DOI: 10.21917/ijivp.2019.0279

# DETECTION AND CLASSIFICATION OF BREAST CANCER USING SUPPORT VECTOR MACHINE AND ARTIFICIAL NEURAL NETWORK USING CONTOURLET TRANSFORM

**Soumya Hundekar and Saritha Chakrasali**
*Department of Information Science and Engineering, BNM Institute of Technology, India*

*Abstract*

***The technique of image processing is applied to diagnose breast cancer from digital mammogram image. The proposed work uses Contourlet transform to decompose the given gray-scale image. The spatial (textual and statistical) features are been extracted along with frequency domain coefficients. GLCM is the method used for extracting the feature values. Classification of image using support vector machine or artificial neural network classifiers is performed.***

***Keywords:***

***Mammographic Images, Support Vector Machine (SVM), Feature Extraction, Contourlet Transform (CT), Gray Level Co-occurrence Matrix (GLCM), Artificial Neural Network (ANN)***

## 1. INTRODUCTION

Breast cancer is a common disease and a most significant cause of death of women generally caused due to congenital disease. Earlier MRI technique was used by radiologists to determine breast cancer. MRI uses powerful magnetic fields and radio waves to produce comprehensive images of body. By using MRI technique it was inconvenient in imaging thick breast tissues and was injurious.

The present work uses mammography approach which uses low power x-rays to scrutinize human breast [1]. The two methods used in mammography are screening the mammogram and diagnostic mammogram. Some patients with breast cancer have a lump or other indication of illness, which can be detected using diagnostic mammogram method [4]. The situation where the patient has developed no sign of illness, is detected by screening mammogram method.

In the present work, digitalized mammographic ROI is passed as an input to histogram equalization process resulting in an equalized image [25]. On these equalized image some inbuilt functions are carried out as a processing steps. The Contourlet transformation (CT) technique used in this work, is a two dimensional transformation technique which contain properties like multi-resolution and directionality. Both spatial and frequency domain features are extracted as feature values using GLCM method [11] [13].

The classification methods used here are SVM and ANN used to distinguish the digital mammogram image as benign also considered as normal or malignant [2]. The SVM is a discriminative classifier that divides one single class into two different groups and is used for labelling purpose. The ANN is another machine learning technique that is basically used for recognizing patterns, regression and classification processes [5]. The presented work can be utilized by doctors for detection of breast cancer at an initial stage and save lives of women [12].

## 2. RELATED WORK

Hundekar et al. [27] presented a technique for classification of mammogram images using ANN as Benign or malignant by employing GLCM feature extraction technique in an image processing domain. Soumya et al. [26] stated the quality of medical images are generally low so mammogram technique is used. The pre-processing method is used to enhance the raw images. This work uses discrete wavelet transform (DWT) to decompose the given gray-scale image. GLCM is feature extraction technique used. Further, using ANN or SVM classifier the images are classified as benign or malignant. The dataset used here is acquired from MIAS. Chakrasali et al. explained CT-based watermarking algorithm to measure the inherent of CT for watermarking on medical images. The CT derived watermarking technique is also compared with watermarking technique in wavelet area with regard to robustness and imperceptibility [9]. Borde et al. discussed retrieval of images using Contourlet Transform [CT], as it provides multiple scaling and multiple directional decomposition details of an image. The CT decompose the medical image into various directions at different scales [6]. Girish et al. [8] presented a work on the process of feature extraction using digital mammogram images, the feature vectors are extracted from pixel value of mammogram image using texture feature extraction method. Arafi et al. [10] presented a classification process for digitalized mammogram images using ANN classifier. At first, the characteristic features of digital mammograms images are extracted and then classification is made using ANN classification technique. Varsha et al. [15] stated a method for identification of breast cancer using SVM classifier method. In SVM classification process the valid divisions are created using hyper plane based on largest distance to a close training set as the higher margins creates lesser generalized error. Chakrasali et al. [17] proposed the use of Contourlet transform that smoothly capture contours and provides directionality details that are predominantly available in medical images. Kumari et al. [18] used Multi-Layer Perceptron (MLP) and SVM for detection and classification of breast cancer disease. The CAD system is used to classify the mammogram images as normal or abnormal. GLCM is the method used for feature extraction. Kaur et al. [19] used mammography technique to detect breast cancer and described various abnormalities types: breast masses. The database used is mini MIAS or DDSM. The segmentation and noise removal are two methods to extract ROI. The feature extraction method used is GLCM, GLDM and geometrical features. Sankar et al. [14] stated a non-sampled Contourlet transform (NSCT) method for enhancing mammogram images and a comparison is made between 2D Haar discrete wavelet transform and Contourlet Transform.

## 3. CONTOURLET TRANSFORM

Transforms are basically used to get the information available in an image as they belong to the frequency domain where all the hidden data can be retrieved. Contourlet transforms contains some unique properties such as anisotropy and directionality. The CT [16] uses Directional Filter Bank (DFB) and Laplacian Pyramid (LP). LP decompose the image into multiple scales and DFB reveals directional details at each scale level. The low and band pass are two outcomes produced by LP. DFB uses band pass output as an input resulting in directional decomposition. The Fig.1 shows the CT block diagram. The Fig.2 and Fig.3 shows LP and DFB decomposition along with frequency divisions [23].

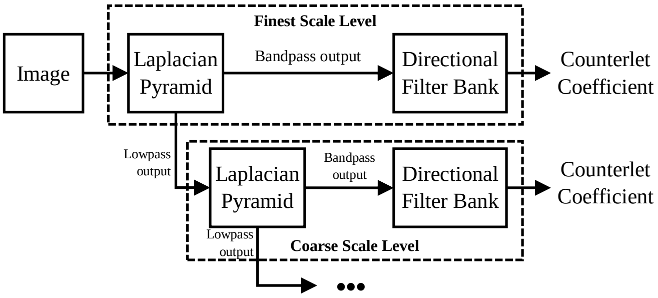


Fig.1. CT block diagram

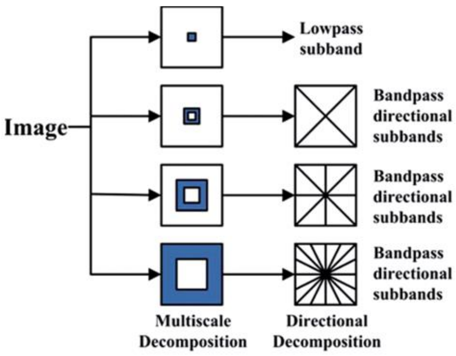


Fig.2. LP and DFB decomposition

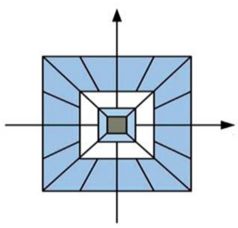

Fig.3. Frequency partitioning

## 4. SUPPORT VECTOR MACHINE

SVM is basically a discrimination classifier that separates one single class into two different groups and is also useful for naming purpose [3]. The hyper-plane is used to bifurcate the two group of classes. For SVM, It is generally not possible to place all the features of an input image in a graph therefore, it stores the feature values of each and every input image by forming vectors for all the values of every image separately. These feature values are further useful to measure accuracy rate and also help conduct numerous iterations. The Fig.4 shows SVM graph.

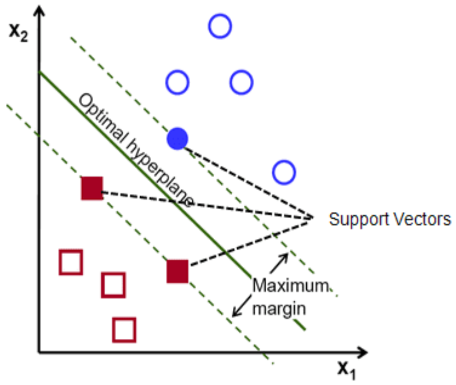


Fig.4. SVM graph for classification

In the Fig.4, $x_1$ and $x_2$ are two features used for plotting. The red squares indicate benign and blue circles indicate malignant values. The optimal hyper-plane is used to differentiate the two classes. The lines above and below the margin are called the boundary region, the pixel value lying on boundary region are called support vectors. If the value of an image falls below the boundary line it's benign and above the boundary line value are malignant image.

## 5. ARTIFICIAL NEURAL NETWORK

ANN is a machine learning method basically utilized for recognizing patterns, classification and regression processes [20].

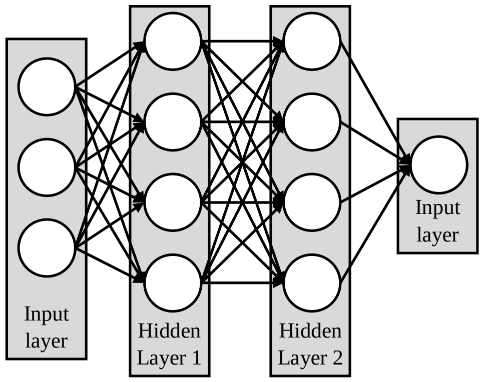


Fig.5. ANN architecture

Artificial neural networks are major parts of machine learning algorithms. Basically, a neural network is constructed based on some processing elements, namely neurons, which are connected together by synapses. Each neuron calculates the sum of weighted input signals and then an activation function is applied to limit the output of neurons to a pre-specified interval. In order to map input vectors to output vectors, the weights of the neural network should be tuned. This process is known as training or learning [24]. Multi-layer neural network (MLP) is composed of one or several hidden layers. MLP is trained using a back propagation (BP) algorithm. In this algorithm, the aim is minimizing the error E between. By applying a 2-layer neural network, an appropriate threshold is obtained for each image. The input and target matrices for training ANN consist of the extracted values related to the intensity. The histogram features from ROI and the obtained threshold of each image for region growing. After training, the obtained neural network is capable of generating a relevant threshold for segmentation. As such, each image is more accurately segmented applying its own generated threshold. The Fig.5 shows the architecture of ANN.

## 6. IMPLEMENTATION

A digital mammogram image is provided as an input to histogram equalization method. This equalization method is used to enhance the intensity values of an image as a result, histogram graph is obtained. The flowchart of the proposed system is shown below in Fig.6 below.

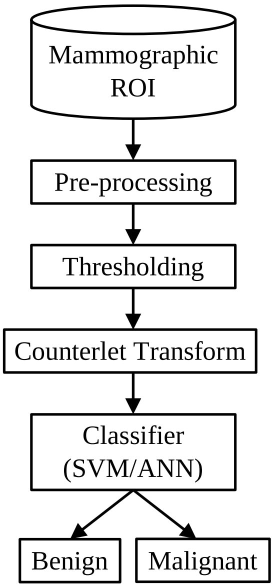


Fig.6. Flow chart of the proposed system

The processing of a digital mammogram image is carried out using inbuilt methods such as thresholding, cross entropy filtration technique and morphological close operation. CT accepts a processed gray-scale image as an input and results in a decomposed image. These transformed images are provided as an input for feature extraction and feature vectors are acquired. The feature values contain both spatial (textual and statistical) and frequency domain feature values. The GLCM is a technique utilized for analysing texture values [21] [22]. Further, two classification techniques SVM and ANN are used to classify the mammogram image as benign or malignant. The results of both SVM and ANN are compared with CT [7].

### 6.1 DATABASE

The mini dataset contains 187 digital mammogram images from which 113 mammogram images are benign and 74 images are malignant. For training purpose, 156 mammogram images are utilised out of total images, here, 56 mammogram images are malignant and 100 are benign images. Similarly, an individual dataset is created for testing purpose contains 31 mammogram images of both benign and malignant, from this set 13 are malignant and 18 are benign images.

## 7. EXPERIMENTAL RESULTS

The original mammogram image has got low contrast and all pixel intensity values are fit to a narrow range, in order to obtain better contrast equalized images are used. In histogram equalization, lighter intensity pixel values of an image are fit to lighter region and darker values are fit to darker area. In Fig.7, the histogram graph are shown where x-axis represent range of pixel values, y-axis indicates intervals and different heights of bar shows frequency of occurrences of data. The graphs are shown for both benign and malignant mammographic pictures.

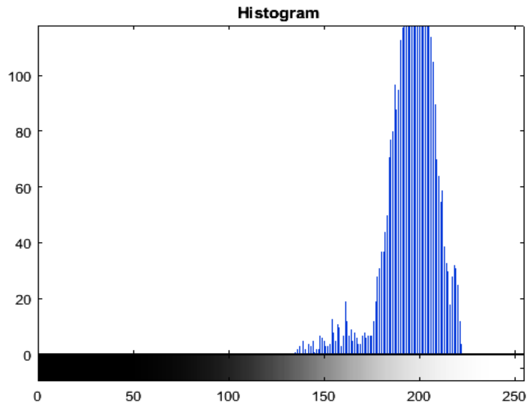


Fig.7. Histogram of benign mammogram image

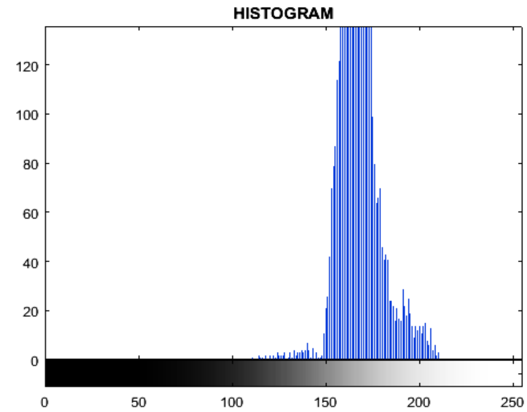


Fig.8. Histogram of malignant mammogram image

The maximum accuracy achieved for SVM at Contourlet level 1 as well as level 2 are tabulated in Table.1. On conducting 35 iterations in total, maximum accuracy achieved at level 1 is 79.64% and at level 2 is 80.54%.

Table.1. SVM accuracy at Contourlet level 1 and level 2

| Iterations | Accuracy at level 1 (%) | Accuracy at level 2 (%) |
|---|---|---|
| 5 | 77.85 | 70.71 |
| 10 | 74.28 | 74.28 |
| 15 | 74.28 | 74.28 |
| 20 | 79.64 | 74.28 |
| 25 | 79.64 | 76.07 |
| 30 | 74.28 | 77.85 |
| 35 | 79.64 | 80.54 |

The maximum accuracy achieved for ANN at Contourlet level 1 and level 2 are tabulated in Table.2 and Table.3 on setting hidden layer value 13 at level 1 and 0 at level 2. At Contourlet level 1, maximum accuracy achieved is 78.5% and at level 2, maximum accuracy obtained is 76.5%.

Table.2. ANN accuracy at Contourlet level 1

| Pattern net value level 1 (%) | Efficiency at level 1 (%) |
|---|---|
| 11 | 59.1 |
| 12 | 66.7 |
| 13 | 78.5 |

Table.3. ANN accuracy at Contourlet level 2

| Pattern net value level 2 (%) | Efficiency at level 2 (%) |
|---|---|
| 10 | 59.7 |
| 20 | 64.7 |
| 30 | 76.5 |

The feature values for Contourlet at level 1 and level 2 are tabulated in Table.4 and Table.5 contain values for both benign and malignant images. From, two tables it is found that feature values less than 0 are malignant and greater than zero are benign.

Table.4. Feature values at Contourlet level 1

| Features | Benign value | Malignant values |
|---|---|---|
| Energy diagonal | 1.4626 | 0.3280 |
| Homogeneity horizontal | 1.1098 | 0.1058 |
| Y horizontal | 1.3892 | 0.2442 |
| Kurtosis horizontal | 1.5142 | 0.0418 |
| Homogeneity vertical | 1.6169 | 0.1394 |
| Kurtosis diagonal | 1.6865 | 0.0713 |
| Energy vertical | 1.6902 | 0.3577 |

Table.5. Feature values at Contourlet level 2

| Features | Benign value | Malignant values |
|---|---|---|
| Energy diagonal | 1.6257 | 0.3229 |
| Homogeneity horizontal | 1.4225 | 0.2004 |
| Y horizontal | 1.8118 | 0.8552 |
| Kurtosis horizontal | 2.0183 | 0.3665 |
| Homogeneity vertical | 2.2394 | 0.5546 |
| Kurtosis diagonal | 2.4414 | 0.1320 |
| Energy vertical | 2.4831 | 0.3120 |

The results of SVM at contour level 1 and 2 achieved by testing after removing one feature at a time are tabulated in Table.6 and Table.7. From Table.6, after removing Y horizontal feature from both benign and malignant images, maximum accuracy obtained for benign is 71.42% and for malignant images, the maximum accuracy achieved is 74.43%. From Table.7, after removing Y horizontal feature from both benign and malignant images maximum accuracy obtained is 74.62% for benign and 71.42% for malignant images.

Table.6. Testing on removing one feature values at SVM Contourlet 1

| Features | Benign value (%) | Malignant values (%) |
|---|---|---|
| Energy diagonal | 57.85 | 59.07 |
| Homogeneity horizontal | 65.22 | 70.10 |
| Y horizontal | 71.42 | 74.43 |
| Kurtosis horizontal | 64.85 | 60.71 |
| Homogeneity vertical | 62.28 | 68.28 |
| Kurtosis diagonal | 70.72 | 73.91 |
| Energy vertical | 66.85 | 62.53 |

Table.7. Testing on removing one feature value at SVM Contourlet 2

| Features | Benign value (%) | Malignant values (%) |
|---|---|---|
| Energy diagonal | 65.86 | 61.17 |
| Homogeneity horizontal | 67.32 | 70.30 |
| Y horizontal | 74.62 | 71.42 |
| Kurtosis horizontal | 63.88 | 67.71 |
| Homogeneity vertical | 64.18 | 69.28 |
| Kurtosis diagonal | 61.71 | 58.72 |
| Energy vertical | 70.85 | 67.58 |

The results for ANN at Contourlet level 1 and level 2 are obtained by testing on removing one feature at a time are tabulated in Table.8 and Table.9. From Table.8, after removing homogeneity vertical from benign the maximum accuracy obtained is 72.7% and on removing Y horizontal feature from malignant values the maximum accuracy achieved is 66.5%. From Table.9, after removing homogeneity vertical from benign values maximum accuracy achieved is 69.6% and for malignant values maximum accuracy obtained is 61.7% on removing Y-horizontal feature.

Table.8. Testing on removing one feature values at ANN Contourlet 1

| Features | Benign value (%) | Malignant values (%) |
|---|---|---|
| Energy diagonal | 71.6 | 49.2 |
| Homogeneity horizontal | 49.9 | 52.7 |
| Y horizontal | 37.5 | 66.5 |
| Kurtosis horizontal | 48.1 | 32.5 |
| Homogeneity vertical | 72.7 | 55.9 |
| Kurtosis diagonal | 39.3 | 60.8 |
| Energy vertical | 44.2 | 53.9 |

Table.9. Testing on removing one feature values at ANN Contourlet 2

| Features | Benign value (%) | Malignant values (%) |
|---|---|---|
| Energy diagonal | 60.6 | 37.6 |
| Homogeneity horizontal | 42.8 | 54.9 |
| Y horizontal | 32.1 | 61.7 |
| Kurtosis horizontal | 40.1 | 32.4 |
| Homogeneity vertical | 69.6 | 54.1 |
| Kurtosis diagonal | 41.2 | 56.6 |
| Energy vertical | 47.2 | 59.7 |

## 8. CONCLUSIONS

The proposed work is mainly used by doctors for early identification of breast cancer which help in treatment at an initial stage and reduce death rates. By considering the above results, it is clear that for SVM, Contourlet level 2 performance is better than level 1 and for ANN, Contourlet level 1 performance is better than level 2. After removing single feature at a time and testing for the results following values obtained. For, SVM, at Contourlet level 2, the maximum accuracy obtained after eliminating Y horizontal feature for both benign and malignant value (benign value accuracy is 74.62% and for malignant value accuracy is 71.42%). For, ANN at Contourlet level 1, after eliminating homogeneity vertical the maximum accuracy obtained for benign value is 72.7% and for malignant value on eliminating Y horizontal feature, the maximum accuracy obtained is 66.5%.

## REFERENCES

[1] Nivedita V. Candade, “Application of Support Vector Machines and Neural Networks in Digital Mammography- A Comparative Study”, PhD Dissertation, Department of Biomedical Engineering, University of South California, 2004.

[2] A. Papadopolulous, D.I. Fotiadis and A. Likas, “Categorization of Clustered Microcalcification in Digitized Mammograms using Neural Network and Support Vector Machine”, *Artificial Intelligence in Medicine*, Vol. 34, No. 2, pp. 141-150, 2005.

[3] Defeng Wang, Lin Shi and Pheng Ann Heng “Automatic Detection of Breast Cancer in Mammograms using Structured Support Vector Machines”, *Neurocomputing*, Vol. 72, No. 13-15, pp. 3296-3302, 2009.

[4] Jose Mejia, O. Humberto, V. Osslan and G. Vianey, “The Non-Subsampled Contourlet Transform for Enhancement of Micro-Calcification in Digital Mammograms”, *Proceedings of Mexican International Conference on Artificial Intelligence*, pp. 292-302, 2009.

[5] Mohammed J. Islam, A. Majid and A. Maher, “An Efficient Automatic Mass Classification Method in Digitalized Mammograms Using Artificial Neural Network”, *International Journal of Artificial Intelligence and Application*, Vol. 1, No. 3, pp. 1-13, 2010.

[6] B. Swapna and Udhav Bhosle, “Image Retrieval using Contourlet Transform”, *International Journal of Computer Applications*, Vol. 34, No. 5, pp. 37-43, 2011.

[7] Jin Chang Ren, “ANN v/s SVM which One Performs Better in Classification of MCCS in Mammogram Imaging”, *Journal Knowledge-Based Systems*, Vol. 26, No. 3, pp. 144-153, 2012.

[8] H. Girish, N. Pradeep, B. Sreepathi and K. Karibasappa, “Feature Extraction of Mammograms”, *International Journal of Bio Information Research*, Vol. 4, No. 1, pp. 241-244, 2012.

[9] Saritha Chakrasali, Ramachandra Murugesan, “A Contourlet Transform Based Versatile Watermarking Algorithm for Medical Images”, *Proceedings of International Conference on VLSI, Communication, Advanced Devices, Signals and Systems and Networking*, pp. 263-271, 2013.

[10] Ayoub Arafi, Youssef Safi and Rkia Fajr, “Classification of Mammographic Images using Artificial Neural Networks”, *Applied Mathematics Sciences*, Vol. 7, No. 89, pp. 4415-4423, 2013.

[11] S. Anand and R. Aynesh Vijaya Rathna, “Architectural Distortion Detection in Mammogram using Contourlet Transform and Texture Features”, *International Journal of Computer Application*, Vol. 74, No. 5, pp. 331-344, 2013.

[12] C. Komal and S. Priti, “Review of Classification of Micro-Calcifications in Digital Mammogram”, *International Journal of Computer Science and Information Technologies*, Vol. 5, No. 2, pp. 1816-1820, 2014.

[13] B. Balasuganya, “Contourlet Based Feature Extraction Method for Classification of Breast Cancer using Thermogram Images”, *International Journal of Scientific and Engineering Research*, Vol. 5, No. 4, pp. 1-7, 2014.

[14] K. Sankar and K. Nirmala, “An Enhanced Mammogram Diagnosis using Shift-Invariant Transform”, *ICTACT Journal on Image and Video Processing*, Vol. 5, No. 2, pp. 920-925, 2014.

[15] Varsha J. Gaikwad, “Detection of Breast Cancer in Mammogram using Support Vector Machine”, *International Journal of Scientific Engineering and Research*, Vol. 3, No. 2, pp. 8-16, 2016.

[16] P. Fatemeh and R.K. Hamidreza, “Improvement of Breast Cancer detection using Non Subsampled Contourlet Transform and Super Resolution Technique in Mammographic Images”, *Iranian Journal of Medical Physics*, Vol. 12, No. 1, pp. 22-35, 2015.

[17] S. Chakrasali, M. Akshata, B.V. Aparna, S. Donthi and N. Jain, “'A Comparative Study between Contourlet and Wavelet Transform for Medical Image Registration and Fusion”, *International Journal of Computer Science and Network Security*, Vol. 6, No. 6, pp. 891-898, 2015.

[18] K. Vani and K. Usha Rani, “Mammogram Classification using Multilayer Perceptron and Support Vector Machine”, *International of Computational Sciences and Engineering*, Vol. 8, No. 1, pp. 46-51, 2016.

[19] Kamaldeep Kaur and E. Pooja, “Classification Through Artificial Neural network and SV of Breast Masses Mammograms”, *International Journal of Advance Research, Ideas and Innovations in Technology*, Vol. 2, No. 4, pp. 1-8, 2016.

[20] Sonal Naranje, "Early Detection of Breast Cancer using ANN", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 4, No. 7, pp. 32-42, 2016.

[21] Ranjit Biswas, Abhijith Nath and Sudiptha Roy "Mammogram Classification Using Gray-Level Co-occurrence Matrix for Diagnosis of Breast Cancer", *Proceedings of International Conference on Micro-Electronics and Telecommunication Engineering*, pp. 1-6, 2016.

[22] P. Samyuktha and D. Sriharsha "Classification of Mammograms using Gray-Level Co-Occurrence Matrix and Support Vector Machine Classifier", *International Journal of Engineering Trends and Technology*, Vol. 6, No. 4, pp. 22-28, 2017.

[23] M .Arfan Jaffar, "Contourlet Transform Domain Based Intelligent Bio-Metric Watermarking System", *International Journal of Computer Science and Network Security*, Vol. 17, No. 5, pp. 65-70, 2017.

[24] M.M. Mehdy, P.Y. Ng, E.F. Shair, N.I. Md Saleh and C. Gomes, "Artificial Neural Networks in Image Processing for Early Detection of Breast Cancer", *Computational and Mathematical Methods in Medicine*, Vol. 2017, pp. 1-15, 2017.

[25] Shankar Thawkar and Ranjana Ingolikar, "Classification of Masses in Digital Mammograms Using Biogeography-based Optimization Technique", *Journal of King Saud University-Computer and Information Sciences*, pp. 1-9, 2018.

[26] Soumya Hundekar and Saritha Chakrasali, "An Observation and Categorization of Breast Cancer utilizing Support Vector and Artificial Neural Networks using Discrete Wavelet Transform", *ICTACT Journal on Soft Computing*, Vol. 9, No. 2, pp. 1851-1855, 2019.

[27] Soumya Hundekar and Saritha Chakrasali, "An Identification of Breast Cancer Disease by using ANN using Contourlet Transform", *International Journal of Scientific Research and Development*, Vol. 6, No. 11, pp. 1-8, 2019.