images, ROC curve, Otsu segmentation, linear Discriminant analysis.

I. INTRODUCTION
Cancer is a class of diseases characterized by uncontrollable cell growth. Cancer harms the body when tumors are formed by the division of damaged cells uncontrollably. Tumors are the form of lumps or masses of tissue but in the case of leukemia the normal blood function is prohibited by the cancer by abnormal cell division in the blood stream. Tumors can develop and intermeddle with the major functional systems of our body. After that they release hormones that alter body function.

Breast cancer is a cancer that starts in the tissues of the breast. Nowadays the most common cancer that occurs in most of the women is breast cancer. And also it was the second commonest cancer after lung cancer. Breast cancer remains the second leading cause of cancer-related death in women. The risk of being diagnosed with breast cancer increases with age. The chances of survival significantly grow if the illness is detected at an early stage.

Breast cancer is a kind of cancer that develops from breast cells. Usually in the inside layer of milk ducts starts off the breast cancer. It may also occur in the lobules too. A malignant tumor when occurs can extend to all other parts of the body. A breast cancer that has occurred in the lobules is known as lobular carcinoma. A breast cancer that has occurred in the ducts is called ductal carcinoma.

Generally medical images are noisy. The noises in the images reduce the quality of the images. In order to improve the quality of the images we normally employ some filtering operations. Median filter is used for filtering. The median filter considers each pixel in the image in turn and looks at its nearby neighbors to decide whether or not it is representative of its surroundings. It replaces the noisy pixel value with the median of neighboring pixel values. The median is calculated by considering the surrounding neighborhood pixel values and sorting all the pixel values into numerical order and with the middle pixel value the pixel being considered is replaced.

The breast mass region is segmented using the Otsu segmentation algorithm. Otsu method is type of global thresholding in which it depend only gray value of the image. Otsu's method Clustering based image thresholding is performed automatically using Otsu method. The gray level image is reduced to a binary image can be done by using Otsu method. In this the algorithm suspects that the image consists of two classes of pixels such as foreground pixels and background pixels. By separating the foreground pixels and background pixels the optimum threshold is calculated such that their intra class variance is minimum. A threshold is defined, and then every pixel in an image is compared with this threshold. Foreground is marked in the image if the threshold value is above the pixel value. Background is marked if the threshold value is below the pixel value.

Amongst the edge detection methods introduced so far, the most widely used is the canny edge detector. It is the most rigorously defined operator. Canny edge detector can have the three criteria of fine detection, excellent localization, and solitary reaction to an edge.

The Hough transform is a feature extraction technique used in various image processing applications. The intention of the hough transform technique is by the using of voting procedure to discover inexact occurrence of objects in a certain class of shapes. For computing Hough transform the voting procedure is carried out in a parameter space called accumulator space where object candidates are obtained as local maxima that is explicitly constructed by the algorithm

Extracted feature are trained with feature matrix from the training database. The Test image is classified based on the trained feature using the LDA classifier.

There are many possible techniques for classification of data. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are two commonly used techniques for data classification and dimensionality reduction. When the within class frequencies are not equal Linear Discriminant Analysis easily handles. Linear Discriminant Analysis easily handles the case where the within-class frequencies are unequal and their it also examine the performance of the classifier on randomly generated test data.

The datasets are taken from the Digital Database for Screening Mammography (DDSM). The database contains normal and abnormal cases.

Chapter 2 focuses on the literature survey of previous works. In Chapter 3, the system design and the main
goals to be achieved using the proposed design are specified. In Chapter 4 implementation of pre processing and segmentation of optic disc regions is detailed. Chapter 5 consists of experimental results. In Chapter 6, conclusion and future work is briefed and in Appendix A, References of the proposed work are shown.

Fig1. Mammogram images

II. RELATED WORK

Shen-Chuan Tai [1] proposed an automatic CADe system that uses local and discrete texture features for mammographic mass detection. This system segments some adaptive square regions of interest (ROIs) for suspicious areas. This study also proposes two complex feature extraction methods based on co occurrence matrix and optical density transformation to describe local texture characteristics and the discrete photometric distribution of each ROI. Finally, this study uses stepwise linear Discriminant analysis to classify abnormal regions by selecting and rating the individual performance of each feature. Results show that the proposed system achieves satisfactory detection performance.

H. D. Cheng [2] proposed CAD systems for breast cancer control. Micro calcifications and masses are the two most important indicators of malignancy, and their automated detection is very valuable for early breast cancer diagnosis. Since masses are often indistinguishable from the surrounding parenchyma, automated mass detection and classification is even more challenging. This paper discusses the methods for mass detection and classification, and compares their advantages and drawbacks.

N. Szekely [3] discusses the hybrid system for detecting masses in mammographic images. The proposed approach analyzes mammograms in three major steps. First, a global segmentation method is applied to find regions of interest. This step uses texture features, decision trees, and a multiresolution Markov random field model. The second stage works on the output of the previous algorithm. Here, a combination of three different local segmentation methods is used, and then, some relevant features are extracted. Some of them refer to the shape of the object; others are texture parameters. The final decision is made using a linear combination of these features.

F. Dehghan [4] presents a computer-aided diagnosis (CAD) system for automatic detection of clustered MCs in digitized mammograms. The proposed system consists of two main steps. First, potential MC pixels in the mammograms are segmented out by using four mixed features consisting of two wavelet features and two gray level statistical features and then the potential MC pixels are labeled into potential individual MC objects by their spatial connectivity. Second, MCs are detected by extracting a set of 17 features from the potential individual MC objects. The classifier which is used in the first step is a multilayer feed forward neural network classifier but for the second step Adaboost with SVM-based component classifier. A free-response operating characteristics (FROC) curve is used to evaluate the performance of CAD system. In particular, 89.55% mean true positive detection rate is achieved at the cost of 0.921 false positive per image.

A. Mencattini [5] proposed a novel algorithm for image denoising and enhancement based on dyadic wavelet processing. In the case of micro calcifications, an adaptive tuning of enhancement degree at different wavelet scales are proposed, whereas in the case of mass detection, a new segmentation method combining dyadic wavelet information with mathematical morphology are developed. The innovative approach consists of using the same algorithmic core for processing images to detect both micro calcifications and masses. The proposed algorithm has been tested on a large number of clinical images, comparing the results with those obtained by several other algorithms proposed in the literature through both analytical indexes and the opinions of radiologists. Through preliminary tests, the method seems to meaningfully improve the diagnosis in the early breast cancer detection with respect to other approaches.

C. L. Huang [6] the purposes of this paper are to obtain the bioinformatics about breast tumor and DNA viruses, and to build an accurate diagnosis model about breast cancer and fibro adenoma. Research efforts have reported with increasing confirmation that the support vector machine (SVM) has greater accurate diagnosis ability. Therefore, this study constructs a hybrid SVM-based strategy with feature selection to render a diagnosis between the breast cancer and fibro adenoma and to find the important risk factor for breast cancer. The results show that (HSV-1, HHV-8) or (HSV-1, HHV-8, CMV) are the most important features and that the diagnosis model achieved high classification accuracy, at 86% of average overall hit rate. A Linear discriminate analysis (LDA) diagnosis model is also constructed in this study. The LDA model shows that (HSV-1, HHV-8, and EBV) or (HSV-1, HHV-8) are significant factors which are similar to that of the SVM-based classifier. However, the classificatory accuracy of the SVM-based classifier is slightly better than that of
LDA in the negative hit ratio, positive hit ratio, and overall hit ratio

M. Salmeri [7] perform the assessment of a CAD for the tumoral masses classification in mammograms by the uncertainty propagation through the system. Carrying on the work of the authors concerning the metrological characterization of the developed CAD, we validate the features extraction, features selection, and classification steps in this paper. In particular, suitable metrics such as the Receiving Operating Curve (ROC) and the Area Under ROC (AUC) are widely used in order to provide a quantitative evaluation of the performance. Finally, we implement a Monte Carlo simulation in order to provide the confidence interval for some coverage probabilities for all involved parameters. The procedure is tested on mammographic images containing both malignant and benign breast masses. In this paper, we perform the assessment of a CAD for the tumoral masses classification in mammograms by the uncertainty propagation through the system.

Carrying on the work of the authors concerning the metrological characterization of the developed CAD, we validate the features extraction, features selection, and classification steps in this paper. In particular, suitable metrics such as the Receiving Operating Curve (ROC) and the Area Under ROC (AUC) are widely used in order to provide a quantitative evaluation of the performance. Finally, we implement a Monte Carlo simulation in order to provide the confidence interval for some coverage probabilities for all involved parameters. The procedure is tested on mammographic images containing both malignant and benign breast masses.

A. Mencattini [8] In this paper, we consider uncertainty handling and propagation by means of random fuzzy variables (RFVs) through a computer-aided-diagnosis (CADx) system for the early diagnosis of breast cancer. In particular, the denoising and the contrast enhancement of micro calcifications are specifically addressed, providing a novel methodology for separating the foreground and the background in the image to selectively process them. The whole system is then assessed by metrological aspects. In this context, we assume that the uncertainty associated to each pixel of the image has both a random and a non-negligible systematic contribution. Consequently, preliminary noise variance estimation is performed on the original image, and then, using suitable operators working on RFVs, the uncertainty propagation is evaluated through the whole system. Finally, we compare our results with those obtained by a Monte Carlo method.

M. Shen [9] Accurate modeling of the multichannel electroencephalogram (EEG) signal is an important issue in clinical practice. In this paper, we propose a new local spatiotemporal prediction method based on support vector machines (SVMs). Combining with the local prediction method, the sequential minimal optimization (SMO) training algorithm, and the wavelet kernel function, a local SMO-wavelet SVM (WSVM) prediction model is developed to enhance the efficiency, effectiveness, and universal approximation capability of the prediction model. Both the spatiotemporal modeling from the measured time series and the details of the nonlinear modeling procedures are discussed. Simulations and experimental results with real EEG signals show that the proposed method is suitable for real signal processing and is effective in modeling the local spatiotemporal dynamics. This method greatly increases the computational speed and more effectively captures the local information of the signal.

K. Hoo [10] Mammography is the most effective procedure for the early detection of breast cancer. In this paper, we develop a novel algorithm to detect suspicious lesions in mammograms. The algorithm utilizes the combination of adaptive global thresholding segmentation and adaptive local thresholding segmentation on a multiresolution representation of the original mammogram. The algorithm has been verified with 170 mammograms in the Mammographic Image Analysis Society MiniMammographic database. The experimental results show that the detection method has a sensitivity of 91.3% at 0.71 false positives per image.

Alto et al. [11] proposed the use of texture, gradient and shape measures as indices for quantitative representation of breast masses in mammograms. They suggested that features that can give high accuracy in pattern classification experiments could also be used as efficient indices for CBIR.

El-Naqa et al. [12] proposed an incremental learning based relevance feedback approach, for mammogram retrieval. It requires use of Support vector machine for developing an online learning procedure for similarity learning. This was implemented on clustered microcalcification images. It is reported that the retrieval is of more effective by using this approach.

III. DATASET

The data used for the experiments in this study were taken from the DDSM [6], provided by South Florida University. The DDSM is a database of digitalized film screen mammograms. The purpose of this resource is
to provide a large set of mammograms in a digital format and compare the performance of algorithms. The DDSM contains mammograms obtained from Massachusetts General Hospital, Wake Forest University School of Medicine, Sacred Heart Hospital and Washington University of St. Louis School of Medicine. The DDSM database includes approximately 2500 cases including normal and abnormal cases.

IV. SYSTEM ARCHITECTURE

The DDSM contains mammograms obtained from Massachusetts General Hospital, Wake Forest University School of Medicine, Sacred Heart Hospital and Washington University of St. Louis School of Medicine. The DDSM database includes approximately 2500 cases including normal and abnormal cases.

IV. SYSTEM ARCHITECTURE

Fig 2. System Architecture

The system architecture shown in figure 3.1 depicts the processes of identification of breast cancer in the mammogram images. The system is trained with mammogram image sets. The mammogram image is preprocessed using median filter to reduce noise. The mammogram image is segmented using Otsu segmentation algorithm. The features are extracted from the segmented image using Canny edge detector and Hough transform. The feature vectors are stored in the training database. LDA classifier predicts whether the image is normal or abnormal from the extracted features.

V. METHOD

This section consists of four major stages: preprocessing, segmentation, feature extraction and classification.

A. IMAGE PREPROCESSING

The median filter has been proven to be very useful in many image processing applications. The median filter is far from being a perfect filtering method since it may remove fine details, sharp corners and thin lines. In median filtering, the neighboring pixels are arranged depending upon the intensity. And for the central pixel the median value becomes the new value. The pixel values in the neighborhood window are ranked according to intensity, and the middle value (the median) becomes the output value for the pixel under evaluation.

In order to perform median filtering in a neighborhood of a pixel [i,j]:
1. Sort the pixels into ascending order by gray level.
2. Select the value of the middle pixel as the new value for pixel [i,j].

B. IMAGE SEGMENTATION

The simplest method of image segmentation is called the thresholding method. This method is based on a clip-level (or a threshold value) to turn a gray-scale image into a binary image. The key of this method is to select the threshold value (or values when multiple-levels are selected). Several popular methods are used for segmenting the images in industry including the maximum entropy method, edge detection, region growing, Otsu’s method (maximum variance), and k-means clustering. Otsu method is based on establishing the optimal threshold that minimizing intra-class variance or maximizing inter-class variance.

In Otsu method the weighted within-class variance is computed as:

\[
\text{desired threshold can be calculated as}
\]

\[
2 - \]

C. FEATURE EXTRACTION

Textural and geometric features from the spatial distribution can be used to characterize the segmented image. Texture analysis can be helpful when objects in an image are more characterized by their texture than by intensity.

Feature extraction methods used were:
1. Canny edge detection
2. Hough transform

**CANNY EDGE DETECTION:**

The edge detection method proposed by canny is based on the image gradient computation. The steps of the canny edge detection method are given below:
1. Noise filtering through a Gaussian kernel.
2. Compute the gradients module and direction.

Computing the gradients magnitude and direction requires the initialization of horizontal and vertical components which can be calculated using:

gradient magnitude and orientation:
3. Non maxima suppression of the gradients module.
Its purpose is the thinning of the edges by retaining only the edge points with the highest gradient module along the direction of the image intensity variation.

4. Edge linking through adaptive hysteresis thresholding.

If the gradient at a pixel is above 'High', declare it an 'edge pixel'. If the gradient at a pixel is below 'Low', declare it a 'non-edge-pixel'. If the gradient at a pixel is between 'Low' and 'High', then declare it an 'edge pixel' if and only if it is connected to an 'edge pixel' directly or via pixels between 'Low' and 'High'.

HOUGH TRANSFORM:
The Hough transform is a method for finding lines, circles, or other simple forms in an image. The underlying principle of the Hough transform is that there are an infinite number of potential lines that pass through any point, each at a different orientation.

The purpose of the transform is to determine which of these theoretical lines pass through most features in an image – that is, which lines fit most closely to the data in the image. In the standard Hough transform, each line is represented by two parameters, commonly called r and \( \theta \) (theta), which represent the length and angle from the origin of a normal to the line.

The parameter represents the distance between the line and the origin, while \( \theta \) is the angle of the vector from the origin to this closest point. This representation of the two parameters is sometimes referred to as Hough space.

The results of this transform were stored in a matrix. One dimension of this matrix is the angles \( \theta \) and the other dimensions are the distances r.

D. CLASSIFICATION

The images are classified into Normal or Abnormal using LDA classifier. The LDA receives inputs, which can be a pattern of some kind. The input is the feature values.

The noise intrinsic to the image can be modeled by a Gaussian distribution and can be suppressed by a Gaussian filter.

Methodology:

1. Formulate the data sets and the test sets which are to be classified in original space.
2. Compute of mean of each data set and mean for entire data set by combining the data sets.
3. In LDA within class and between class scatter are used to formulate criteria for class seperability.

VI. EXPERIMENTAL RESULTS

The figure 5.1 shows an input mammogram in JPEG format of size 512 x 512 is considered for study of the system.

Fig3 Input mammogram image

Image preprocessing can be done by various filtering operations.

PSNR is most commonly used to measure the quality of reconstruction of lossy compression codecs (e.g., for image compression). The signal in this case is the original data, and the noise is the error introduced by compression. When comparing compression codecs, PSNR is an approximation to human perception of reconstruction quality. Although a higher PSNR generally indicates that the reconstruction is of higher quality.

Computing PSNR values for various filters:
Fig4 Comparison Of Filters

In spite of this, the median filter is far from being a perfect filtering method since it may remove 'ne details, sharp corners and thin lines.

The breast region are segmented from the background by using Otsu thresholding are shown in figure 5.

Fig5 Segmented image

To evaluate the performance accuracy, sensitivity and specificity are measured.

Following table shows the performance measures.

<table>
<thead>
<tr>
<th>Performance measures</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>96.667</td>
</tr>
</tbody>
</table>
Sensitivity
100
Specificity
93.333

Table 1 Performance Measures

VII. CONCLUSION
The mammogram images are segmented using Otsu thresholding after preprocessing with median filter, and a set of features extracted using canny edge detector and Hough transform a set of features are extracted and they are classified by a Linear Discriminant Analysis (LDA) classifier. The system was able to predict whether the mammogram was Normal or abnormal in nature accurately in cases of diseased ones with minimal processing time.

VIII. FUTURE WORK
The proposed system would be effective in assisting the physician in identifying whether the mammogram image is normal or abnormal. The proposed system can be enhanced in the following ways. The accuracy can be improvised by using different classifier. The classification can be extended to predict benign or malignant.

IX. REFERENCES

About Essay Sauce

EssaySauce.com is a completely free resource to help students research their academic work and learn from great essays!

View all posts by Essay Sauce

...(download the rest of the essay above)

About this essay:

This essay was submitted to us by a student in order to help you with your studies.

If you use part of this page in your own work, you need to provide a citation, as follows:


Review this essay:

Please note that the above text is only a preview of this essay. The full essay has 4061 words and can be downloaded free in PDF format, using the link above.

<table>
<thead>
<tr>
<th>Name *</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Email</td>
<td></td>
</tr>
<tr>
<td>Rating *</td>
<td>⭐⭐⭐⭐⭐</td>
</tr>
<tr>
<td>Comments (optional)</td>
<td></td>
</tr>
</tbody>
</table>
Latest reviews:

- Health essays
  - Knowledge sharing and knowledge management
  - Computer arrangement and networks

Search for student essays:

Search ...

About EssaySauce, the student essay site:

EssaySauce.com is a free resource for students, providing thousands of example essays to help them complete their college and university coursework. Students can use our free essays as examples to write their own.
Latest student essays:

Ocular disease
HUMAN action recognition
Analysing data production
Desorption study
Surfactants (surface active agents)
Islamic Finance and Its Impact on Customer Satisfaction
Persian gulf
Feminist approach (Bhumika) (notes)
What does it mean to be a Muslim woman in 21st century? (Shari’ah)
Appellate Body’s analysis under section XIV(c)

Student essay categories:

Accounting essays
Architecture essays
Business essays
Average review:

Overall rating: 0 out of 5 based on 0 reviews.
Q: Is EssaySauce.com free?

Yes! EssaySauce.com is a completely free resource for students. You can view our terms of use here.

Why use Essay Sauce?

The brightest students know that the best way to learn is by example! EssaySauce.com has thousands of great essay examples for students to use as inspiration when writing their own essays.

Is Essay Sauce completely free?

Yes! EssaySauce.com is a completely free resource for students. You can view our terms of use here.

Info:

About
Content policy
Essay removal request
Privacy
Terms of use