



Breast Cancer Detection using Image Processing Techniques

PRANNOY GIRI* and K SARAVANAKUMAR

Department of Computer Science, Christ University, India.

*Corresponding author E-mail: prannoygiri@mca.christuniversity.in

<http://dx.doi.org/10.13005/ojcs/10.02.19>

(Received: April 03, 2017; Accepted: April 18, 2017)

ABSTRACT

Breast Cancer is one of the significant reasons for death among ladies. Many research has been done on the diagnosis and detection of breast cancer using various image processing and classification techniques. Nonetheless, the disease remains as one of the deadliest disease. Having conceive one out of six women in her lifetime. Since the cause of breast cancer stays obscure, prevention becomes impossible. Thus, early detection of tumour in breast is the only way to cure breast cancer. Using CAD (Computer Aided Diagnosis) on mammographic image is the most efficient and easiest way to diagnosis for breast cancer. Accurate discovery can effectively reduce the mortality rate brought about by using mamma cancer. Masses and microcalcifications clusters are an important early symptoms of possible breast cancers. They can help predict breast cancer at it's infant state. The image for this work is being used from the DDSM Database (Digital Database for Screening Mammography) which contains approximately 3000 cases and is being used worldwide for cancer research. This paper quantitatively depicts the analysis methods used for texture features for detection of cancer. These texture features are extracted from the ROI of the mammogram to characterize the microcalcifications into harmless, ordinary or threatening. These features are further decreased using Principle Component Analysis(PCA) for better identification of Masses. These features are further compared and passed through Back Propagation algorithm (Neural Network) for better understanding of the cancer pattern in the mammography image.

Keywords: Breast Cancer, Image Processing Techniques, Neural Network, Algorithm.

INTRODUCTION

Breast cancer is one of the frequent diagnosis diseases among women. It can be detected by clinical breast examination, yet the detection rate endures to be very low. Additionally, the abnormal areas that cannot be felt can be quite challenging to check using traditional techniques but

can be easily seen on a conventional mammogram or with ultrasound. Mammography is currently the best method for detecting breast cancer at its early stage. The problem with mammography images are they are complex. Thus, image processing and features extraction techniques are used to assist radiologist for detecting tumour. Features extracted from suspicious regions in mammography images

can help doctors to discover the existence of the tumour at real time thus speeding up treatment process. Detecting breast cancer can be quite a challenging job. Specially, as cancer is not a single disease but is a collection of multiple diseases. Thus, every cancer is different from every other cancer that exist. Also, the same drug may have different reaction on similar type of cancer. Thus, cancer vary from person to person. Depending on only one technique or one algorithm to detect breast cancer may not provide us with the best possible result. As one cancer differ from another, similarly every breast appears differently from another. The mammography image can also be compromised if the patient has undergone some breast surgery.

Breast Cancer has been a big topic in research field for the last two decades. It has been well funded medical research topic across the globe. Many people have been cured of it, due to early detection. Still the progress in diagnosis and treatment for it remains expensive and time consuming. Automated detection of mass still remains a difficult task, this might be due to the fact that every cancer is different like it's host and each requires customized medication to be cured. So, a lot of work is still left to be done. Some of the reasons for the challenges in automated detection as follows.

First to begin with, the object of interest can be to an extraordinary degree small, inciting to potential miss-identification.

Second, unique sizes, different shapes, and variable appropriations of microcalcifications show up in mammograms, therefore, sample matching seems to be impossible.

Third, the Region of Interest (ROI) might be of low contrast. The refinement between suspicious reaches and their enveloping tissues can be thin.

Fourth, the thick tissues as well as skin thickening, particularly in young women, cause suspicious territories to be practically undetectable.

Finally, dense tissues may easily be confused as calcifications, resulting in high false-positive cases.

Literature Survey

Detecting macrocalcification in dense breast tissue can be a difficult task as both tends to depict white pixel on the mammogram. The number of false positive cases on dense breast tissue are higher. Indicators of cancer symptom are generally, masses and microcalcifications. Detecting masses are more challenging task than detection of microcalcifications. As their size and shape varies in large variation and they often exhibit poor image contrast. The utilization of grouping frameworks in classification and pattern recognition system, in medical diagnosis, specially cancer diagnosis are growing rapidly. Evaluation and decision making based on machine learning for medical diagnosis is a key factor. Intelligent classification algorithm may help doctor in identifying symptoms that may not be possible through traditional approaches⁸.

Any Image processing and analysis applications would require a unique function for alignment of feature for classification and segmentation. Mainly texture features and statistical features are of more suitable in pattern recognition area to find this alignment.

Screening Mammography is the easiest and affordable way to diagnosis for breast cancer. The mammography image is checked through several techniques like finding edges, smoothing border, finding structures & shapes among matrixes. Finally finding the size distribution of tissues in an Image without explicitly segmenting each object⁵.

Digital mammography is the standard procedure for breast cancer diagnosis, various classification problem is applied on the digital mammography image. Various features are extracted as per standard procedure for breast cancer diagnosis. These features are calculated from the sensitive part of the breast to avoid any unwanted features to affect the classification problem. Area of tumour is calculated by the Maximum Likelihood Estimation (MLE). All the features extraction techniques are applied on the stored database image¹³.

This paper mainly studies the multiple image processing algorithms which can be extensively used for finding cancerous cells. The techniques in computer aided mammography includes image

pre-processing, image segmentation, feature extraction, feature selection and classification. Further developments are required to extract more features to find pattern in tumour to have a better understanding on them. Texture analysis method can be used to classify between benign and malignant masses by means to identify the micro-calcification in the mammography.

Research In The Field Of Cancer

Many research has been done in the field of image processing to find the cancer. Yet, the accuracy rate lies between 75% - 92%. Thus, there is still a gap of 8% to 25% of accuracy to be achieved. The new research analysis and techniques to find the cancerous cells and eradication methodology to cure the cancer from any person. However, even cancer cells have evolved them to hide from drugs and medications. As cancer cells are immortal they are not affected by the immune system. There is a research for curing the cancer tumour, the methods are as follows.

CRISPR

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) also known as Cas9 is a simple and powerful gene editing tool. Genetic engineering has allowed cancer researchers to screen the drug to target the cancer cells in efficient manner. There is also a vast door for direct treatment of cancer through gene interference or activation.

Artificial Immune System

Artificial Immune Systems resembles the natural properties of our biological immune system. Natural immune system has the property to pattern matching which is used to distinguish between normal and abnormal cell¹⁹.

Nano Technology

Nano technology can give fast and delicate location of cancer cells in the breast tissues. Empowering researchers to identify molecular changes not withstanding when they happen just in a smaller amount of cells.

The nanodevices can be programmed to annihilate infected cells and kill those infected cells¹⁷.

METHODOLOGY

Pre-processing

Mammography images are selected and converted into grayscale image of 2D matrix by the scale of 1024X1024. All the images of the database are of the same size, if image size differs from others then the image enchancement algorithm is applied to match the image resolution. These images are filtered through Noise removal algorithm. Then, they are filtered and adjusted to increase pixel intensity.

Segmentation

The image goes through thresholding process for the purpose of segmenting the ROI of the image. A Global thresholding value can be applied to remove the unwanted part of the image and segment the part with higher pixel density. Histogram can be used to check for pixel distribution. Thus, using value from all these fields the best ROI of the image is acquired⁵.

Feature Extraction

As each image is captured by different angles of the patient, the different features of these images may vary among themselves. Also, the past mammography or future mammography image of

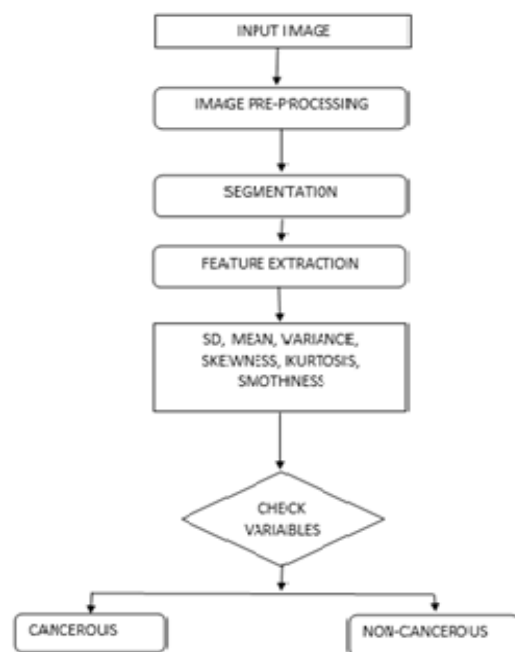


Fig. 1: Flow of the System

the patient could tell a lot about the tumour details. So, at each projected angle, a matrix, $A(i)$, with 250 images and with 15 texture and statistical features, i is the number of projections.

Edge Detection

The Region of Interest of the image is extracted from the mammography image. The different edge detection techniques Sobel, Canny, Prewitt, Roberts and fuzzy logic methods is applied to detect the edges of tumour cells.

Texture Features

In imaging terms, texture can be described as the spatial arrangement and there is a variation of intensities (gray values) within the image. Texture features have been very useful for identifying micro-calcification. The image acquisition can be used to create two different regions. The first region is the central area of the breast where the thickness is nearly uniform, and is referred to as the constant thickness region. The other consists of tissue near the edge of the breast where the thickness gradually tapers due to the breast geometry. Features are to be extracted from the segmented image for region acquiring accurate classification results. The Gray Level Concurrence Matrix (GLCM) is generated from the ROI of the image, which is a gray level concurrence matrix by calculating how frequently pair of pixels with particular qualities occurs in an image. The Spatial Gray Level Dependence matrix (SGLD) also known as the concurrence matrix of ROI can be used to classify benign and malignant cells. These features have different elements which have different values.

Neural Network

Neural Network (NN) is one of the best machine learning techniques for classification,

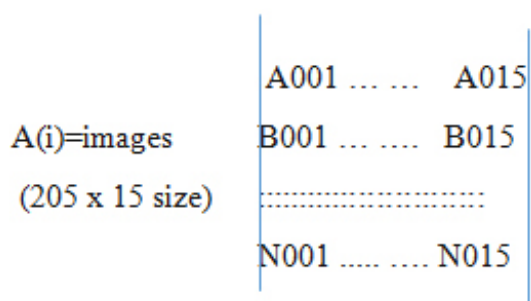


Fig 2: Matrix of the database with features

regression and pattern recognition. NN have discovered numerous applications in capacity guess and signal processing. A lot of research work on detection of cancerous cells shows that the number of false positive cases have decreased drastically. However, there are several limitations to the machine learning techniques.

First of all the few parameters to be configured for the training process through Artificial Neural Network (ANN). The process is to find the number of hidden layer, hidden nodes and learning rates. Second it takes long time to train the system through complex architecture and parameters updates in each iteration.

Third, it can be caught to neighbour minima so that the optimal performance cannot be guaranteed.

Most popular Neural Network (NN) algorithms used in classification is feed forward with back propagation algorithm. After choosing the weights of the network randomly, the back-propagation algorithm computes the necessary evaluations.

Feed Forward Networks with Backpropagation

In ANN the nodes are organized into three layers. The input layer, output layer and one

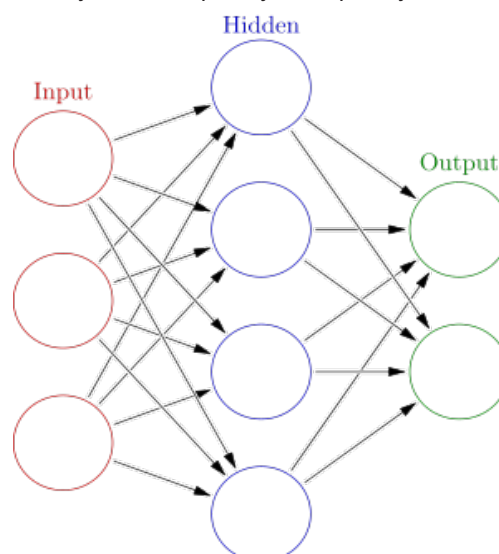


Fig. 3: Back Propagation Neural Network

or more hidden layers as shown in the figure below. Every node is interconnected by weights and data engenders starts from one layer to another through a sigmoidal activation function. The Neural Network is a supervised learning process as the data are feed to the ANN, the weights are adjusted through backpropagation procedure to meet the desired output.

All the texture features of the images are taken as input and pass through Neural Network with back propagation algorithm to find pattern among the cancerous and non-cancerous tissue. These patterns help us to understand the cancer behaviour better. It also help us predict and to identify cancer.

Proposed System

In the proposed system a new approach for the classification of texture features through neural networks. The main objective of the paper is to generate highly accurate texture features by means of curving out ROI of the mammography image. These acquired ROI are further used for extracting the texture features. These features are fed into a neural network that classifies the images as cancerous or non-cancerous. The neural network is trained using backpropagation algorithm for adjusting weight. The following steps are to be followed for the proposed system. They are:

Image Processing

The general methods for image pre-processing are divided into various branches such as image enhancement, noise removal, image smoothing, edge detection and enhancement of contrast.

Thresholding Techniques

Thresholding is an old, simple and popular technique for image segmentation.

Global Thresholding (GT) is one of the most common and mostly used techniques in image segmentation. As masses usually have greater intensity than the surrounding tissue. A global thresholding value can be found based on the histogram of the image. On the histogram, the regions with an abnormality impose extra peaks while a healthy region has only a single peak.

In the Local Thresholding (LT) a threshold value is defined locally for each pixel based on the intensity values of its neighbour pixels. Pixels belonging to the same class are not always homogenous and may be represented by different feature values. Local adaptive thresholding to segment mammographic image into parts belonging to same classes and an adaptive clustering to refine the results.

Image Segmentation

Partitioning an image into regions such that each region is homogeneous with respect to one or more properties (such as brightness, colour, texture, reflectivity, etc.). Common image segmentation methods are thresholding, edge based segmentation, region based segmentation, clustering, classifier based segmentation and deformable model based segmentation.

Feature Extraction and Selection

Feature extraction is a very important process for the overall system performance in the classification of micro-calcifications. The features extracted are distinguished according to the method of extraction and the image characteristics. The features which are implemented here is texture features and statistical measures like Mean, Standard deviation, Variance, Smoothness, Skewness, Uniformity, Entropy and Kurtosis.

Classification and Evaluation

Evaluation is done based on the acquired features and these features are compared to the respective reference to draw final conclusion.

Neural Networks

All these values of the texture features are stored and passed through the Neural Network. Back Propagation algorithm can be used to find a pattern within the datasets to automatically finding a cancer. Back propagation algorithm can be designed to self-learn and adjust the weight accordingly. As more number of data are entered into Neural Network the better the pattern recognition and accuracy.

RESULTS

The research and technology in the area of medical science is constantly updated because

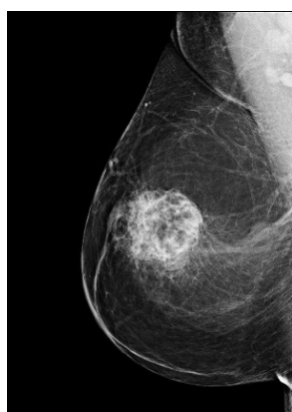
of fast changing field but the progress for detection and cure of cancer tissue remains primitive.

The above image is passed through pre-processing to obtain label free image.



Original Image

Fig. 4: Original Image



After Removing the label

Fig. 5: AfterProcess

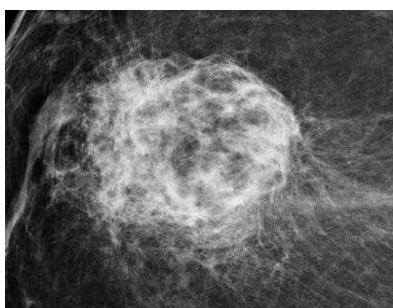


Fig. 6: ROI of the Image

The ROI of the image is cropped to run analysis to find symptoms of masses.

The above figure shows Kurtosis and Skewness from cropped image.

Accrual is the texture features of the mammography image. In the process the Statistical Features, SGLDM features, GLDC and Laws features are extracted, evaluated and added to Neural Network.

As these features are selected based on Principle Component Analysis (PCA) for selecting the decision-making parameters. These parameters are selected for subsequent classification using NN. ANN consist of a great number of neurons which are connected with each other like the neurons in our brain. The neural network has three layers input layer, hidden layer and output layer. The input layer consists of the number of features selected propionate to the number of neurons. The hidden layer have two hidden layers used for back propagation algorithm to work. The output layer will

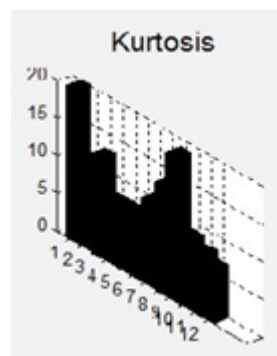


Fig. 7: Kurtosis

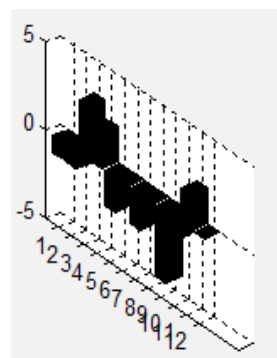


Fig. 8: Skewness

have two neurons, one representing the benign and other malignant. During the learning period, the weights are initialized between -0.1 to 1.0. In the Feed Forward Back propagation algorithm it consists of two passes through different layers of network: a forward pass and backward pass.

In the forward pass the input vector is connected to the sensory nodes of the network and its impact propagates through the network layer by layer. At last, a set of outputs is created as the genuine reaction of the system. In the forward pass, the synaptic weights of the system are altogether fixed.

During the backward pass, the weights of the connected nodes are adjusted with error correction rule. The actual reaction of the network is subtracted from a target reaction to create an error

signal. This error signal is then backpropagated from output layer to hidden layer to the input layer.

Backpropagation algorithm can be constantly enhanced and improved by considering momentum and self-learning mechanism to adjust weight. Continuous flow of information allows a network to respond not only to local gradient, but also on recent trends of surface error. Acting like a low pass channel, momentum permits the network to ignore little features in the error surface.

In backpropagation with continuous dataset, the weight is constantly changed into the direction combination of present and previous gradients. This is an adjustment of gradient whose points of interest emerge primarily when some preparation information is altogether different from most of other information. The performance of algorithm is very sensitive to the self-learning rate. If the learning rate is too high the algorithm may oscillate and unstable. If the self-learning rate is too low, the algorithm may take too much time to converge. It is practically impossible to determine the optimal setting for self-learning rate. However, the performance of the backpropagation can be improved drastically by allowing the learning rate to adapt and change based on complexity of the error and variance in the input received.

CONCLUSION

Breast cancer is one of the major causes of death among women with 1 woman affected by breast cancer out of 8 women. In the diagnosis process, due to the wide range of features associated to breast abnormalities some abnormalities may be missed or misinterpreted. There is also a number of false positive findings and therefore a lot of unnecessary biopsies may be required. Computer-aided detection and diagnosis algorithms have been developed to help radiologists give an accurate diagnosis and to reduce the number of false positives. In this study, typical steps in image processing algorithms have been extensively studied. The techniques in the field of computer aided mammography include image pre-processing, image segmentation techniques, feature extraction, feature selection, classification techniques and features for mammograms. Texture

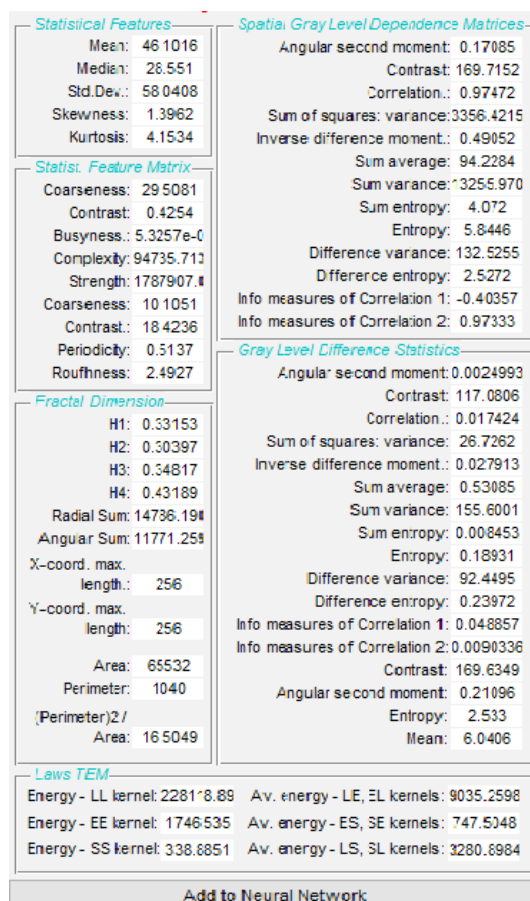


Fig. 9: Texture Features in display box

feature are obtained to distinguish between normal cell and cancerous cell. Cancer being one of the oldest disease and lot of research has been carried out in this field. Cancer is not a single disease rather a collection of multiple diseases thus a single

medicine to cure cancer is not possible. The key input to cure a cancer is customization of medication based on the type of cancer and it can be cured if found it early.

REFERENCES

1. Yui Man Lui, Rita I. Freimanis: "A Novel Approach to Microcalcification Detection Using Fuzzy Logic Technique," *International Journal on Graphics Vision and Image Processing*, **9**, pp: 69-74, 2009.
2. Ranjeet Singh Tomar, Tripty Singh, Dr SulochanaWadhvani, Dr. Sarita Singh Bhadoria: "Analysis of Breast Cancer Using Image Processing Techniques," *International Journal of Soft Computing and Engineering (IJSCE)*, ISSN: 2231-2307, **4**(2), May 2014.
3. Tobias ChrisiianCahoonA Melanie A. Suttorf James e. Bezdek: "Breast Cancer Detection Using Image Processing Techniques," *International Journal of Computer Applications*, **87**(14), February 2014.
4. Dr. M. Karnan¹, K. Rajiv Gandhi: "Diagnose Breast Cancer Through Mammograms, Using Image Processing Techniques and Optimization Techniques", Fifth International Conference on Computational Intelligence, Communication Systems and Networks, 2013.
5. Albert Gubern-M´erida, MichielKallenberg, Ritse M. Mann, Robert Mart´y, and Nico Karssemeijer: "Breast Segmentation and Density Estimation in Breast MRI," *IEEE Journal of Biomedical and Health Informatics*, **19**(1), January 2015.
6. Mussarat Yasmin, Muhammad Sharif and Sajjad Mohsin: "Survey Paper on Diagnosis of Breast Cancer Using Image Processing Techniques," *Research Journal of Recent Sciences*, **2**(10), 88-98, October 2013.
7. Cristina Juarez, Maria Elena Castillo, and VolodymyrPonomaryov: "Classification of Masses in Mammography Images Using Wavelet Transform and Neural Networks,"
8. K.SaravanaKumar and A.M. Arthanariee "Evaluate the Multiple Breast Cancer Factors and Calculate the Risk by Software Tool Breast Cancer Risk Evaluator" *Indian Journal of Science and Technology*, **8**(S7), 686–691, 2015.
9. S. Sahan, K. Polat, H. Kodaz, and S. Güne^o, "A new hybrid method based on fuzzy-artificial immune system and k-nn algorithm for breast cancer diagnosis.," *Comput. Biol. Med.*, **37**(3), pp. 415–23, 2007.
10. Y. H. Chang, BinZheng, Xiao-Hui Wang, W.F.Good,"Computer-Aided Diagnosis of Breast Cancer Using Artificial Neural Networks: Comparison of Backpropagation and Genetic Algorithms," *IEEE*, 1999.
11. Verma, B., "Novel network architecture and learning algorithm for the Classification of Mass Abnormalities in Digitized Mammograms," *Elsevier*, **42**(1), pp 67-79, 2008.
12. Peter McLeod, BrijeshVerma, "Clustered Ensemble Neural Network for Breast Mass Classification in Digital Mammography", WCCI 2012 IEEE World Congress on Computational Intelligence, June, 10-15, 2012 - Brisbane, Australia.
13. B. Verma and P. Zhang, "A novel neural-genetic algorithm to find the most significant combination of features in digital mammograms," *Applied Soft Computing*, **7**(2), 2007, pp. 612-625.
14. Arden SagiterrySetiawan, Elysia¹, Julian Wesley and YudyPurnama, "Mammogram Classification using Law's Texture Energy Measure and Neural Networks," *ICCSCI* 2015, 59.
15. Sheshadri HS &Kandaswamy A, "Computer aided decision system for early detection of

- breast cancer", *Indian J Med Res* **124**, pp. 149-1, 2006.
16. Heang-Ping Chany, Berkman Sahiner, Nicholas Petrick, Mark A Helvie, Kwok Leung Lam, Dorit D Adler and Mitchell M Goodsitt : "Computerized classification of malignant and benign microcalcifications on mammograms: texture analysis using an artificial neural network," *Iopscience, Phys. Med. Biol.* **42**, 549-567, (1997).
 17. Jie Chen, Stephen T.C. Wang: "Nanotechnology for genomic signal processing in cancer research - A focus on the genomic signal processing hardware design of the nanotools for cancer ressearch", *IEEE Signal Processing Magazine*, **24**(1), pages 111 – 121, 2007.
 18. G. Magna, S. VelappaJayaraman, P. Casti, A. Mencattini, C. Di Natale, E. Martinelli : "Adaptive classification model based on artificial immune system for breast cancer detection", AISEM Annual Conference, 2015 XVIII.
 19. J. Al-Enezi, M.F. Abbod, S. Al-Sharhan , "Artificial Immune Systems Applications in Cancer Research" IEEE Xplore.