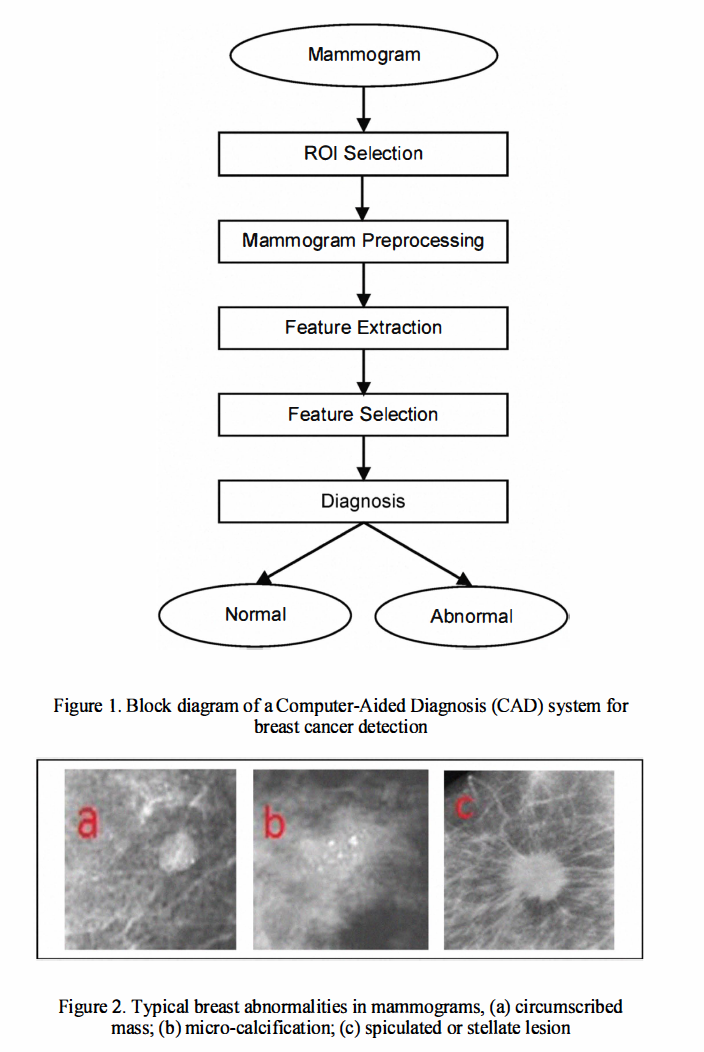
**Enhanced Accuracy of Breast Cancer Detection in Digital Mammograms using Wavelet Analysis**

Abstract - **It is an imaging system that uses low dose x-rays for examining the breasts, by the electrons reflected from the tissues.** This study was aimed at enhancing the current accuracy (diagnostic) of digital mammograms using industry standard simulation software tool, MA TLAB and the **MIAS dataset**. The technique involves identification of tumor cells to segment them in terms of different stages of the disease. We consider the process of object detection, recognition and classification of mammograms with the aim of differentiating between normal and abnormal (benign or cancerous) cells. It is reported that **dense breasts can make traditional mammograms more difficult to interpret. Although newer digital mammography techniques claim for better detection in dense breast tissues, the availability of such expensive digital mammograms is not widespread.** This problem can be minimized by analyzing different breast structures (mammograms) using the MATLAB numerical analysis software for image processing applications. **The results indicated up to 91 % accuracy, compared to 70% at present.**

The primary role of imaging technique is thus **the detection of primary lesions in the breast** [3]. Currently, the most effective method for the detection of early breast cancer **is X-ray mammography.** The underlying principle of a CAD system is illustrated in Fig. 1. In this framework, the region of interest (ROI) is first selected from the mammogram as an input 153 sample image. The sample image is then preprocessed by noise reduction and image enhancement. A large number of features, such as texture features, multi-resolution features, and shape features, are obtained using feature extraction algorithms. The redundant features are removed using a feature selection method to improve the classification efficiency.



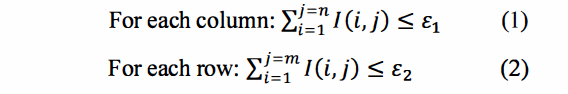
METHODOLOGY

In this study, the work is divided into three stages [9]. **The first stage covers the mammogram preprocessing and detection of the breast region.** The second stage uses the **output of the first stage to detect suspicious densities (abnormalities) in the breast region.** This stage covers feature extraction and selection. The set of features obtained from second stage is used as an input for the next stage. **The third stage is the classification stage, where a mammogram is checked for the presence of tumor based on the feature vectors obtained from the previous stage**. Any suspicious density is classified further as one having a benign or malignant tumor. This stage helps in the diagnosis of breast cancer.

A. Stage 1

In the preprocessing stage, the input image (mammogram data) is segmented to limit the search for abnormalities from the background of the mammograms in order to improve the quality of the image and reduce noise. This stage consists of the following steps:

1) Step 1**: Removal of Black Background:** For every mammogram image, the sum of intensities for all rows and columns is calculated. A set of threshold values is applied to these sums to remove the black background. A specific column or row will be removed if its sum falls below the predefined threshold value, as shown in formula (I) and (2).



where I(i,j) is the intensity of pixel, Pi,j; £ represents the predefined threshold and n, m are the dimensions of the window.

2) Step 2**: Removal of Label:** The label present in a mammogram image is removed by using the connected component technique. The biggest region is maintained for further analysis which happens to be the breast region of interest.

3) Step 3: **Removal of Pectoral Muscle**: The pectoral muscle is removed from the breast region so that only the specific region of interest (RO!) is considered. This can be established by the following two steps:

a. **Breast Orientation:** The mammogram is now divided into two sides (left and right) and the sum of each side is calculated. This helps in classifying the mammogram as left or right breast based on the biggest sum between the two sides.

b. **Pectoral Muscle Suppression**: Once the mammogram is classified to be left or right,

the pectoral muscle is estimated using a region growing method. Then, the pectoral muscle is suppressed from the breast region. To achieve better segmentation, step (1) is repeated and the normal tissues neighboring the pectoral muscle is separated.

B. Stage II

In this stage, the output from Stage I, i.e., the **breast region (ROI) is used as the input image**. Features are extracted from **the enhanced image based on the wavelet decomposition process.** The following steps summarize the feature extraction and selection stages:

1) Step 1: **Wavelet Decomposition**: The Wavelet Transfonn is probably the most recent solution to overcome the shortcomings of the Fourier Transfonn. A wavelet is a wavefonn of effectively limited duration that has an average value of zero. The decomposition of an image using wavelets results in a set of coefficients which act as features for further analysis. The type of wavelet used for this study is **Daubechies**. The outputs from this wavelet decomposition are the decomposition vector, C and the corresponding bookkeeping matrix, S. The decomposition vector consists of three detail coefficients - **horizontal detail coefficient, H, vertical detail coefficient, V and**

**diagonal detail coefficient, D.**

2) Step 2**: Extraction of Coefficients**: In this step the three detail coefficients - **H, V and D** at the decomposition scale of N are extracted from the wavelet decomposition structure (C, S). These **vectors are extracted at each scale without scale I; since the details coefficients at scale 1 contain high frequency details and noise.** The quality of the resultant image is checked and validated to 98% of the original image with the exclusion (zeroing) of scale 1 coefficients.

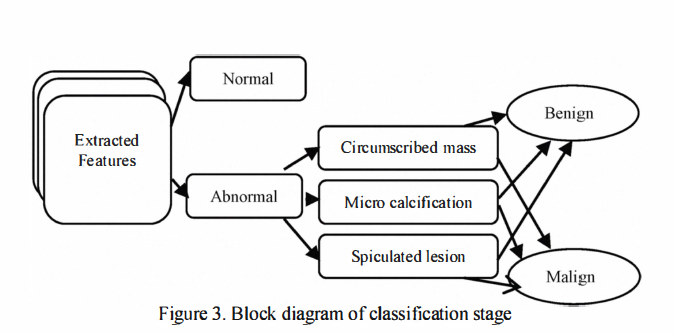
3) Step 3: **Normalization of vectors**: The coefficients vector (H, V and D) for scale 2 to scale 5 are extracted and normalized . Normalized is achieved by dividing each vector by it maximum value. This results in all the vector values being less than or equal to one. This step **is used to simplify the coefficient values.**

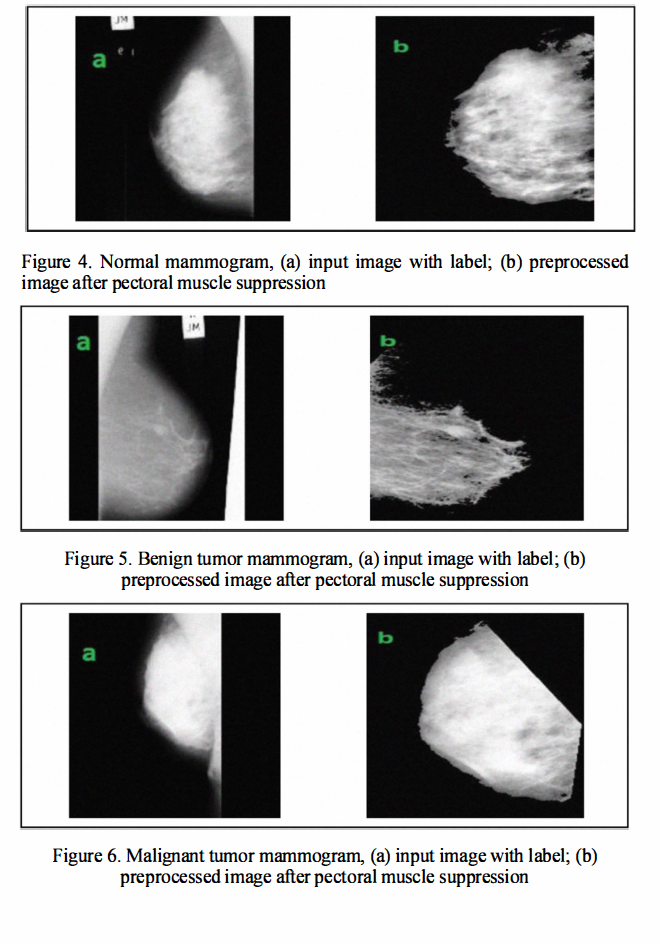
4) Step 4: **Computation of Energy**: The energy is computed for each vector by **squaring every element in the vector. These energy values are considered as features for the classification process.**

5) Step 5: **Feature Reduction**: The abnormalities present in a mammogram can be further classified into **benign and malignant tumors based on the energy values.** Since the images used have a large size, it produces high number of coefficients. This can be reduced by summing a predefined set of energy values to aid the feature reduction stage.

C)Stage III

In this stage, the output from Stage II, i.e., **the features obtained from the energy values is used for classification and diagnosis.** These feature vectors are used for identifying and differentiating a nonnal mammogram from an abnonnal case. The abnonnal mammogram can be further classified into a benign or malignant case. This classification holds good irrespective of the type of abnonnality identified. This is depicted in Fig. 3. The benign and malignant masses are differentiated through mass attributes of margin, density and location**. Round, low density masses with smooth sharply defined margins are considered benign; high density, speculated masses with poorly defined margins are considered malignant tumors.**





We have presented wavelet transform as a tool to extract features for classification purpose. After experimenting with a large number of features for each technique, our experiments show that this system can **achieve a detection rate of about 91 %.**

