

# Breast Cancer Detection using Histopathology and Thermography Images

Swathy.C.K<sup>1</sup>, Resmi K Rajan<sup>2</sup>

P.G. Student, Department of ECE, Vidya Academy of Science and Technology, Kaiparambu, Kerala, India<sup>1</sup>

Assistant Professor, Department of ECE, Vidya Academy of Science and Technology, Kaiparambu, Kerala, India<sup>2</sup>

**ABSTRACT:** Nowadays, Breast cancer diseases are common cancer among women. It is doubtlessly an abominable and life-threatening disease. If it is distinguished and by rights treated in its early stage, the chance of healing increases. Different imaging techniques are there which plays a critical role in the diagnosing of breast cancer. This paper presents breast cancer detection in digital thermography and histopathology images using Image Processing Techniques by Artificial Neural Networks. A database of 90 previously verified patient cases are employed and randomly partitioned into one set for training and five independent sets for testing. Gray Level Co-occurrence Matrix (GLCM) and chip histogram features extracted from the known thermography and histopathology images are used to train Artificial Neural Network based detection system. In Testing Phase the extracted features of known and unknown thermography and histopathology images are compared for classification of images into malignant and benign. Feed-forward back propagation Artificial Neural Network structure had been trained for detection. The performance is evaluated on the basis of five-fold cross validation method Mean Square Error (MSE). The verification results show that the proposed algorithm gives the best classification results of thermography and histopathology images using Artificial Neural Network classifier and an accuracy of 98.2 % and 96.5% respectively.

**KEYWORDS:** Breast cancer, Thermography, Histopathology, GLCM, Chip histogram.

## I. INTRODUCTION

In India, the second most common disease is cancer. Breast cancer is highly predominant among women. Early detection is the important key for treatment; according to numerous pathological studies, more abnormalcies are benign at the beginning stages. In recent years, many studies and wide research have been done to determine techniques for early detection of breast cancer that have more eminent precision and increased accuracy [1,2]. There are many techniques such as mammography, ultrasound and MRI are used for breast cancer diagnosis. Each has considerably-known advantages and disadvantages. Mammography is regarded the gold standard for sensing of breast cancer, but still has limited sensitivity, specificity, and among restrictions as a screening tool. Ultrasound is used as an adjunctive tool in diagnosing women with thick breasts along with mammography. The demerits of MRI are that it is not in effect at detection of ductal carcinoma in situ, slow, expensive, and goes wrong to show all calcifications. However, these sensory systems all suffer from one or more usually disadvantage of radiation exposure, breast compression, only structural imaging, and limitations with breast density [3, 4]. In order to remove the disadvantages of the present methods, infrared thermography can be used for detecting breast abnormalities with thermal images.

Infrared thermography uses an extremely sensitive infrared camera to get the image of the temperature distribution in the human body. This imaging technique due to the use of non-ionizing radiation (passive), non-contrast raised injection, lack of contact, and non-invasiveness is of value in many biological and medical applications compared or in combination with other imaging methods. Temperature dispersion patterns in thermography lead to the physiological reading of tissues (temperature variations based on interaction patterns) which can bring out suspected areas of cancer tissue, even when anatomical abnormalcies on the tissue surface are not shown and the tissue seems

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 4, April 2017

healthy [5,6]. Research on the temperature, dynamic changes of the body surface after a cold stimulus to patients has been taken on which may be advantageous for physicians, particularly for early disease diagnosis. Stimulation of the sympathetic nervous system causes a constriction of the normal blood vessels in the breast. The blood vessels that provide a cancerous tumor will resist constriction or fail to constrict. Therefore, dynamic digital thermal subtraction gives us the possibility of isolating the vessels that provide nutrients to the tumor. This method finally raises the suspiciousness that a cancerous tumor may be present [7].

The Histopathological plays a serious role in the treatment of sickness. Histopathology is the examination of tissues affected by a disease, may be abominably helpful in making a diagnosing and in deciding the asperity and progression of a sickness. It also used to interpret the normal structure and function of various tissues is a necessity of interpreting the changes that occur during disease. Histopathological examination is the most dependable procedure and believed as the gold standard for cancer diagnosis and grading. In this examination, a minuscule part of a tissue is taken out from a patient by surgery and analyzed under a microscope by pathologists. In this examination, pathologists should be able to distinguish the variations in cellular structures and the deformations in tissue distribution. Moreover, tissue training procedures such as staining and sectioning operations may present noise and artifacts to the image. Therefore, histopathological examination is subject to a considerable amount of Intra- and inter-observer variability. To meliorate the reliability of cancer diagnosis, it is significant to develop computational tools for automatized cancer diagnosis that operate on quantitative measures[8,9].

The paper gives an idea to implement an effective classification technique to detect and locate signs of possible breast cancer or abnormal heat readings and patterns in a body through thermal imaging (breast thermography). As cancerous cell has a more eminent temperature as compared to a normal cell, these temperature deviations are reflected onto thermal images. Similarly, Mitosis count in histopathological slides, play an important role in invasive breast cancer grading using the Nottingham grading system. Instead of analyzing the biopsy slide manually, automated mitotic cell detection method is used to reduce the workload of the pathologists. The diagnosis is done by applying the techniques of digital image processing on histopathological and thermogram images. The histopathological image is biopsy slide image by using high-power microscopy and Thermogram is an infrared thermal image. This work is carried out with the help of a database of normal and abnormal histopathological and thermography images.

Paper is organized as follows. Section II describes automatic text detection using morphological operations, connected component analysis and set of selection or rejection criteria. The flow diagram represents the step of the algorithm. After detection of text, how text region is filled using an Inpainting technique that is given in Section III. Section IV presents experimental results showing results of images tested. Finally, Section V presents conclusion.

## II. LITERATURE SURVEY

During the past few years, several contributions have been made in literature considering the application of pattern identification techniques for breast cancer diagnosis in tissue level. Rejani et al. [10] proposed a pattern recognition process to classify the breast tumor. They used the image segmentation to segment the breast tissue corresponding to the tumor. The discrete wavelet transform (DWT) as a feature extraction method to select various features from the segmented images and used an SVM classifier to separate the breast tissue corresponding to the features and attained an accuracy of 88.75%. Martins et al. [11] mentioned the technique for detection of mass on digitized mammograms. They used K-means clustering algorithm for image segmentation and GLCM feature extraction method is used to describe and analyze the texture of segmented structures in the image. SVM classifier used for classification of structures, which separate them into two groups; using shape and texture descriptors: masses and non-masses. The classification accuracy obtained from that method was 85%.

Marina Milosevic et al. [12] present a system based on feature extraction techniques and image segmentation techniques for detecting and diagnosing abnormal patterns in breast thermograms. GLCM are used to evaluate the

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 4, April 2017

effectiveness of textural information possessed by mass regions and differentiating abnormal from normal tissue is investigated using an SVM classifier, Naive Bayes classifier, and KNN classifier. Five-fold cross-validation method and Receiver operating characteristics are used to evaluate the performance of the classifier. The verification results show that the proposed algorithm gives the best classification results using K-Nearest Neighbour classifier and an accuracy of 92.5 %.

Satish Saini et al. [13] proposed a breast cancer detection in digital mammography using Image Processing Techniques by Artificial Neural Networks. GLCM features extracted from the segmented mammogram images and Feed-forward back propagation and Cascade forward back propagation Artificial Neural Network structures had been trained for detection. The performance is evaluated on the basis of Mean Square Error (MSE). Angshuman Paul et al. [9] propose a fast and accurate approach for automatic mitosis detection from histopathological images. They employ area morphological scale space for cell segmentation. The scale space is made in a novel manner by constraining the scales with the maximization of relative entropy between the cells and the background. This results in accurate cell segmentation. The segmented cells are classed in a mitotic and non-mitotic category using the random forest classifier.

### III. PROPOSED METHODOLOGY

The designed breast cancer detection system using ANN works in two phases: Learning/Training Phase and Recognition/Testing Phase. In Learning/Training Phase the ANN is trained for recognition of breast cancer from known benign and malignant Histopathology and thermography Images. Features, extracted from the known Histopathology and thermography images, are stored in the knowledge base and given as input to the ANN based Diagnosis System. In Testing/Recognition Phase the extracted features of known and unknown Histopathology and thermography images are compared for classification of images into malignant and benign using ANN. Training phase and testing phase consists of pre-processing, segmentation and feature extraction processes. Finally classify the patient results into benign or malignant.

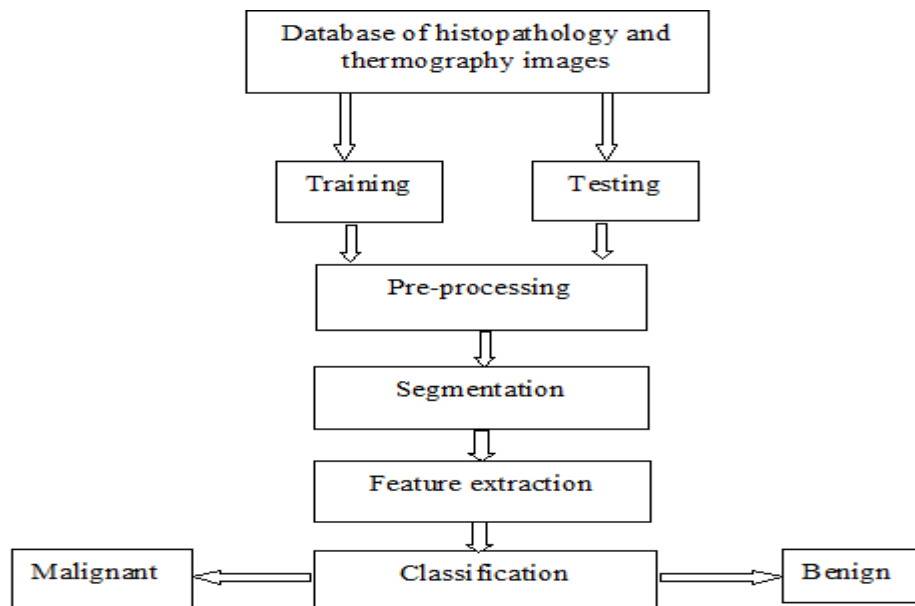


Figure 1: Block diagram of proposed method

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 4, April 2017

## A. DATASET

The thermography image data used in this paper are obtained from the publicly available Iranian Thermography dataset. The thermography images are available in the website /irthermo.ir/. The images are acquired by using FLIR E30 Infrared Camera having a spectral response of 8m to 14m and 160X120 IR resolution. The images are received in JPEG Format. This website provided a set of 150 thermography images. The 90 thermography Images were selected for an ANN based detection system. The thermography images consist of 54 benign and 36 malignant images. The histopathology images were obtained from the MITOS-ATYPIA-14 dataset. The pathology slides are stained with standard hematoxylin and eosin (H&E) dyes and scanned by two slide scanners: Aperio Scanscope XT and Hamamatsu Nanoscooper 2.0-HT. Aperio scanner image data at X40 magnification utilized for experiments. These slide frames are 24-bit RGB bitmap images with a dimension of  $1539 \times 1376$  pixels and resolution of 4.073 pixels/ $\mu\text{m}$ .

## B. FEATURE EXTRACTION

The medical image categorization procedure consists of two steps: Texture Feature Extraction and Classification. Feature extraction is the most significant step in breast cancer detection. A feature is used to denote a piece of data which is relevant for working the computational task linked to a certain application. Some of the most usually used texture measures are deduced from the Gray Level Co-occurrence Matrix (GLCM) and Chip histogram based texture features.

### Gray Level Coherence Matrix (GLCM)

Gray Level Coherence Matrix (GLCM) is widely employed to characterize texture images. In our experiments, four adjacency directions  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$  and 8 gray levels are applied to figure the GLCM. On the GLCM, 13 Haralick parameters are computed [14]: angular second moment, contrast, correlation, sum of squares, variance, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, information measures of correlation 1, and information measures of correlation 2. Finally, we get a final feature vector by calculating the mean of the 13-dimensional feature vectors in the four directions.

### Chip histogram features

For the image with number of gray levels as 'L', so there exists the gray vector, which varies from 1 to L. The histogram of the image is given as: Histogram =  $\mathbf{h}(\mathbf{r}_k) = \mathbf{n}_k$ , where ' $\mathbf{r}_k$ ' is the  $k^{\text{th}}$  gray level,  $\mathbf{n}_k$  the number of pixels in the image having gray level ' $\mathbf{r}_k$ ' and  $\mathbf{h}(\mathbf{r}_k)$  is the histogram of a digital image with gray level  $\mathbf{r}_k$  (Kaiwei et al. 2011). The image having the size of  $M \times N$ , the Gray level probability function is defined as:

$$\text{Gray level probability} = P(\mathbf{r}_k) = \mathbf{h}(\mathbf{r}_k) / M * N$$

The various chip histogram features are extracted from each cancer and non-cancer segmented image like mean, variance, kurtosis, skewness, energy, entropy [15].

## C. CLASSIFICATION

The training and testing of the classifier for textural feature set were performed using the cross-validation methodology. Classification of the normal and cancer cases was conducted by using the ANN (Artificial Neural Network) classifier with diagonal covariance matrix estimate. In this paper, breast cancer detection based on ANN has been designed by Feed-forward back propagation Network model. Levenberg-Marquardt back-propagation algorithm was used to train both the networks for classification of thermogram and histopathology images into benign and malignant.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 4, April 2017

The most popular Neural Network algorithms is back propagation algorithm, which consists of four main steps:

1. Feed-forward computation.
2. Back propagation to the output layer.
3. Back propagation to the hidden layer.
4. Weight updates.

Initial weights are commonly set at just about random numbers and then they are corrected during NN training. After selecting the weights of the network arbitrarily, the back propagation algorithm is used to compute the essential corrections. During the NN training weights are updated subsequent iterations. If the results of NN subsequent weights updates are better than the former set of weights, the new values of weights are saved and iteration goes on. The algorithm stops when the value of the erroneous function has become sufficiently small.

Figure 2, shows how the breast mass characteristic features are fed to the Neural Network Classifier. In the feature extraction stage, a valid feature vector of thirty-eight features (are extracted from ROI images of thermogram and similarly a valid feature vector of seventy-six features are extracted from the red and green channel of segmented cells of histopathology image to represent the breast mass characteristic features. This feature vector is passed to the Neural Network (NN) Classifier, which is the last stage.

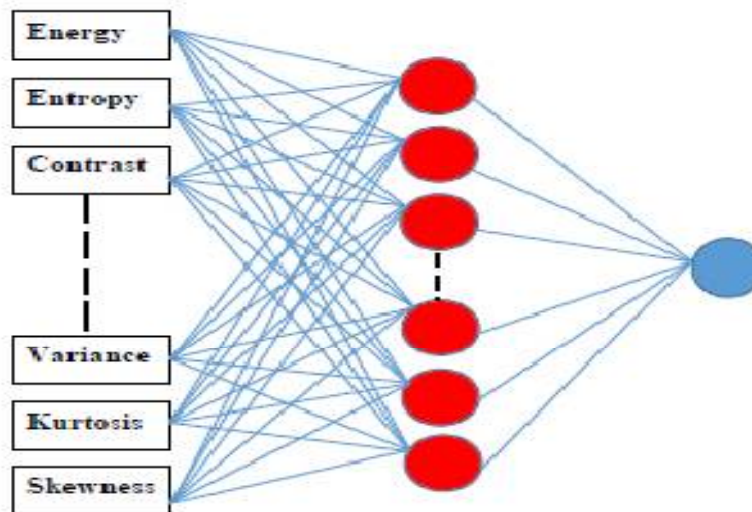


Figure 2: Artificial Neural Network

## IV. RESULTS AND ANALYSIS

A total of 90 thermography and histopathology images of the breast (54 benign and 36 abnormal patterns) is analyzed. The twenty textural features based on GLCM and chip histogram based features are used to classify breast into malignant by containing masses and benign by containing normal tissues. Every texture feature value was computed as the average of values got from the four GLC matrices matching to four different directions ( $\theta = 0^\circ, 45^\circ, 90^\circ, \text{ and } 135^\circ$ ) and one distance = 1 pixel. The classifier employed in this research is an Artificial Neural Network.

The performance in identification can be measured by the following factors: Accuracy (AC), Sensitivity (SE) and Specificity (SP) of detection. They are determined as follows:

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 4, April 2017

1. Classification accuracy is dependent of the number of samples correctly classified.

$$AC = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Sensitivity is a proportion of positive cases that are well detected by the test.

$$SE = \frac{TP}{TP + FN}$$

3. Specificity is a proportion of negative cases that are well detected by the test.

$$SP = \frac{TN}{TN + FP}$$

Where, TP is the amount of true positives, FP is the number of false positives, TN is the number of true negatives, FN is the count of false negatives.

Figure 3a and 3b shows the confusion matrices for the classification results received with ANN classifier. The confusion matrix is defined between target class and output class. The diagonal cells in each table show the amount of cases that were correctly classified, and the off-diagonal cells show the misclassified cases. The blue cell at the bottom right shows the total percent of correctly classified cases (in green) and the total percent of misclassified cases (in red). Gray cells numbered 1 and 2 show sensitivity and specificity, respectively. Gray cells in the third column of Confusion Matrix represent the Precision or Positive Predictive Value (PPV) and Negative Predictive Value (NPV). The positive and negative predictive values are the proportions of positive and negative answers of classification tests that are true positive and true negative results. Mathematically, PPV and NPV can be expressed as:

$$PPV = \frac{TP}{TP + FP}$$

$$NPV = \frac{TN}{TN + FN}$$

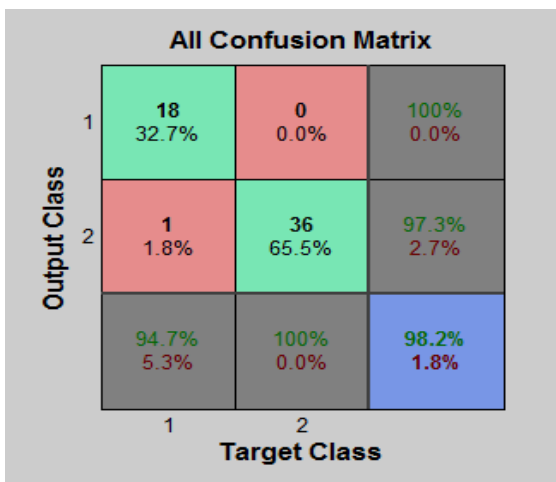


Figure 3a:ANN classification of breast Thermography images.

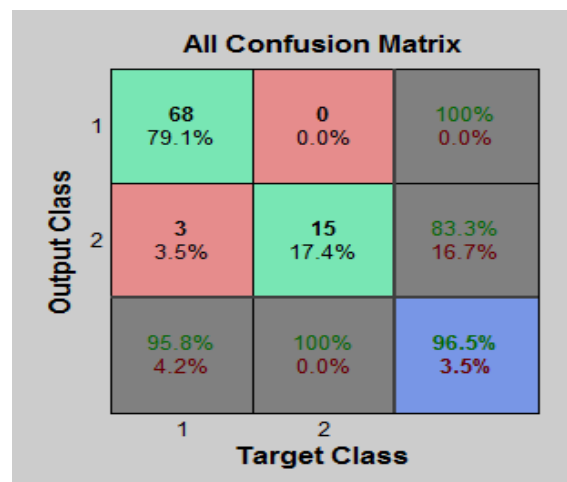


Figure 3b:ANN classification of breast Histopathology images.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 4, April 2017

In Figure 3a, the results in confusion matrix indicate that the accuracy ratio of 98.2%, sensitivity of 94.7% and Negative Predictive Value (NPV) of 97.3% are obtained by classifying thermography using ANN classifier and in Figure 3b results indicate that the accuracy of 96.5%, sensitivity of 95.8% and Negative Predictive Value (NPV) of 83.3% are obtained by classifying histopathology using ANN classifier. High sensitivity and high NPV of ANN classification of thermography and histopathology indicate that texture features can be parameters for breast cancer screening.

The Performance graph of Thermography training process for Feed forward Backpropagation is shown in Figure 4a. The MSE for training, testing and validation of thermography is shown in corresponding curves. Similarly the Performance graph of Histopathology training process for Feed forward Backpropagation is shown in Figure 4b and MSE for training, testing and validation of histopathology is shown in corresponding curves.

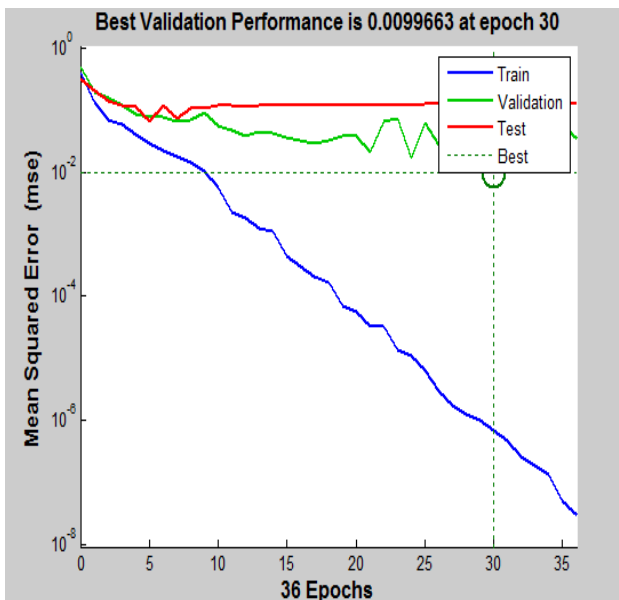


Figure 4a: Performance graph of Thermography

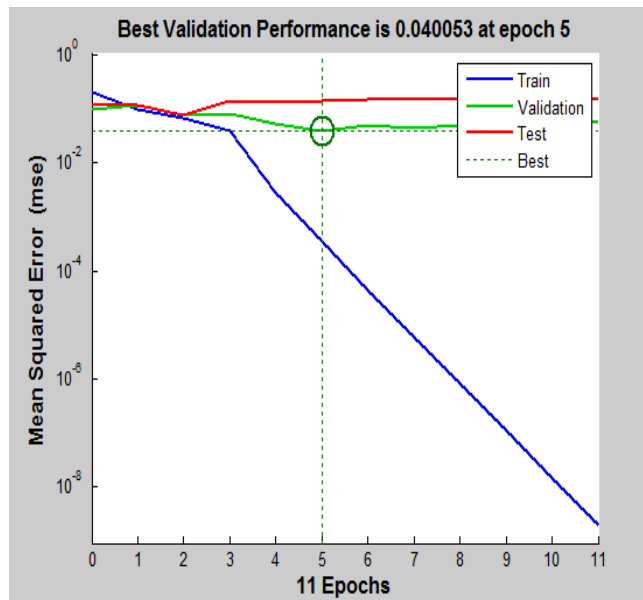


Figure 4b: Performance graph of Histopathology

## V. CONCLUSION

The proposed system is an effective technique for detection of breast cancer in early stages. The project proposes an approach which deals with Analysis of Breast Thermograms based on colour analysis for abnormality Detection based on K-means Clustering and mitotic segmentation and abnormality detection in histopathological images based on FuzzyC-Means Clustering. The proposed technique is expected to reduce the workload of pathologists and thermologists when they evaluate the cancer grade.

## REFERENCES

- [1] Liberman L, Abramson AF, Squires FB, Glassman JR, Morris EA, Dershaw DD, "The breast imaging reporting and data system: positive predictive value of mammographic features and final assessment categories", *AJR Am J Roentgenol*, 171(1):35-40, 1998.
- [2] Liberman L, "Clinical management issues in percutaneous core breast biopsy", *Radiol Clin North Am.*, 38(4):791-807, 2000.
- [3] Rane KP, Joshi RR, Chaudhari LA, " IR Imaging: A new approach for breast cancer detection", *IJARECE, Special issue IJAECs 2014 ICMSET Proceedings*, ISSN NO: 2277-7318, 2014.
- [4] Arena F, Barone C, DiCicco T, "Use of digital infrared imaging in enhanced breast cancer detection and monitoring of the clinical response to treatment", *Proceedings of the 25th Annual International Conference of the IEEE 2003*; New York, NY, USA, Vol:2, pp.1129-1132, 2003.
- [5] M.Gautherie, "Atlas of breast thermography with specific guidelines for examination and interpretation", *PAPUSA, Milan*, 1989.

## International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirset.com](http://www.ijirset.com)

Vol. 6, Issue 4, April 2017

- [6] Ghayoumi Zadeh H, Haddadnia J, Hashemian M, Hassanpour K, “ Diagnosis of breast cancer using a combination of genetic algorithm and artificial neural network in medical infrared thermal imaging”,*Iran J Med Phys*. 9(4):265-74, 2012.
- [7] Zhang H, Li KY, Sun SR, Wan YW, Yao XL, Zhang XL, “The Value-exploration of the Clinical Breast Diagnosis by Using Thermal Tomography”, Fourth International Conference on Natural Computation; 2008 Oct; Jinan, China: IEEE, pp.138-142, 2008.
- [8] J. H. University, “Overview of Histologic Grade,” 2012. [Online]. Available: <http://pathology.jhu.edu/breast/grade.php>.
- [9] Angshuman Paul, Dipti Prasad Mukherjee, “Mitosis Detection for Invasive Breast Cancer Grading in Histopathological Images”, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 24, NO. 11, NOVEMBER 2015.
- [10] Rejani YIA, Selvi ST, “Early detection of breast cancer using SVM classifier technique”, IJCSE , Vol: 1, Issue 3, pp:127-130, 2009.
- [11] Martins LDO, Junior GB, Silva AC, Paiva ACD, Gattass M , “Detection of Masses in Digital Mammograms using K-means and Support Vector Machine”, Electronic Letters on Computer Vision and Image Analysis , Vol:8, issue 2, pp:39-50, 2009.
- [12] Marina Milosevic, Dragan Jankovic, Aleksandar Peulic, “Thermography based breast cancer detection using texture features and minimum variance quantization”, EXCLI Journal, 13:1204-1215 – ISSN 1611-2156, 2014.
- [13] Satish Saini, Ritu Vijay, “Mammogram Analysis Using Feed-Forward Back Propagation and Cascade-Forward Back propagation Artificial Neural Network” 2015 Fifth International Conference on Communication Systems and Network Technologies, IEEE, pp 1177-1180, 2015.
- [14] R. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classification,” IEEE Transactions On Systems Man And Cybernetics, vol. 3, no. 6, pp. 610–621, 1973.
- [15] Kaiwei C, Xiaoqing L, Jianguo S, Xiao W , “Blind Image Tampering Identification Based on Histogram Features”. 2011 Third International Conference on Multimedia Information Networking and Security IEEE, p 300-303, 2011.