AN OBSERVATION AND CATEGORIZATION OF BREAST CANCER UTILIZING SUPPORT VECTOR AND ARTIFICIAL NEURAL NETWORK USING DISCRETE WAVELET TRANSFORM

Soumya Hundekar[1] and Saritha Chakrasali[2]
Department of Information Science and Engineering, BNM Institute of Technology, India

Abstract
Digital mammogram images are generally used in medical field as a standard tool for enhancing, transmission and restoring of data. The procedure of image processing is applied to diagnose breast cancer from mammographic ROI image. The quality of mammogram pictures are very low and are sometimes influenced by X-Ray absorption properties of an anatomic parts, size as well as shape. The method of pre-processing help to enhance the raw mammogram image obtained from sensors to aid in an identification of tumors. The proposed work uses Discrete Wavelet Transform (DWT) to decompose the given gray-scale image. The textual and statistical features are being extracted from spatial domain coefficients along with frequency domain coefficients. The feature extraction method used in this work is Gray-Level Co-occurrence Matrix (GLCM). Classification of image is performed using support vector and artificial neural network as benign or malignant. The proposed method is applied on Mammographic Image Analysis Society (MIAS) database. The images of the database have to undergo training, testing and validation stages.

Keywords:
Mammography, Feature Extraction, Support Vector Machine, Artificial Neural Network, Gray-Level Co-occurrence Matrix, Mammographic Image Analysis Society, Discrete Wavelet Transform

1. INTRODUCTION

Breast cancer is a most common disease and a leading cause of death of women generally caused due to inherited abnormalities. Earlier Magnetic Resonance Imaging (MRI) technique was used by radiologists which uses radio waves and strong magnetic fields to produce images of body. By using MRI technique it was difficult in imaging heavy breast tissues and was harmful. The proposed work uses mammography technique which uses small power x-rays to examine human breast. The two techniques used in mammography are screening mammogram and diagnostic mammogram. Some victims with breast cancer do have a lump or other symptoms of illness, which is identified by diagnostic mammogram technique. The other situation is screening mammogram technique in which victim develops no symptoms of illness. The existing work states the performance of SVM with DWT.

The major drawback of existing system is false-negative images sometimes appear normal even though cancer is present. This motivated for an implementation of the proposed approach wherein an experiment is performed on mammogram images using DWT on both SVM and ANN machine learning techniques. The DWT is a mathematical function that splits the given function into numerous measureable components and exhibits DWT co-efficient.

The feature extraction [7] method uses spatial domain and frequency based features. The technique used for extraction of textual and statistical feature values is Gray level co-occurrence matrix (GLCM) [10]. The classifier used are SVM and ANN to classify the mammogram image as malignant or benign [11] [15]. The submitted work could be used by radiologists for an early identification of breast cancer and reduce death rates of women.

To effectively detect breast cancer using SVM and neural network machine learning by performing validation on different kernel and hidden network. The machine learning effectiveness can be tested on frequency domain feature called DWT.

The organization of this paper starts with brief introduction, literature survey from which most of the references are made, detailed explanation on wavelet transformation technique and then comes the implementation section consist of both dataset information and experimental details. The final section of this work is conclusion which contain final result as well as future enhancement ideas which is followed by acknowledgement section.

2. LITERATURE SURVEY

Chakrasali et al [1] proposed the use of wavelet transform on medical images to extend support to rural areas by lessening distance barriers.

Faye et al. [2] proposed a technique for feature extraction by using wavelet coefficient values for classification. Wavelet technique shows an efficient representation for mammographic ROI image. The multi-resolution analysis is the method used to improve the effectiveness of system based on discrete wavelet coefficients.

Girish et al. [3] deliberates about a method of feature extraction using mammogram images, the feature values are extracted from the abnormalities of mammograms using texture feature extraction method.

Arafi et al. [4] showed a classification method for mammograms using Artificial Neural Network (ANN) classifier. Initially, the characteristic of mammograms are extracted and classification is carried out using ANN.

Varsha et al. [8] stated a method for breast cancer detection using SVM. In SVM the good partitions are created using hyper plane based on having largest distance to a closest training set as the highest margin creates lesser generalized errors.

Roopashree et al. [14] stated the use of wavelet transform for decomposition of a given retinal image, based on texture and color which are obtained from frequency area coefficients. Classification process is carried out using Euclidean distance along with SVM.

3. PRELIMINARY WORK

Earlier radiologists used ultrasound, Computed Tomography (CT), radiography, Position Emission Tomography (PET) techniques to determine the cancerous tumor present in the breast area that used strong magnetic fields and radio waves and sometimes false-negative images were found to appear normal which usually made radiologists seek second opinion. To overcome this issue the mammography technique is used.

Mammography is a method that uses low-energy X-Rays to examine the human breast. Mammograms can be either benign (normal) or malignant (cancerous). Benign is a condition where there is no tumor or a tumor is found due to the controlled growth of abnormal cells. Malignancy is a condition where the tumor is formed due to uncontrolled growth of abnormal breast cells.

4. PROPOSED METHOD

The architecture of the proposed work demonstrates the process where the raw mammographic ROI is passed as an input for pre-processing as shown below in Fig.1.

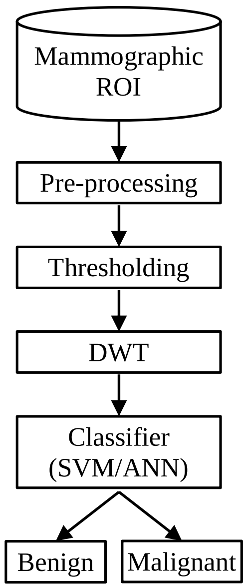

Fig.1. Architecture of the proposed method

During pre-processing, the histogram equalization, thresholding and filtration processes are performed on the mammogram image. The Otsu [5] is the thresholding method that automatically separates background from breast tissues. The filtration process used in this work is cross entropy that filter all the artefact's present in the mammogram ROI image. The morphological close operation is also performed on the image i.e. dilation followed by erosion. The next step is a transformation process, the method used is Discrete Wavelet Transform (DWT) as a result the transformed values are obtained. These are further fed as an input for feature extraction process and feature values are obtained. The method used for extracting the feature is Gray-Level Co-occurrence Matrix (GLCM) which uses both frequency and texture based feature for image retrieval. The normalization process further set the range of pixel intensity values between 0 (benign) to 1 (malignant). Finally, the two classifiers SVM and ANN are utilized to classify the given image as benign or malignant [12] [13]. SVM help to differentiate the given image as benign or malignant and is also used for labelling purpose whereas, ANN is a machine learning process used for regression, classification and pattern recognition.

5. WAVELET TRANSFORM

Transforms are used to know the information present in an image as they refer to the frequency domain where all the hidden information can be retrieved. Wavelet transforms contain a series of both high and low pass filters which extracts all the high and low frequencies separately [2]. The image should be pre-processed before transforming. The wavelet transform uses processed gray-scaled image as input and results in a transformed image. The Fig.2 and Fig.3 shows wavelet decomposition where $x[n]$ is an input image, $g[n]$ and $h[n]$ are low and high pass filters respectively resulting in level 1 wavelet coefficients which will then be down sampled by 2 and once again lead to high and low pass filter banks. During decomposition at level 1 image is divided into four sub-bands that is Approximation details (LL), Horizontal details (HL), Vertical details (LH) and Diagonal details (HH) of an original image in first level decomposition. The LL sub-band will be further decomposed into four more sub-bands for second level decomposition. All the required features will be extracted from LL sub-band of second level decomposition.

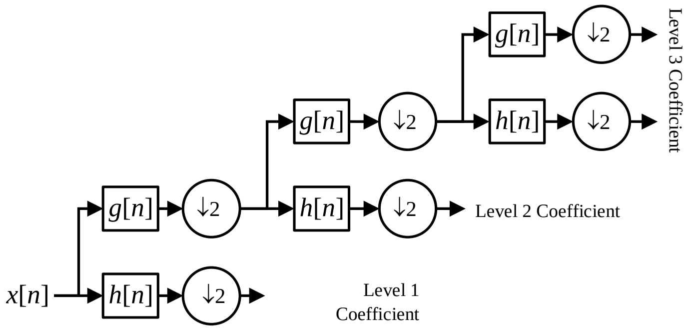

Fig.2. Wavelet Decomposition

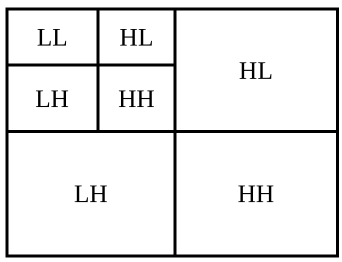

Fig.3. Second level decomposition

5.1 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. 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.

In 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.

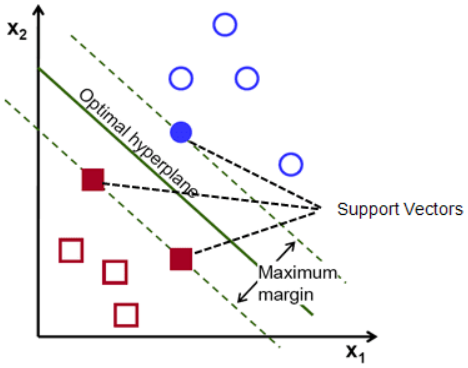

Fig.4. SVM graph for classification

5.2 ARTIFICIAL NEURAL NETWORK

ANN is a machine learning method basically utilized for recognizing patterns, classification and regression processes [9] [16]. The Fig.5 shows the architecture of ANN.

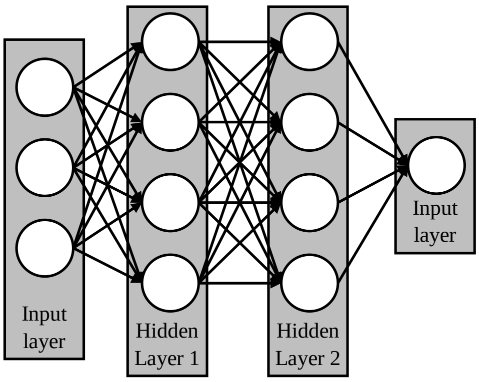

Fig.5. ANN architecture

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. 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.

6. IMPLEMENTATION

A gray-scale mammographic ROI is passed as an input to histogram equalization process. This histogram equalization is used to increase the intensity value of an image resulting in a histogram graph. The pre-processing is carried out using some integral functions such as Otsu thresholding which automatically separates background and breast tissues, cross entropy filtration help remove redundant data from raw mammogram image and morphological close operation. The Discrete Wavelet Transform (DWT) is a mathematical function that splits the given image into various scaling components resulting in a transformed image. These are fed as an input for feature extraction and feature vectors are obtained. The feature values consist of both spatial and frequency domain values. The GLCM is a method used for texture analysis. The normalization function is performed to set pixel values to a certain range. Further, two machine learning classifiers called SVM and ANN are used to classify the image as benign or malignant.

6.1 DATABASE

There are gross 187 mammographic ROI images used, out of which 113 images are for benign and 74 are for malignant. In the first instance, for training 156 images are used out of total images, from this 56 are malignant ROI images and 100 are benign. Similarly, a discrete dataset is created for testing from total set of images containing 31 images of both benign and malignant images from these 13 are malignant and 18 are benign ROI images. The size of each and every mammogram image is 1024*1024 pixels in the dataset at 256 gray levels and are in (.pgm) format. This database holds both left and right breast images of the patients.

6.2 EXPERIMENTAL RESULTS

The maximum accuracy achieved for SVM at DWT level 1 and level 2 are shown in Table.1. On conducting 30 iterations overall, the maximum accuracy achieved at level 1 is 80.35% and at level 2 is 69.64% on 15th iteration.

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

Iterations	Accuracy at level 1 (%)	Accuracy at level 2 (%)
5	73.21	64.2
10	75	67.85
15	80.35	69.64
20	75	62.5
25	76.38	66.7
30	78.57	66.7

The maximum accuracy achieved for ANN [6] at DWT level 1 and level 2 are shown in Table.2 below on performing 30 iterations. At DWT level 1, the maximum accuracy achieved is 64.7% and at level 2, maximum accuracy obtained is 68.7% on pattern net value 20.

Table.2. ANN accuracy at DWT level 1 and level 2

Pattern net value	Efficiency at level 1 (%)	Efficiency at level 2 (%)
10	52.9	58.9
20	64.7	68.7
30	58.9	61.9

The feature values for DWT at level 1 and 2 are shown in Table.3 and Table.4 below which contains value for both benign and malignant images. From, two tables below it is found that kurtosis horizontal and diagonal values for benign and malignant images are higher when compared with all other feature values.

Table.3. Feature values at DWT level 1

Features	Benign value	Malignant values
Energy diagonal	0.1331	0.1764
Homogeneity horizontal	0.1484	0.1715
Y horizontal	0.6284	0.5422
Kurtosis horizontal	2.6113	2.8236
Homogeneity vertical	0.1521	0.1539
Kurtosis diagonal	2.5482	2.5688
Energy vertical	0.1684	0.1817

Table.4. Feature values at DWT level 2

Features	Benign value	Malignant values
Energy diagonal	0.0038	0.0040
Homogeneity horizontal	0.2330	0.2509
Y horizontal	0.2473	0.3674
Kurtosis horizontal	1.9274	2.7136
Homogeneity vertical	0.2301	0.2568
Kurtosis diagonal	2.0934	2.2840
Energy vertical	0.0038	0.0042

The results for SVM at DWT level 1 and 2 obtained by testing on removing one feature at a time are represented in Table.5 and Table.6 below. From Table.5, after removing kurtosis horizontal for benign and homogeneity vertical from malignant values, the maximum accuracy achieved is 78.57%. From Table.6, on removing homogeneity horizontal from both benign and malignant values, the maximum accuracy obtained is 74.21%.

Table.5. Testing removing one feature at a time at DWT level 1

Features	Benign value (%)	Malignant values (%)
Energy diagonal	71.42	71.42
Homogeneity horizontal	74.33	74.33
Y horizontal	75	75
Kurtosis horizontal	78.57	69.64
Homogeneity vertical	73.21	78.57
Kurtosis diagonal	75	73.21
Energy vertical	73.21	67.85

Table.6. Testing removing one feature at a time at DWT level 2

Features	Benign value (%)	Malignant values (%)
Energy diagonal	67.85	66.07
Homogeneity horizontal	74.21	74.21
Y horizontal	71.42	71.42
Kurtosis horizontal	67.85	60.71
Homogeneity vertical	64.28	64.28
Kurtosis diagonal	60.71	58.92
Energy vertical	67.85	62.50

The results for ANN at DWT level 1 and level 2 are obtained by testing on removing one feature at a time as shown in Table.7 and Table.8 below. From Table.7, on removing Y horizontal and homogeneity vertical for benign the maximum accuracy obtained is 58.8% and on removing homogeneity horizontal and kurtosis diagonal from malignant values the maximum accuracy achieved is 52.9%. From Table.8, on removing energy diagonal and homogeneity vertical from benign values maximum accuracy achieved is 70.6% and for malignant values the maximum accuracy obtained is 64.7% on removing Y-horizontal feature.

Table.7. Testing removing one feature at a time at ANN DWT level 1

Features	Benign value (%)	Malignant value (%)
Energy diagonal	35.3	21.8
Homogeneity horizontal	38.8	52.9
Y horizontal	58.8	50.7
Kurtosis horizontal	52.9	29.4
Homogeneity vertical	58.8	47.1
Kurtosis diagonal	41.2	52.9
Energy vertical	52.9	29.4

Table.8. Testing removing one feature at a time at ANN DWT level 2

Features	Benign value (%)	Malignant value (%)
Energy diagonal	70.6	47.1
Homogeneity horizontal	52.9	58.8
Y horizontal	35.3	64.7
Kurtosis horizontal	47.1	29.4
Homogeneity vertical	70.6	52.9
Kurtosis diagonal	35.3	58.8
Energy vertical	41.2	52.9

7. CONCLUSIONS

The proposed system is used by radiologists for early detection of breast cancer which aids in treatment at an initial stage and reduce death rates. By observing the above results it is clear that for accuracy:

- *SVM at DWT level 1 and level 2*: The level 1 performs better than second level i.e. 80.35% on 15th iteration.
- *ANN at DWT level 1 and level 2*: The maximum accuracy obtained is 68.7% at pattern net value 20 at level 2.
- On testing by removing one feature at a time following values obtained:
- *SVM at DWT level 1*: The maximum accuracy gained is 78.57% for both benign and malignant images on removal of kurtosis horizontal feature for benign and homogeneity vertical feature for malignant images.
- *SVM at DWT level 2*: on removal of homogeneity horizontal feature the maximum accuracy gained is 74.21% for both benign and malignant values.
- *ANN at DWT level 1*: The highest value gained for benign images is 58.8% for both homogeneity vertical and Y-horizontal feature and for malignant images the maximum value achieved is 52.9% on removal of kurtosis diagonal and homogeneity horizontal feature.
- *ANN at DWT level 2*: The highest accuracy obtained for benign images is 70.6% for both homogeneity vertical and energy–diagonal feature and for malignant images the maximum accuracy gained is 64.7% on removal of Y-horizontal feature.

Hence, as stated in this work for smaller dataset SVM still performs better than ANN. The future enhancement can be conducted for an identification of various stages of cancer such as early, advanced and metastatic stages.

REFERENCES

[1] Saritha Chakrasali, M Sumathi, R Bhaskaran and R Murugesan, "Secure Dual Watermarking in the Wavelet Domain for Teleradiology", *Proceedings of International Conference on Advances in Information and communication Technologies*, pp. 41-48, 2007.

[2] Ibrahima Faye and Brahim Belhaouari Sami, "Digital Mammogram Classification using Wavelet Based Feature Extraction Method", *Proceedings of International Conference on Computer and Electrical Engineering*, pp. 23-29, 2009.

[3] 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.

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

[5] 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. 1-13, 2015.

[6] C.S. Roopashree and Saritha Chakrasali, "Detection of Diabetic Retinopathy using Wavelet Transform and SVM Classifier", *Indian Journal for Research in Applied Sciences and Engineering Technology*, Vol. 5, No. 5, pp. 32-38, 2017.

[7] S. Satish and R. Vijay, "Design and Analysis of Breast Cancer Detection System Using Mammogram Features Extraction", *International Journal of Electronics and Computer Science Engineering*, Vol. 3, No. 4, pp. 406-408, 2014.

[8] 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. 1-7, 2015.

[9] Sonal Naranje, "Early Detection of Breast Cancer using ANN", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 4, No. 7, pp. 14008-14013, 2016.

[10] 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.

[11] Muhammad Talha, "Classification of Mammograms for Breast Cancer Detection using Fusion of Discrete Cosine Transform and Discrete Wavelet Transform Features", *Biomedical Research*, Vol. 27, No. 2, pp. 322-327, 2016.

[12] 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. 3, No. 1, pp 131-137, 2017.

[13] Deepthi Sehrawat, Abhishek Sherawat, Dhirendra Jaiswal and Anima Sen, "Detection and Classification of Tumor in Mammograms using Discrete Wavelet Transform and Support Vector Machine", *International Research Journal of Engineering and Technology*, Vol. 4, No. 5, pp. 1328-13334, 2017.

[14] C.S. Roopashree and Saritha Chakrasali. "Detection of Diabetic Retinopathy using Wavelet Transform and SVM Classifier", *Indian Journal for Research in Applied Sciences and Engineering Technology*, Vol. 5, No. 5, pp. 121-127, 2017.

[15] M.M. Mehdy et al., "Artificial Neural Networks in Image Processing for Early Detection of Breast Cancer", *Computational and Mathematical Methods in Medicine*, Vol. 2017, pp. 1-15, 2017.

[16] Nadeem Tariq, "Breast Cancer Detection using Artificial Neural Network", *Journal of Molecular Biomarkers and Diagnosis*, Vol. 9, No. 1, pp. 1-6, 2017.