



# Efficient Early Detection of Breast Cancer Detection using ANN and DWT

Parvati N. Angadi  
Research Scholar,  
Rayalaseema University,  
Kurnool, India

M. Nagendra, PhD  
Professor & HOD,  
Dept of Computer Science & Technology,  
S K University,  
Anantapuramu, India

## ABSTRACT

Excavating the prior literature shows that there has been an abundant prior studies in the area of breast cancer detection but, very little work has been put towards 'Early Detection of Breast Cancer.' In the country like US, where majority of the women has higher vulnerabilities of becoming a victim of breast cancer, as evident from history, early detection of breast cancer can play a boon in medical science. This paper therefore makes an attempt where the system is designed considering a dataset of mammogram from DDSM where feature extraction is performed using Discrete Wavelet Transform (DWT), and the feature vectors are then efficiently trained by Artificial Neural Network (ANN). The final trained results are stored in matrix and validation is performed using real time mammogram image to exhibit that the proposed model has successfully accomplished more than 90% in accuracy, sensitivity, and specificity.

## Keywords

Breast Cancer Detection, Artificial Neural Network, Discrete Wavelet Transform, DDSM

## 1. INTRODUCTION

In the U.S., breast cancer mortality is second only to that of lung cancer. For women ages 40-55, it is the leading cause of death. In 2005, an estimated 40,410 women died of breast cancer [1]. Because of its role in early tumor detection, mammography has played a substantial role in reducing this mortality rate by 20% in the last decade. While screening mammography is recognized as the most effective method for early detection of breast cancer, this modality has the following limitations:

- Up to 20% false-negative rates, where mammography fails to detect cancers.
- Up to 12% false-positive rates, where mammography detects a cancer when there is none.
- Inadequate detection of cancer in women with dense breast tissue. Mammography detects approximately 90% of tumors in women over 50, but only 60% of tumors in women under the age of 50.4 [2].
- Lack of specificity, where mammography cannot distinguish between benign and cancerous breast lesions.

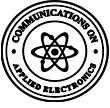
These limitations are the driving force behind efforts to refine existing mammography technologies and develop new ones offering improved detection of breast cancer. Some of the more promising new breast imaging technologies includes digital mammography (DM), computer-aided detection (CAD), breast

tomosynthesis, and new uses for breast magnetic resonance imaging (MRI) and ultrasound. Emerging technologies include contrast mammography; nuclear imaging modalities such as breast positron emission tomography (PET) and scintimammography; near infrared optical imaging platforms; new applications for breast thermography; and bioelectrical imaging technologies. Another trend is acquiring co-registered images from complementary modalities, such as full-field ultrasound coupled to DM. It is necessary to detect the presence or absence of lesions from the mammograms. However, the appearance of breast cancer is very subtle and unstable in their early stages. Therefore, doctors and radiologists can miss the abnormality easily if they only diagnose by experience. Computed aided detection technology can help doctors and radiologists get a more reliable and effective diagnosis. There are numerous tumor detection techniques that have been analyzed [3]. The key task in designing such image processing and computer vision application is the accurate segmentation of medical images. Image segmentation is the process of portioning different regions of the image based on different criteria [4]. Breast cancer image segmentation from mammography images is complicated and challenging but its precise and exact segmentation is necessary for tumor detection and their classification of tissues for early detection of abnormalities in breast. Mammography imaging is the most efficient imaging technique. Mammography is highly accurate but like most medical tests, it is not perfect [5]. On an average, mammography will detect above 80% - 90% of breast cancer in women without symptoms. Accurate image segmentation of the mammography images is very important and crucial for the exact diagnosis by computer aided clinical tools. Variety of image segmentation algorithms have been developed for mammography images. In this paper, we present a review of the methods used in mammography image segmentations. The review covers mammography images and noise reduction and image segmentation approaches. The paper concludes with a discussion on the upcoming trend of advanced researches in early breast cancer detection.

## 2. RELATED WORK

The preliminary work on Early Detection of Breast cancer has been done by Qi et al. [6] where the authors have used thermal infrared imaging (TIR). The study has presented a method for analyzing a thermal system based on an analogy to electrical circuit theory called as thermal-electric analog.

Xie et al. [7] have presented two improved Multi-static Adaptive Microwave Imaging (MAMI) methods: MAMI-2 and MAMI-C, for early breast cancer detection. MAMI [8] is one of the microwave imaging modalities based the significant contrast between the dielectric properties of normal and



malignant breast tissues and employs multiple antennas that take turns to transmit ultra wideband (UWB) pulses while all antennas are used to receive the reflected signals.

Xiao et al. [9] have proposed a method of extracting calibration waveform during detection in which the tumor is arranged randomly located in the breast tissue.

Xiao et al. [10] in their studies, the tumor detection in a twodimensional breast model with dispersive characteristic properties of the breast organisms is carried out numerically by the finite difference time domain method (FDTD).

Xiao et al. [11], in their study, the tumor detection in a two dimensional breast model with hemi-elliptical configuration is carried out numerically. The influences from the skin, breast gland and the chest wall are involved in the study. The dispersion characteristics of the breast organisms are taken into account to approach the actual properties of the human breast. Results show that the tumor could be clearly recognized from the reconstructed images created by the confocal algorithm after the appropriate signal processing.

Rejani and Selvi [12] have presented an early tumor detection algorithm from mammogram. The proposed system focuses on the solution of two problems. One is how to detect tumors as suspicious regions with a very weak contrast to their background and another is how to extract features which categorize tumors. The tumor detection method follows the scheme of (a) mammogram enhancement. (b) The segmentation of the tumor area. (c) The extraction of features from the segmented tumor area. (d) The use of SVM classifier.

Abbosh [13] has presented a theoretical model to investigate the possibility of using a hybrid imaging method for early breast cancer detection. The model exploits the high contrast in elasticity between healthy and malignant tissues. It uses an acoustic excitation using Doppler frequency to cause vibrations in the different tissues of the breast depending on their elasticity.

Halloran et al. [14] have outlined two modifications to the MIST system for the early detection of breast cancer, resulting in a quasi-multistatic MIST beamformer (multi-MIST). Multistatic MIST beamforming involves illuminating the breast with an ultra wide band (UWB) signal from one antenna while collecting the reflections at an array of antennas, as opposed to traditional monostatic MIST beamforming where only the transmitting antenna records the reflections from the breast.

Patel and Sinha [15] present a method for medical image enhancement based on the well established concept of fractal derivatives and selecting image processing techniques like segmentation of an image with self similar properties for early detection of breast cancer.

Mencattini et al. [16] present an automated procedure for bilateral asymmetry detection composed of the following steps: (1) mammography density analysis and fibro-glandular disc detection through adaptive clustering techniques, (2) analysis and implementation of bilateral asymmetries detection algorithms based on Gabor filters analysis, (3) use of a linear Bayes classifier with the leave-one-out method to assess the asymmetry degree of the two breasts, (4) metrological evaluation of the whole system through random and systematic measurement uncertainty contributions modeling.

Chiu et al. [17] has investigated the feasibility of visualization of microcalcification using opto-acoustic (photo-acoustic)

imaging (PAI) technique.

Guardiola et al. [18] have proposed a work on early breast cancer detection using microwave imaging and the worked on algorithm was termed as 3-D quasi real time algorithm that was basically inspired from the computational ability of efficient modified –Born method for assessing the early stage of the cancer.

Spandana et al. [19] implemented algorithm for 1) Image enhancement using wavelets and adaptive histogram equalization technique 2) Segmentation of masses is done using region growing technique 3) Extraction of border of the mass using canny edge detection and morphological operations. Bilateral asymmetry was detected using fluctuating asymmetry. The work presents case studies of four patients, though fourteen patient breast images are processed having different mammographic features.

Although there are abundant works on detection of breast cancer in the literature, it is evident from [6] to [19] that there is only few work which has been addressed the issues of ‘Early breast cancer detection.’

### 3. PROPOSED SYSTEM

The proposed system targets to introduce a framework that can efficiently determine the early stages of breast cancer from the radiological images. The final accuracy is enhanced using ensembling of classifiers. The potential involuntary abnormality detection method is proposed in this research work. The proposed system deploys the usage of Discrete Wavelet Transform (DWT) for extracting features to be used for classification purpose. Various rules for integrating the results of ensemble classifiers have been experimented with the purpose of enhancing the probability of choice of exact class. There exist different techniques like majority voting, weighted majority voting, min, max, product and median rule. We have compared the min, product and median rule against all of these techniques. The median rule provides better results. However, the proposed system adopts using Artificial Neural Network (ANN) for early detection of breast cancer. The most significant and novel work done in this study is use of DWT features and ensembling of classifiers using ANN. Also SOM has better feasibility in this study. The self-organizing map (SOM) is a type of artificial neural network (ANN) that is trained using unsupervised learning to produce a low-dimensional (typically two-dimensional), discretized representation of the input space of the training samples, and called a map. Self-organizing maps are different from other artificial neural networks in the sense that they use a neighborhood function to preserve the topological properties of the input space. The proposed system achieved good accuracy for the classification of mammograms as cancer stage and non-cancer stage.

The proposed approach presents an effective visual processing system for early detection of breast cancer. Digital mammogram images are taken as input and passed through the system. The system identifies malignant mammograms. The whole system is divided into six major blocks namely preprocessing, feature extraction, ensembling of classifiers, and malignancy detection. Image pre-processing techniques are necessary, in order to find the orientation of the mammogram, to remove the noise and to enhance the quality of the image [3]. Before any image-processing algorithm can be applied on mammogram, preprocessing steps are very important in order to limit the search for abnormalities without undue influence

from background of the mammogram. The complete system is illustrated in Figure 1. In first block, image of a mammogram is input to the system for pre-processing. This block perform various steps on the image and make the image ready for feature extraction. These steps are noise removal, background removal, image enhancement and pectoral muscle removal. The feature extraction is done using DWT. The technique is made time efficient by dimension reduction using SOM. Using DWT features, the classification accuracy of ensemble classifier is highly beneficial from all other benchmark techniques. The classification type of malignant mammogram is also detected as cancer stage and non-cancer stage. In this block, classifier is intelligently used for multi-classification using one against all approach.

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 accomplished by the selection of optimum feature set. Use of all the pixel values in classification creates a computational overhead because image is a large data set. To improve the efficiency of the classifier, dimensionality reduction techniques is a good approach. There are many techniques for feature extraction for example e.g., texture features, Gabor features [20], feature based on wavelet transform, principal component analysis and spectral mixture analysis. The proposed system considers using DWT features. The dimensionality reduction [21] is process of elimination of closely associated data with other data items in a set, as a result a smaller set of features is generated which preserves all the properties of the original large data set. Commonly used dimensionality reduction techniques are Principal Component Analysis (PCA), Discrete Cosine Transformation (DCT), DWT and Random Projection. In the proposed system, we have applied DWT and SOM on local blocks rather than on the complete image for dimensionality reduction. The concept of artificial neural network (ANN), can be visualized as a system of exploitation of the biological basis of neural networks, in other words, ANN is an emulation of biological neural system. The key objective of the development of neural network is the to develop a computation model which work like human brain and be able to solve hard problems in less computation time than traditional approach [22]. Artificial neural networks are useful to solve various problems like data clustering, optimization, pattern matching and classification. The structure of a neural network is like a directed graph in which different nodes, called neurons, in layers are connected to each other with some associated weights. The output of the neuron is determined through an activation function which is sum of the product of inputs with their associated weight to that neuron.

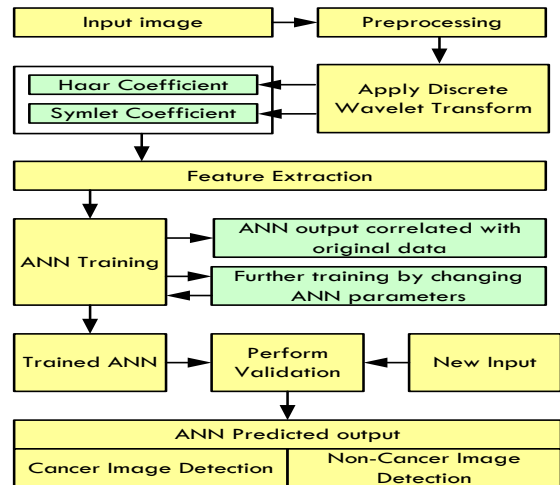


Fig 1 Architecture of the Proposed System

#### 4. IMPLEMENTATION & RESULTS

The proposed system is executed on 32 bit Windows OS with 1.84 GHz (min) processor speed and programming platform is considered in Matlab. It is also known that 64-bit architectures perform 40% faster computation than 32 bit architectures. Although the computer industry is encountering transition from 32-bit architectures to 64-bit architecture, but still majority of the users (academicians, researchers, commercial user, corporate users) still uses 32-bit windows OS and therefore the proposed study is chosen to experimented over 32 bit windows OS. The experimentation considers the use of Digital Database for Screening Mammography (DDSM) [23] as a resource for use by the mammographic image analysis research community. The proposed database is considered for evaluating 10 cases with 70 mammograms image that were used to assess the detection effect of the masses. All of these mammograms were considered as a input to the framework design in Matlab. For the purpose of the experiment, all the mammograms corresponding to non-cancer stage are segregated from cancer stage.

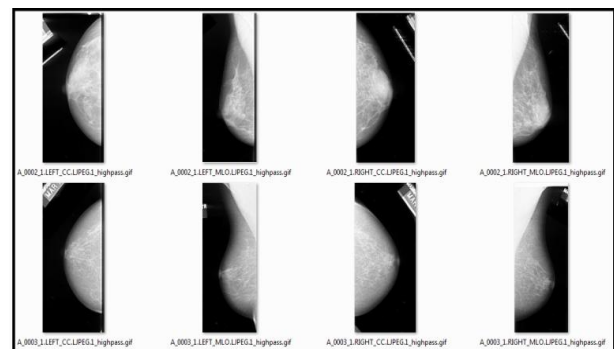
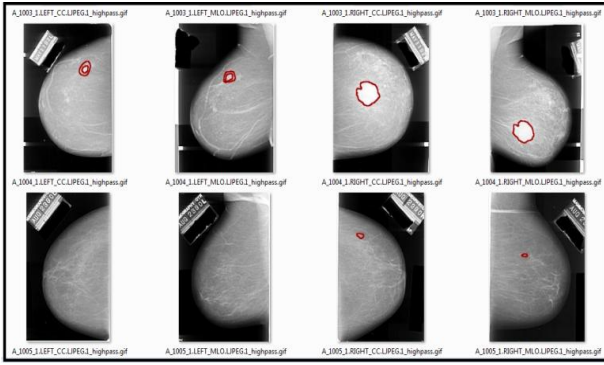


Fig 2 Mammograms for non-cancer stage



**Fig 3 Mammograms for cancer stage**

It is to be noted from Figure 2 and 3 (the mammograms for both cancer and non-cancer stage) that both categories of images are considered in the dataset and thereby used in training by Artificial Neural Network. In these mammograms, besides the breast region, where the focus of breast cancer lies, there are still other regions, including the mark region (consisting of serial number, shooting position mark, horizontal and vertical frames, etc.), and the background region. Before the formal detection process of the masses was carried out, the mark region must be removed with a seed filling method, and all the pixels of the background region is set to zero-intensity. Thus, there is only the breast region left in the mammogram, where all the detection process of proposed system, including detection of pectoral muscle, nipple, MCs and masses, could be applied.

The steps for implementing the proposed study are as follows.

**Step-1:** Feeding the input image

The input image is obtained from the DDSM database which is of size 165x363, 150x330, 159x345

**Step-2:** Median Filtering

In the next process, the input image is filtered by median filtering.

**Step-3:** Perform Discrete Wavelet Transformation

The reconstructed image from the  $n^{\text{th}}$  level (1-8) of Decomposition.

**Step-4:** Deploy Neural Network Classifier

The Neural Network Fitting Tool trains a network and evaluates its performance using mean square error and regression analysis. The steps involved in classification using nftool:

- Sub-Step 4.1: Load the input data into nftool and fix the target. In this step, the obtained data set by using the extracted DWT features is loaded as the input data and the target file where the output data is to be stored is loaded as output.
- Sub-Step 4.2: Divide the samples into training and testing. The loaded input samples are divided randomly into three groups of data namely testing, validation and training.
- Step-Step 4.3: Define the number of hidden neurons to create the Network

In this step, define the number of neurons, which will give the efficient output. This can be done by first assigning the number of neurons randomly and then, the network is trained. By

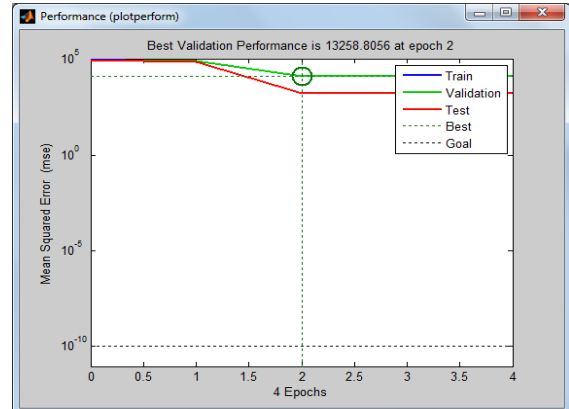
analyzing the performance plot, the number of neurons can be increased or decreased in order to get the desired output.

**Step 5:** Perform Training of the network.

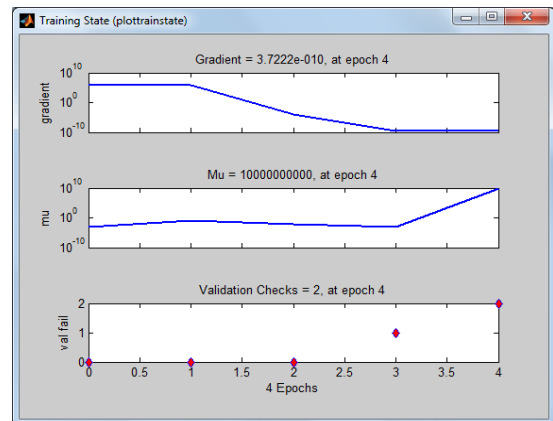
According to the input given and the number of samples assigned, the network is trained by selecting the “train” option

**Step 6:** Evaluate the network by selecting train.

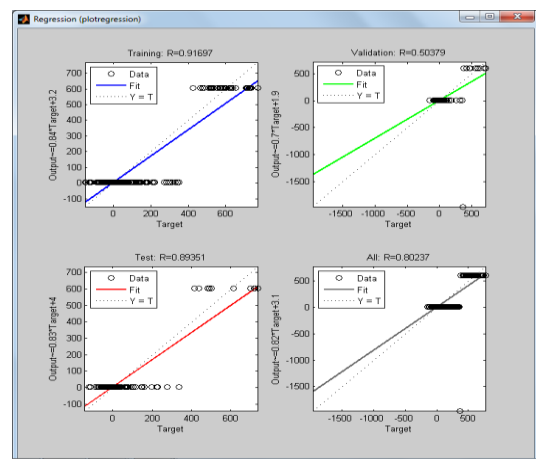
By selecting performance plot in Figure 4, the performance of the trained network is obtained which is shown in Figure 5. It indicates that the best validation performance is obtained at epoch 2.



**Figure 4 Best Validation Performance of the overall training by ANN**



**Figure 5 Training States with gradient, Mu, and validation checks**



**Figure 6 Regression Results**

We have tested the performance of these classifiers by calculating and analysis of accuracy, sensitivity and specificity for malignancy detection. These are defined as follows:

**Accuracy:** number of classified mass / number of total mass

$$(TP+TN) / (TP+TN+FP+FN) \quad (1)$$

**Sensitivity:** number of correct classified malignant mass / number of total malignant mass

$$(TP) / (TP+FN) \quad (2)$$

**Specificity:** number of correct classified benign mass / number of total benign mass

$$(TN) / (TN+FP) \quad (3)$$

Accuracy, sensitivity and specificity of different algorithms are evaluated and are given in Figure 7. The evaluation is conducted for multiple sets of real time images where thumbnails are shown in Figure 8. 20 images of cancerous origin and 20 images of non-cancerous origin are selected for the purpose of validation.

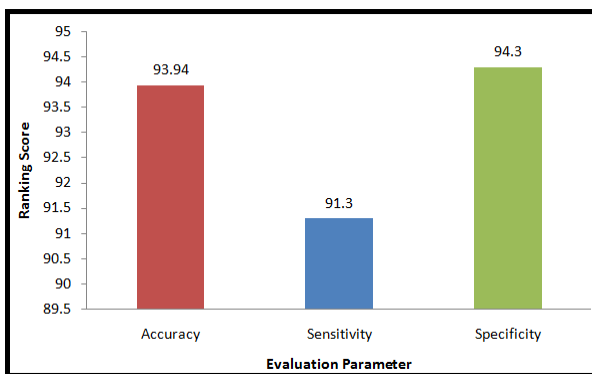


Figure 7 Performance Analysis of the proposed study

Figure 7 discusses the results of extent of cancer detection in mammogram images from the performance analysis considering accuracy, sensitivity, and specificity. The evaluated score of the validation of ANN model is considered as final result. Three important measures accuracy, specificity and sensitivity are taken to measure the performance of the proposed method. It has been observed that in case of ensembling of classifiers the accuracy is quite good as compared to single classifier. Table 4 shows the abnormality detection results of ensemble classifier using DWT features. These results are compiled using mean, median and product rule and it has been seen that the performance of proposed method approximately remain consistent in case of mean and median rule for the detection of abnormalities in the mammogram. The abnormality detection rate is also satisfactory.

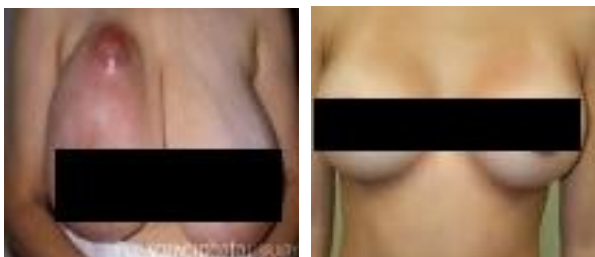


Figure 8 a) Input image with cancerous lesions b) Input image with non-cancerous stage

Table 1 Accomplished statistical results of ANN model

Parameters	Cancer Stage	Non-Cancer Stage
MSE_training	- 51.68457	- 8.02368
MSE_testing	- 160.87978	-29.15358
Regression_training	9.2802	-3.07696
Regression_testing	9.2199	-1.09436

## 5. CONCLUSION

The proposed system is developed for diagnosing the early detection of breast cancer from real time mammogram images. The system performs the diagnosis in multiple phases of ANN training. In the initial phase, preprocessing on mammogram image is done which drastically reduces the computational cost and maximize the probability of evaluation parameters e.g. accuracy, sensitivity, and specificity. In next consecutive stage, Discrete Wavelet Transform (DWT) features are extracted. These extracted features are used for categorization of mammogram into cancerous stage and non-cancerous stage. Later, the cancerous stage images are again validated using real time mammogram images. In preliminary classification phase when non-cancerous stage and cancerous stage images are separated, different classifiers are experimented but ensemble classifier is found better for the DDSM dataset. From the evaluation, it can be seen that linear classifiers are performing optimally better than the other ones. This is because of the linearity present in the dataset.

## 6. REFERENCES

- [1] [http://www.kdheks.gov/edw/ks\\_women\\_facts.html](http://www.kdheks.gov/edw/ks_women_facts.html)
- [2] <https://www.clinicalkey.com/topics/surgery/breast-cancer.html>
- [3] R.N. Strickland, Image-Processing Techniques for Tumor Detection, Marcel Dekker, 2002
- [4] P-L. Chang and W-G. Teng, Exploiting the Self-Organizing Map for Medical Image Segmentation, Twentieth IEEE International Symposium on Computer-Based Medical Systems, 2007
- [5] G. M. Brake and N. Karssemeijer, Single and Multiscale Detection of Masses in Digital Mammograms, IEEE Transactions on Medical Imaging, Vol. 18, No. 7, July 1999
- [6] H. Qil, P.T. Kuruganti, Z. Lid, Early Detection of Breast Cancer using Thermal Texture Maps, IEEE, 2002
- [7] Y. Xie, B. Guo, J. Li, P. Stoica, On multi-static adaptive microwave imaging methods for early breast cancer detection, IEEE, 2006
- [8] Y. Xie, B. Guo, L. Xu, J.Li, P. Stoica, Multistatic Adaptive Microwave Imaging for Early Breast Cancer Detection, IEEE Transactions On Biomedical Engineering, Vol. 53, No. 8, August 2006
- [9] X. Xiao and T. Kikkawa, Extraction of Calibration Waveform for Confocal Microwave Imaging for Early Breast Cancer Detection, International Symposium on Microwave, Antenna, Propagation, and EMC Technologies For Wireless Communications, IEEE, 2007



- [10] X. Xiao, X. Liu, B.Liu, Study on Microwave Imaging for the Early Breast Cancer Detection by FDTD with PML Boundary Condition, International Conference on Future BioMedical Information Engineering, IEEE, 2009
- [11] X. Xiao, T. Kikkawa, Early Breast Cancer Detection with Hemi-elliptical Configuration by UWB Imaging, IEEE, 2009
- [12] Y. Ireaneus, A.Rejani, S.Thamarai Selvi, Early detection of breast cancer Using SVM classifier technique, International Journal on Computer Science and Engineering Vol.1(3), 2009, 127-130
- [13] A.M. Abbosh, Early Breast Cancer Detection Using Doppler Frequency Shift, Proceedings of Asia-Pacific Microwave Conference, 2010
- [14] M. Halloran, E. Jones, M. Glavin, Quasi-Multistatic MIST Beamforming for the Early Detection of Breast Cancer, IEEE Transactions On Biomedical Engineering, Vol. 57, No. 4, April 2010
- [15] B. C. Patel, G.R.Sinha, Early Detection of Breast Cancer using Self Similar Fractal Method, International Journal of Computer Applications (0975 – 8887), Volume 10– N.4, November 2010
- [16] A. Mencattini, M. Salmeri, P. Casti, Bilateral asymmetry identification for the early detection of breast cancer, IEEE, 2011
- [17] T-I., Chiu, T-C., Hsiao, Shi-Bing, Luo, WanTing Tien, A feasibility study of the optoacoustic imaging of microcalcification for early breast cancer detection, IEEE 2011
- [18] M. Guardiola, S. Capdevila, J.Romeu, L. Jofre, 3-D Microwave Magnitude Combined Tomography for Breast Cancer Detection Using Realistic Breast Models, IEEE antennas and wireless propagation letters, Vol. 11, 2012
- [19] P. Spandana, K.M.M. Rao, Novel Image Processing Techniques for Early Detection of Breast Cancer, Matlab and Lab view implementation, IEEE Point-of-Care Healthcare Technologies, 2013
- [20] C-H., Wei, Y. Li, C-T. Li, Effective extraction of Gabor features for adaptive mammogram retrieval, IEEE, 2007
- [21] K. N. Batmanghelich, D.H., Ye, K.M. Pohl, B. Taskar, C. Davatzikos, and Adni, Disease classification and prediction via semi-supervised dimensionality reduction, IEEE, 2011
- [22] Sivandam SN, Deepa SN (2007). Principles of Soft Computing, Wiley India, pp. 11-29.
- [23] <http://marathon.csee.usf.edu/Mammography/Database.htm>  
1