Extreme learning machine for cancer classification in Mammograms based on Fractal and GLCM Features

Saraswathi.D[1], Srinivasan.E[2], Ranjitha.P[3]

[1,3]*Manakula Vinayagar Institute of Technology, Puducherry.*

[2]*Pondicherry Engineering College, Puducherry.*

E-mail: saraswathiece@mvit.edu.in

ranjitha18993@gmail.com

ABSTRACT

Breast cancer is becoming a leading cause of death among women in the whole world; meanwhile, it is confirmed that the early detection and accurate diagnosis of this disease can ensure a long survival of the patients. In this work, classification of Malignant, Benign and normal Images of Mammogram is performed. Extreme Learning Machine is developed for single hidden layer feedforward neural network and connection between the input layer and hidden neurons are randomly assigned and remain unchanged during the learning period. In this work, noises are removed by using Curvelet transform and Region of Interest (ROI) mammogram images using manual segmentation. By using Fractal and GLCM features are extracted from the Region of Interest (ROI) and the classification is performed by using Extreme Learning Machine (ELM) Algorithm. ELM is used to classifies the input images into normal and abnormal. A set of images (150) from Mammographic Image Analysis Society (MIAS) database is used for evaluating the system. The performance is analyzed using classification accuracy rate and reducing the false positive rate. The experimental results are compared with wavelet transform and Support Vector Machine Method. The result shows that Curvelet Transform with Extreme Machine Learning (ELM) has high classification accuracy rates of 98.3% than the existing method for diagnosing the breast cancer. This system will help the radiologist to diagnose the breast cancer in an efficient way.

Keywords: *Mammogram, Region of Interest, Extreme Machine Learning, Mammographic Image Analysis Society, Curvelet Transform.*

1. INTRODUCTION

Mammogram image is considered as the most reliable, low cost, and highly sensitive technique for detecting small lesions. The radiologists are searching for signs of abnormality, but the signs of early disease are often small or subtle. That is the main cause of many missed diagnoses that can be mainly attributed to human factors such as subjective or varying decision criteria, distraction by other image features, or simple oversight. Nevertheless, a false positive detection causes unnecessary biopsy. It has been estimated that only 20–30% of breast biopsy cases are proved to be cancerous. On the other hand, in a false negative detection, an actual tumor remains undetected. Studies have shown that 10– 30% of the visible cancers are undetected. Thus, there is a significant necessity for developing methods for automatic classification of suspicious areas in mammograms for aiding radiologists to improve the efficacy of screening programs and avoiding unnecessary biopsies.

Computer aided detection (CAD) systems use computer technologies to detect abnormalities in mammogram such as micro calcification, mass, architecture distortion and asymmetry, can play a key role in early detection of breast cancer and help to reduce the mortality rate among women with breast cancer [8].

One of the main points that should be taken under serious consideration when implementing a robust classifier for recognizing breast tissue is the selection of the appropriate features that describe and highlight the differences between the abnormal and the normal tissue in an ample way. Feature extraction is an important factor that directly affects the classification result in mammogram classification. Most systems extract features to detect and classify the abnormality as benign or malignant from the textures, statistical properties, spatial domain, fractal domain and wavelet bases.

classification of malignant and benign is still very challenging and a difficult problem for researchers. Researchers spend a lot of time in attempting to find a group of features that will aid them in improving the classification for malignant from benign. There are various feature transforms that serve to condense input data and to reduce redundancies by highlighting important characteristics of the image. Wavelet transform (WT) is the most well-known two dimensions and multi-resolution transform that decompose an image in horizontal, vertical and diagonal directions. Wavelet provides an effective representation for images. In recent years, numerous schemes for mammogram analysis using wavelet were introduced. M.Thirumalesh et al. presented a multilevel wavelet decomposition process by using different types of wavelet (Haar, db8, bior3.7, and sym8). Experimental result showed that Daubechies wavelet with the 4th level decomposition achieved the best detecting result with 96% [6]. Essam A. Rashed et al. provides a multiresolution analysis system for taking digital mammograms is proposed and tested. This system is based on fractional amount of largest wavelet coefficients in multilevel decomposition. The results shown that the system has a high efficient compared by other systems [16]. Following the success of wavelets, a series of multi resolution, multidimensional tools, namely contourlet, curvelet, ridgelet have been developed in the past few years. Curvelet was developed by Candes and Donoho for providing efficient representation of smooth objects with discontinuities along curves. Curvelet efficiently represents discontinuities along curves or edges in images or objects. Some studies using curvelet transform in image processing have been proved out. Mohamed Meselhy Eltoukhy et al. presented a feature extraction method by using multiresolution representation. Experimental results shown that curvelet is better than wavelet and produces 91% accuracy [4]. Mohamed Meselhy Eltoukhy et al. presents a comparison between wavelet and curvelet for breast cancer diagnosis. The results shown than Curvelet transform outperforms wavelet transform [14]. Researchers attempt to find new two dimensions and multi-resolution transforms as the traditional WT with more directionality in contrast with WT. Due to, Contourlet transform (CT) [9][12]. Although CT was originally introduced in discrete domain, ST similar to continues WT represents an affine system obtained by scaling and translation and in addition unlike WT it has an extra parameter called shear. The original WT, CT, and ST because of using up- and down-sampling are shift variant. That means, the coefficients are changing whenever the original signal is translating.

From the above papers, it has some disadvantages such as 1. The lack of transparency in results. ***2.*** The biggest limitation of the support vector approach lies in choice of the kernel. 3. Limitation in speed and size, both in training and testing. 4. Although SVMs have good generalization performance, they can be abysmally slow in test phase. 5. Choosing appropriately hyper parameters of the SVM. To overcome all these disadvantages, curvelet transform via wrapping with Extreme Learning Machine algorithm is introduced.

The remainder of this paper is organized as follows. Section 2 describes the proposed method. Section 3 contains experimental results and discussion while section 4 concludes the article.

2. Proposed system

This paper deals with the problem of presenting images with a reduced and

iv

discriminative feature vector. The proposed method consists of the following steps: noise reduction, feature extraction and classification (Fig.1)

2.1 Noise Reduction

Curvelet transform was developed by candes and Donoho in 1999. Its development was motivated by the need of image analysis. The transform represents an image at different scales and angles[5].

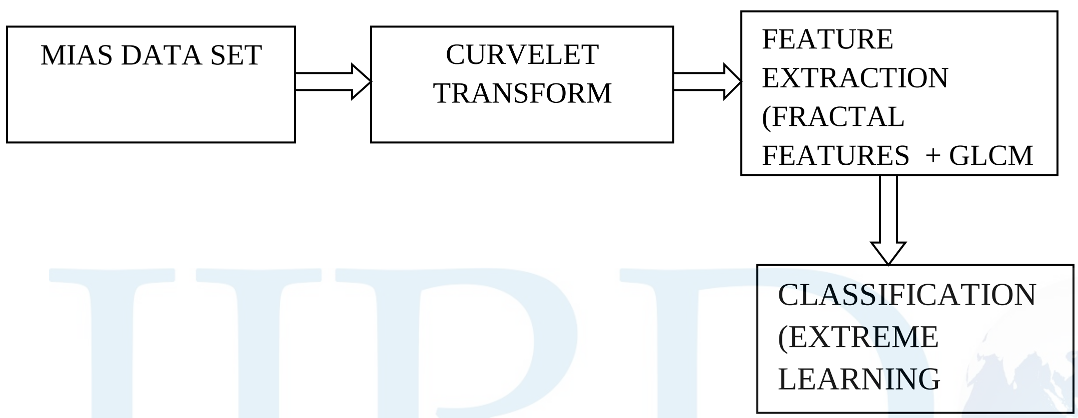

Fig. 1 The basic model of the proposed algorithm

It provides an efficient representation of smooth objects with discontinuities along curves. As an extension of the wavelet concept, the curvelet transform was not only successfully applied in several image processing field but also outperformed waved- based Methods [13]. Curvelet will be superior over wavelet in following cases:

(i) Subband Decomposition,
(ii) Smooth partitioning,
(iii) Renormalization,
(iv) Ridgelet Analysis.

Curvelet transform is multiscale and multidirectional. Curvelets exhibit highly anisotropic shape obeying parabolic-scaling relationship. In order to implement curvelet transform, first a 2D-FFT of the image is taken. Then the 2D fourier frequency plane is divided into parabolic wedges. Finally an inverse FFT of each wedge is taken to find the curvelet coefficients at each scale j and angle l. Figure 2 shows the division of wedges of the fourier frequency plane.

v

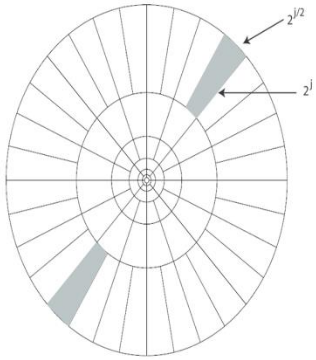

Fig. 2 Curvelets in Fourier frequency

Curvelet transform via wrapping is chosen for this paper as it is the fastest curvelet transform currently available. Steps involved in curvelet denoising method:

(i) Estimate the noise Standard Deviation (σ) in the input image

(ii) Evaluate the curvelet transform of the input image. We get a set of bands W_j, each band W_j contains N_j coefficients and corresponding to a given resolution level.

(iii) Calculate noise SD σ_j for each band j of the curvelet transform

(iv) For each band j: calculate the maximum of the band and multiply each curvelet coefficient.

(v) Reconstruct the image from the modified curvelet coefficient.

2.2 Feature Extraction:

2.2.1 Fractal theory:

Fractal geometry was introduced and supported by the mathematician Benoit Mandelbrot to characterize objects with unusual properties in classical geometry [15,3]. This concept helps to interpret the inexplicable subjects by any law. It is shown to study irregular objects in the plan or space, which is actually better suited to handle the real world. Fractal geometry has experienced significant growth in recent years.

A texture can be seen as a repetition of similar patterns randomly distributed in the image. Fractal approach allows measuring the invariant translation, rotation and even scaling. Many approaches have been developed and implemented to characterize real textures using concepts of this geometry [7]. Fractal dimension is a fractional parameter and greater than the topological dimension. The fractal dimension is a helpful of an image by describing the texture irregularity [11] [26].

$$F = \frac{\log N}{\log {}^{1}\!/_{s}}$$

(1)

$N = r^d$ is the number of self similar parts in the image and ${}^{1}\!/_{s}$ is the scaling factor in equation

(1) The best methods have been introduced to reduce quantization levels and it was demonstrated the utility of the fractal parameter and its variants in the transformed texture characterization images [7]. However, given that this parameter alone is not sufficient to discriminate visually different surfaces; the concepts of homogeneity of fractal dimension were examined. The Hausdorff- imension is the most studied mathematically. We are used in other work the Hausdorff -multifractal spectrum to describe the different fractalities recovering the medical image. It is generally the best known, but individually, it is probably the least calculated.

$$Lac = \frac{\frac{1}{KL}\sum_{K=0}^{K-1}\sum_{l=0}^{l-1} I_f(k,l)^2}{(\frac{1}{KL}\sum_{K=0}^{K-1}\sum_{l=0}^{l-1} I_f(k,l))^2} - 1$$

(2)

F measure the complexity of texture, but if there are two textures with the same F then by lacunarity, we are able to differentiate between two texture with the same F values and it also use as estimator in classification the texture of image, it is calculated by equation (2) and if it's value near to zero ,this mean that there are homogeneity (i.e. there is no tumor) and vice versa, that if it is other than we deduce that there is a tumor but, after selection step that is applied manually.

2.2.2 GLCM Features

Grey-Level Co-occurrence Matrix (GLCM) texture measurements have been the workhorse of image texture since they were proposed by Haralick [17], and 14 statistical features were introduced. GLCM is a statistical method of examining texture that considers the spatial relationship of pixels. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image. These features are generated by calculating the features for each one of the cooccurrence matrices obtained by using the directions 0°, 45°, 90°, and 135°, then averaging these four values.

The some GLCM Features used in this work are:

1. Energy

   The energy returns the sum of squared elements in the GLCM. Energy is *1* for a constant image.

$$\sum_{i,j=0}^{N-1}\left(P_{ij}\right)^2$$

2. Contrast

   The contrast returns a measure of the intensity contrast between a pixel and its neighbour over the whole image.

$$\sum_{i,j=0}^{N-1} p_{ij}(i-j)^2$$

iv

3. Homogeneity

   The homogeneity returns a value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. Homogeneity is 1 for a diagonal GLCM.

$$\sum_{i,j=0}^{N-1} \frac{p_{ij}}{1+(i-j)^2}$$

4. Correlation

   The correlation returns a measure of how correlated a pixel is to its neighbour over the whole image.

$$\sum_{i,j=0}^{N-1} p_{ij} \frac{(i-\mu)(j-\mu)}{\sigma^2}$$

Pij = Element *i, j* of the normalized symmetrical GLCM.

N = Number of gray levels in the image as specified by number of levels in under quantization on the GLCM.

μ = The GLCM mean, calculated as:

$$\mu = \sum_{i,j=0}^{N-1} iP_{ij}$$

σ^2 = The variance of the intensities of all reference pixel in the relationships that contributed to the GLCM, calculated as:

$$\sigma^2 = \sum_{i,j=0}^{N-1} P_{ij}(i-\mu)^2$$

2.3 Classification

Different pattern classifiers have been applied for object recognition, including After dimensionality

reduction, we used Extreme learning machine (ELM) classifier for classification.

ELM as a new learning algorithm for single layer feedforward neural networks (SLFNs), was first introduced by Huang el al. [1,2] ELM seeks to overcome the challenging issues faced with the traditional SLFNs learning algorithms such as slow learning speed, trivial parameter tuning and poor generalization capability. ELM has demonstrated great potential in handling classification and regression tasks with excellent generalization performance. The learning speed of ELM is much faster than conventional gradient based iterative learning algorithms of SLFNs like back propagation algorithm while obtaining better generalization performance. ELM has several significant features [2] which distinguish itself from the traditional learning algorithms of SLFNs:

– ELM is easily and effectively used by avoiding tedious and time-consuming parameter

tuning.

v

– ELM has extremely fast learning speed.

– ELM has much better generalization performance than the gradient based iterative learning algorithms in most cases.

– ELM is much simpler and without being involved in local minima and over-fitting.

– ELM can be used to train SLFNs with many non-differentiable activation functions.

A standard ELM classifier, whose M hidden nodes use infinitely differentiable activation functions, could approximate arbitrary samples with zero error [1], which means given a training data $\vartheta = \{(xi, ti)\}N\ i=1$, the output function of the single hidden layer feed forward neural network (SLFN) with L hidden neurons can be expressed as:

$$f(x) = \sum_{i=1}^{N} \beta_i G(a_i, b_i, x_i) = \beta \cdot h(x)$$

(3)

Other than updating the network parameters iteratively as done in conventional gradient decedent algorithms, ELM employs random hidden node parameters and the tuning free training strategy for feedforward neural networks. The learning is then transferred to solving a linear system which has been well suited via the minimal norm least square approach [2]. As shown in the universal approximation capability theorems, ELM is flexible with hidden activation functions. Almost any nonlinear piecewise continuous functions and their linear combinations work well in ELM algorithms [1]. Thanks to these advantages, ELM has shown superiority of the fast learning speed and reasonable generalization performance over SVM and its variants [1,2].

3. Simulation Results

3.1. Input Mammogram Image:

Initially, the mammogram input image is given to the image preprocessing is shown in Fig.3.1.

vi

Fig.3.1. Input Mammogram Image

3.2. Cropped Image:

Fig.3.2 shows the cropped image. Here, cropping operation is applied to the image to cutoff the black parts of the image. It also eliminates the label on the image and the black background.

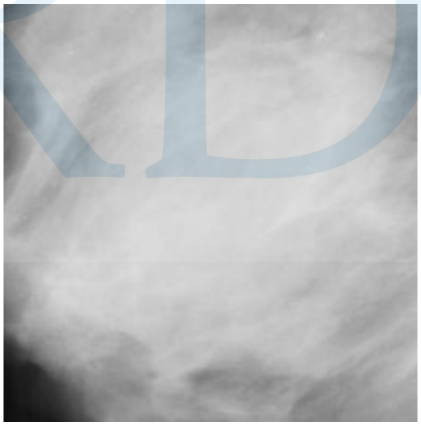

Fig.3.2. Cropped Image

iii

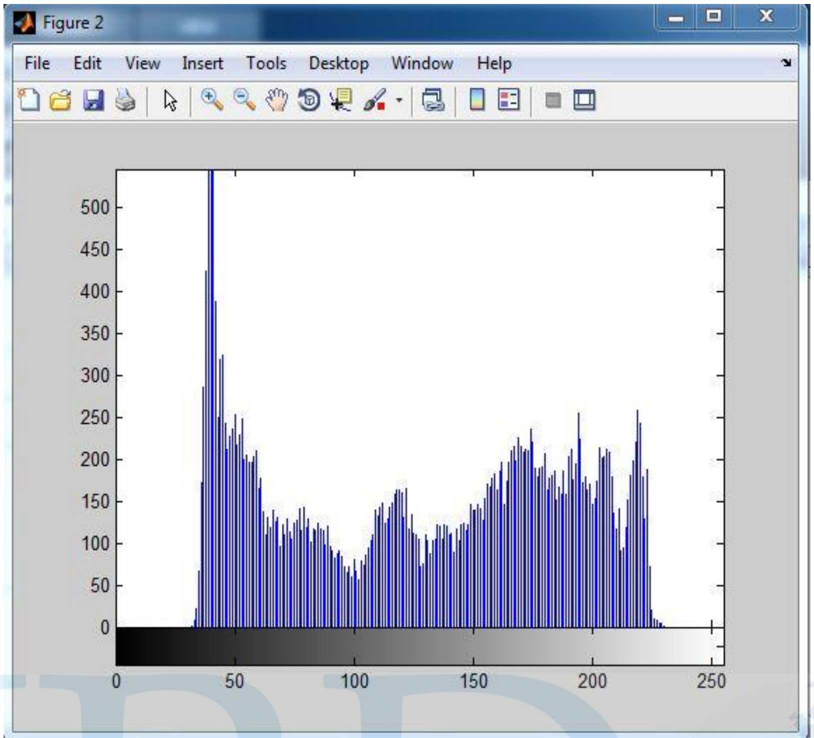

Fig. 3.2.1 Histogram of an input image

3.3. CURVELET TRANSFORM OUTPUT:

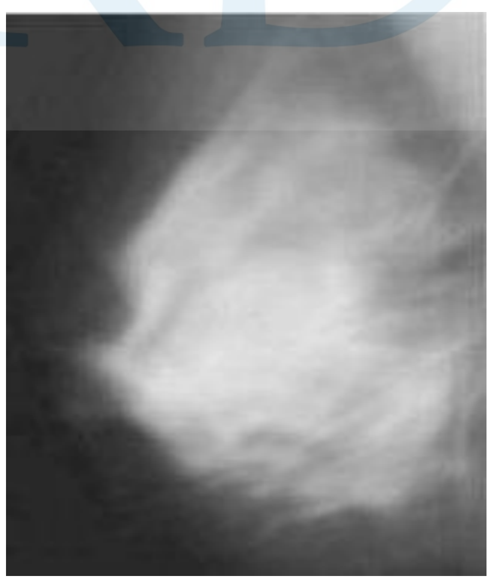

Fig.3.3. Curvet Transform

iv

Fig.3.3 shows the Curvelet Transform based on Wrapping Method. Firstly, it remove the noise present in the Mammogram. Then the features are extracted. After that training process is done for classification. Finally, ELM classifier is applied to classify the mammogram images. If the given ROI image is normal, there is no need for further classification. If it is abnormal, then the process is again classified into benign and malignant for breast cancer detection.

3.4. ROC Curve:

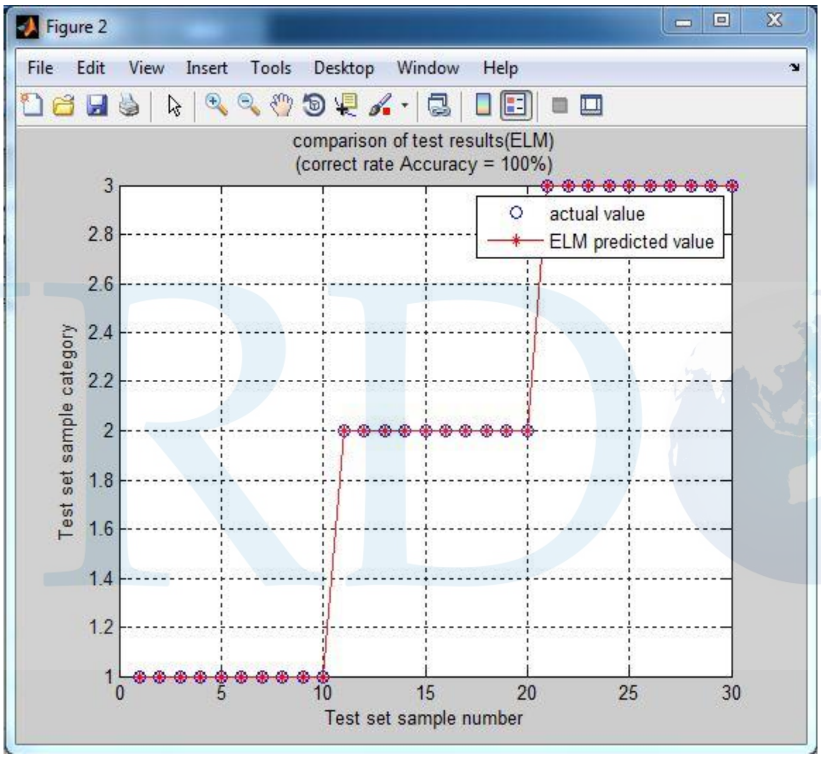

Fig.3.4.1. Accuracy of the proposed Method

Fig.3.4.1 shows the ROC curve of proposed system. From the figure, it is observed that the proposed method has higher classification accuracy, sensitivity and specificity of 100%.

v

3.5 comparison results:

Table 3.5 comparison of extreme learning machine with other transforms

METHODS	ACCURACY (%)	SPECIFICITY (%)	SENSITIVITY (%)
Wavelet + ELM	78%	78%	77.5%
Curvelet+ELM	93.4%	93.2%	93%
Curvelet+ GLCM, Fractal+ ELM	**98.53%**	**98%**	**97%**

4.Conclusion:

In this work, the difference between wavelet and Haar for analysis of digital mammogram is carried out. Firstly, each mammogram image is decomposed using Curvelet Transform (CT) decomposition process. The features are extracted from each transform output using Fractal and GLCM Features. These features are given to the Extreme Learning Machine classifier. The performance of the classifier is evaluated using classification accuracy rate, sensitivity and specificity. The experimental results show that the extracted features based on Curvelet transform with GLCM and Fractal gives a better performance as compared with existing techniques such as wavelet and haar transform.

The results show that Curvelet transform with GLCM and Fractal achieves a higher classification accuracy rates of 98.33% with sensitivity and specificity of 98% and 97.4% than the existing method for diagnosing the breast cancer. In future, ELM has to be combined with various existing feature selection techniques and other transforms for diagnosing the breast cancer.

REFERENCES

[1] Zhou, Yicong, Jiangtao Peng, and CL Philip Chen. "Extreme learning machine with composite kernels for hyperspectral image classification." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8.6 (2015): 2351-2360.

[2]Cao, Jiuwen, and Zhiping Lin. "Extreme learning machines on high dimensional and large data applications: a survey." *Mathematical Problems in Engineering* 2015 (2015).

[3]Korchiyne, Redouan, et al. "A combined method of fractal and GLCM features for MRI and CT scan images classification." *arXiv preprint arXiv:1409.4559* (2014).

vi

[4]Zong, Weiwei, Guang-Bin Huang, and Yiqiang Chen. "Weighted extreme learning machine for imbalance learning." *Neurocomputing* 101 (2013): 229-242.

[5]Gedik, Nebi, and Ayten Atasoy. "A computer-aided diagnosis system for breast cancer detection by using a curvelet transform." *Turkish Journal of Electrical Engineering & Computer Sciences* 21.4 (2013): 1002-1014.

[6]Lu, Chi-Jie, and Yuehjen E. Shao. "Forecasting computer products sales by integrating ensemble empirical mode decomposition and extreme learning machine." *Mathematical Problems in Engineering* 2012 (2012).

[7]Florindo, João Batista, and Odemir Martinez Bruno. "Fractal descriptors based on Fourier spectrum applied to texture analysis." *Physica A: statistical Mechanics and its Applications* 391.20 (2012): 4909-4922.

[8] Huang, Guang-Bin, et al. "Extreme learning machine for regression and multiclass classification." *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 42.2 (2012): 513-529.

[9]Jasmine, JS Leena, S. Baskaran, and A. Govardhan. "An automated mass classification system in digital mammograms using contourlet transform and support vector machine." *International Journal of Computer Applications* 31.9 (2011): 54-60.

[10]Lopes, R., et al. "Prostate cancer characterization on MR images using fractal features." *Medical physics* 38.1 (2011): 83-95.

[11]Miao, Qi-guang, et al. "A novel algorithm of image fusion using shearlets." *Optics Communications* 284.6 (2011): 1540-1547.

[12] Islam, Mohammed J., Majid Ahmadi, and Maher A. Sid-Ahmed. "An efficient automatic mass classification method in digitized mammograms using artificial neural network." *arXiv preprint arXiv:1007.5129* (2010).

[13]]Eltoukhy, Mohamed Meselhy, Ibrahima Faye, and Brahim Belhaouari Samir. "A comparison of wavelet and curvelet for breast cancer diagnosis in digital mammogram." *Computers in Biology and Medicine* 40.4 (2010): 384-391.

[14]Rong, Hai-Jun, et al. "Online sequential fuzzy extreme learning machine for function approximation and classification problems." *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 39.4 (2009): 1067-1072.

[15]Lan, Yuan, Yeng Chai Soh, and Guang-Bin Huang. "Ensemble of online sequential extreme learning machine." *Neurocomputing* 72.13 (2009): 3391-3395.

[16]Rashed, Essam A., Ismail A. Ismail, and Sherif I. Zaki. "Multiresolution mammogram analysis in multilevel decomposition." *Pattern Recognition Letters* 28.2 (2007): 286-292.

[17]Haralick, Robert M., and Karthikeyan Shanmugam. "Textural features for image classification." *IEEE Transactions on systems, man, and cybernetics* 3.6 (1973): 610-621.