ISSN: 0976-9102(ONLINE)
DOI: 10.21917/ijivp.2014.0123

WAVELET BASED SEGMENTATION USING OPTIMAL STATISTICAL FEATURES ON BREAST IMAGES

A. Sindhuja[1] and V. Sadasivam[2]

[1]*Department of Computer Science and Engineering, Manonmaniam Sundaranar University, India*
E-mail: a.sin2ja@gmail.com
[2]*PSN College of Engineering and Technology, India*
E-mail: vs_msu@yahoo.com

Abstract

Elastography is the emerging imaging modality that analyzes the stiffness of the tissue for detecting and classifying breast tumors. Computer-aided detection speeds up the diagnostic process of breast cancer improving the survival rate. A multiresolution approach using Discrete wavelet transform is employed on real time images, using the low-low (LL), low-high (LH), high-low (HL), and high-high (HH) sub-bands of Daubechies family. Features are extracted, selected and then finally segmented by K-means clustering algorithm. The proposed work can be extended to Classification of the tumors.

Keywords:
Daubechies Wavelet, Feature Selection, SFFS, K-means

1. INTRODUCTION

Breast cancer is a worldwide threat and common malignancy among women. It is the second most common cancer among women. All around the world there are 1.6 million new cases of breast cancer in woman in 2010 [31]. The WHO has suggested early detection and treatment through education and screening programs [32]. To improve the survival rate, the treatment should be provided at an early stage. Therefore there is a need for a system that detects the tumor automatically that helps in diagnosis. From 2000 to 2007 there is a 7% increase in breast cancer survival. Mammography and sonography are currently used techniques in diagnosis of breast cancer. Physicians have come to accept the grayness of Mammographic images and CT scans (which are based on X-rays), where structures appear as lights and shadows are considered poor substitute for "photographs" of our insides and are difficult to find tumors in small dense breast. Sonography has emerged as a very useful modality in suspecting breast masses. However sonography has limitations on the features of benign and malignant tumors [30]. This leads to 70% to 90% breast biopsies on benign diseases, leading to discomfort and anxiety including unnecessary expenses to the patients. There is a great need for a reliable imaging modality and detection method to improve the diagnosis rate to avoid unnecessary biopsies. Elastography, a tissue strain imaging obtained from tissue compression described by Ophir et al [24], along with sonography is a new emerging modality for identifying tumors. It has been used in a wide range of different applications, including prostate, breast, thyroid, and intravascular ultrasound [13, 17, 19, 28]. The ultrasound scanner is equipped with an elastography unit. The probe is moved inferior and superior with only light pressure to obtain the elasticity images.

Detecting the tumor at an early stage is important for the treatment of cancer. Medical image segmentation is tedious since there is inhomogeneity in image acquisition that affects the intensity of the images. The shape and size of the tumor is necessary for diagnosing the type of tumor. This information can be obtained manually which is a tedious, time consuming and expensive task. An automated segmentation is in demand. Research on Segmentation and classification of breast images using statistical texture features obtained from various modalities such as Mammograms, Sonography, CT, MRI are going on for decades [7, 1, 16]. However texture analysis has not been considered seriously on the sonoelastographic breast images.

The wavelet transform is used in the proposed work for characterizing the texture features since the wavelet transform provides multiscale image analysis [18, 8, 16, 27]. For image decomposition and feature extraction, the Daubechies wavelet [11] has been used as a basic tool in the wavelet transform. For the proposed work the input image is decomposed into sub-bands (three detail and one approximate sub image). GLCM (Gray Level Co-occurrence Matrix) [26], LBP (Local Binary Pattern) [23], edge features [6] are extracted from the sub-bands. The proposed work is to select the optimal feature set that best segments the image. The optimal feature selection algorithm SFFS (Sequential Floating Forward Selection Algorithm) [25, 14] examines every combination of the features extracted. Unsupervised clustering on the selected subset of features is done by the k-means algorithm [12].

The paper is organized as below: Section 2 describes the transformation of the image after the enhancement using Daubechies Wavelet. Section 3 introduces the feature extraction and the selection of the subset of feature vector. Section 4 introduces segmentation using the K-means clustering. The experiments and the results are given in section 5. Section 6 gives conclusion.

2. WAVELET TRANSFORM

Sticks algorithm is a pre-processing step applied for edge enhancement before transforming the images. Czerwinski et al. [10] proposed the stick technique to reduce speckle also to improve edge information. Stick is a sub-optimal but very useful technique for real images and requires simple computation to implement.

The Discrete Wavelet Transform is among the most popularly used wavelet transforms [20]. The Discrete Wavelet Transform (DWT) based on sub-band coding yields a fast computation of Wavelet Transform. The discrete wavelet transform provides the information useful for image processing and is easy to implement, reduces the computation time and resources required. 2D wavelet decomposition is performed by applying the 1D wavelet transform along the rows of the image first, and then finally, corrected interpolated high frequency sub bands and interpolated input image are combined with the help of inverse DWT (IDWT) to achieve a high resolution output image. This results in four decomposed sub-band images low-low (LL), low-high (LH), high-low (HL), and high-high (HH) sub-bands.

Daubechies wavelet is considered for the proposed work because of its compact support and less computational complexity. By varying between six Daubechies filters with one to six taps (abbreviated as DB1, DB2, DB3, DB4, DB5, DB6) images are decomposed into three detail and one approximate sub-band images. The features obtained from these wavelet transformed images are used for texture analysis.

3. FEATURE SELECTION

Extraction and selection of features from the transformed images can be used to detect tumors. Extraction and selection is to find an appropriate feature set that can distinguish tumor / non-tumor tissues. Feature extraction is collecting the descriptors or quantitative feature measurements, for quantifying the texture characteristics. When the feature space is large and complex, extracting and selecting the most effective features is very important. This process reduces the redundancy of feature space to produce an optimum feature set.

The texture features used in the proposed work are GLCM, LBP. Edge features are also included for the proposed work since edges plays a major role in segmentation. All the features cannot be used at the same time. The best performing feature set has to be found. The features are extracted from each of the wavelet decomposition sub-band. Eighteen features of GLCM and one feature from LBP and thirteen edge features are calculated for the six Daubechies wavelet taps for the four wavelet decomposed sub-bands and formed 768 feature vectors. The features are selected by the popular feature selection technique SFFS, that avoids the "curse of dimensionality". In the proposed work three subsets with the first five, ten, twenty features are taken for experimentation in finding the optimal set for segmentation. The 768 feature vectors are reduced to three subsets with 120, 240, 480 feature vectors.

GLCM are the number of transitions between all pairs of two pixels with grey levels i, j by varying distance d and angular orientation θ. $N_{d,\theta}(i, j)$ is the number of pixel-pairs at location (x,y) and (w,z) satisfying the following conditions: $G(x,y) = i$, $G(w,z) = j$, and the distance measure $|(x,y) - (w,z)|_{dm} = (d, \theta)$. GLCM is the widely used texture analysis method in medical imaging. Literature shows that GLCM has been used in medical image segmentation [2, 3, 4, 9].

LBP provides a unified description including both statistical and structural characteristics of a texture patch, so that it is more powerful for texture analysis. LBP 3 × 3 model is where the difference between the central pixel and its neighbors are used for thresholding, then the pixels are labeled resulting in 28 patterns and the 256 bin occurrence histogram computed over a region is then employed for texture description. Ojala and peitikainen [23] came up with efficient multi resolution approach to grayscale and rotation invariant texture based on LBP.

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c), & U(LBP_{P,R}) \le 2, \\ P+1, & otherwise. \end{cases} \quad (1)$$

where, the center pixel of the local neighbourhood is g_c and $g_p(p = 0\ldots P\text{-}1)$ corresponds to the gray values of P equally spaced pixels on a circle of radius R ($R > 0$) that forms a circularly symmetric neighbour set and the uniformity measure U, corresponds to the number of spatial transitions that describes the "uniform" pattern.

$$U(LBP_{P,R}) = |s(g_{p-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)| \quad (2)$$

The patterns that have U value of at most 2 are "uniform". Research is going on in the field of medical image processing where LBP is extracted for use in various imaging modalities [5, 15, 22].

Edge Features based on the magnitude and direction is used for segmentation. The gradient is used for extraction of the edge based features. The features are also captured from the edge direction and intensity homogeneity used by Boland [6].

SFFS allows dynamic number of features that have been selected to be removed in a dynamic number of posterior steps (forward and backtracking mechanisms). So, SFFS is very efficient and effective on problems of high dimensionality with non-monotonic feature selection criterion functions.

4. SEGMENTATION

Unsupervised texture segmentation based on the k-means clustering has been reported in the literature for the past few decades [21, 29]. It is one of the simplest and widely used unsupervised learning algorithms that can solve the well-known clustering problems with low computational complexity. K-means has the ability to cluster huge data points very quickly. The clustering process is carried on the chosen feature vectors with random initial centers. K-means is suitable for the medical image segmentation since the number of clusters is known in prior for particular regions of human anatomy. The proposed work focuses on segmenting the image into two regions, the tumor and non-tumor region.

K-means clustering algorithm assigns each feature vector of the image to the cluster that minimizes the distance between the feature vector and the cluster center. Re-compute the cluster centers by averaging the entire feature vector in the cluster. Repeat these steps until no feature vector change clusters. The output of the K-means is a binary image for the proposed work.

Morphological processing is used to extract few properties of the image. Morphological operations are used for post processing in the proposed work. Morphological operations are suited for binary images since they rely upon the relative ordering of the pixels, not the numerical values. Morphological operations such as dilation, erosion, etc are performed on the resultant image to remove the negligible spots present and to provide the exact boundary of the tumors. The shape obtained can be used for further classification.

5. RESULTS

5.1 DATA SET

Real time images of 15 Patients are acquired for experimentation. The ground truth or the hand segmented image data set are also taken. These images have both benign and malignant lesions. The images are acquired from Siemens Acuson Antares ultrasound scanner with a 7.3 MHz linear array transducer. A light compression produced by the beating heart and the breathing action gives an elasticity image from a sonographic image. The gray scale elasticity image is considered for the proposed work. The ground truth images are used in evaluating the automated segmented image.

5.2 EXPERIMENTS AND RESULT

The images are denoised by sticks filter by varying the length and thickness. The processed image is then decomposed by Daubechies wavelet from one to six taps resulting in 24 sub-bands, each tap resulting in four sub-bands (LL, LH, HH and HL). The GLCM, LBP, Edge features are extracted from each sub-band. SFFS technique ranks the extracted 768 feature vectors. Three subsets with 480, 240, 120 (DS1, DS2, DS3) feature vectors are the final data sets which are taken for experimentation. K-Means clustering algorithm is iterated using random cluster centers until the maximum values of performance measures are found. The clusters are formed by minimizing the Euclidean distance between the data and the corresponding cluster centroid. The resultant image is evaluated by comparing it with the ground truth image provided by the expert radiologist.

There is no gold standard evaluation for medical image segmentation. The proposed work is evaluated by the measures accuracy, overlap and combination metric that include metrics such as overlap, over-segmentation and under-segmentation. The evaluation measures are given as:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{3}$$

Overlap (also known as AOM, Jaccard similarity measure, or the Tannimoto coefficient)

$$P_1 = \frac{|S \cap G|}{|S \cup G|} \tag{4}$$

Under-Segmentation

$$P_2 = \frac{|G - (S \cap G)|}{|G|} \tag{5}$$

Over-Segmentation

$$P_3 = \frac{|S - (S \cap G)|}{|S|} \tag{6}$$

Combined Metric

$$P = (P_1 + (1\text{-}P_2) + (1\text{-}P_3))/3 \tag{7}$$

where,

TP is True Positive - Tumor Pixel diagnosed correctly as Tumor Pixel,

TN is True Negative - Non-Tumor Pixel diagnosed correctly as Non-Tumor Pixel,

FP is False Positive - Non-Tumor Pixel diagnosed incorrectly as Tumor Pixel,

FN is False Negative - Tumor Pixel diagnosed incorrectly as Non-Tumor Pixel,

S - System Segmented Image,

G - Ground Truth Image.

The Optimal feature set is selected from the evaluated performance of the segmentation from the above measures. The sub-band from the six versions of Daubechies that gives the best results for the optimal feature set is also shown. The under-segmentation and over-segmentation results are negligible since the numerical values are very low. Graphical representation of the evaluation measures combination metric, accuracy and overlap for the datasets DS1, DS2, DS3 are given in Fig.1, Fig.2 and Fig.3 respectively.

The results show that the dataset with the first 20 features (DS1) performs well than the other two datasets when decomposed with the Daubechies fourth tap filter (DB4) in the approximate CA sub-band. The visual results are also shown for the DB4 wavelet for the first 20 features in the CA sub-band in Fig.4. Fig.4(a) and Fig.4(d) are the input images that contain tumors. Fig.4(b) and Fig.4(e) are the delineated tumor boundary by the proposed work. Fig.4(c) and Fig.4(f) are the ground truth images where the boundary is hand drawn by an expert radiologist.

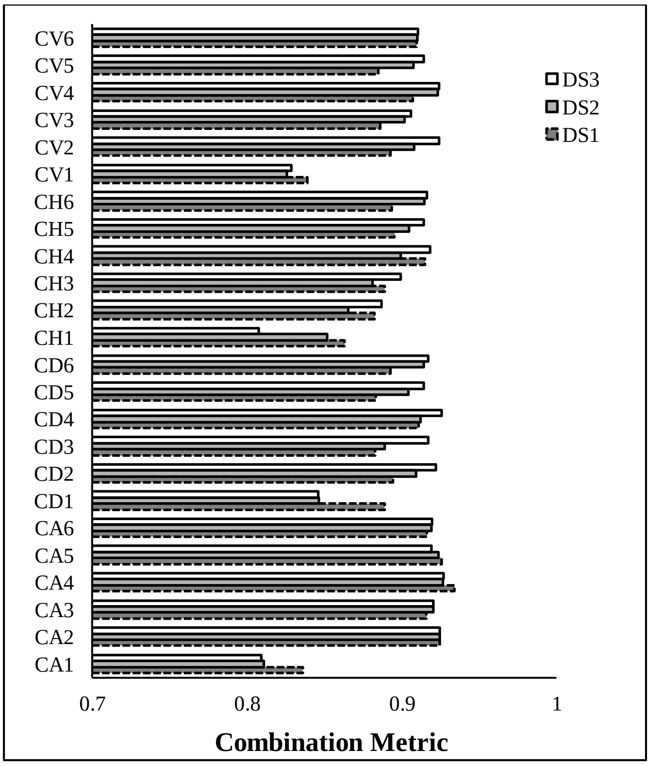

Fig.1. Combination metric for Daubechies wavelet filter (six taps, four sub-bands)

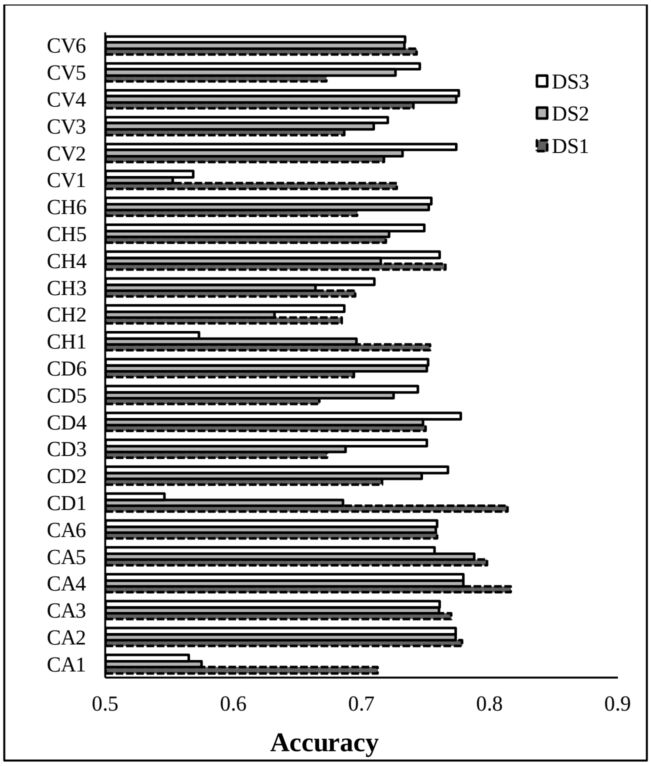

Fig.2. Accuracy for Daubechies wavelet filter (six taps, four sub-bands)

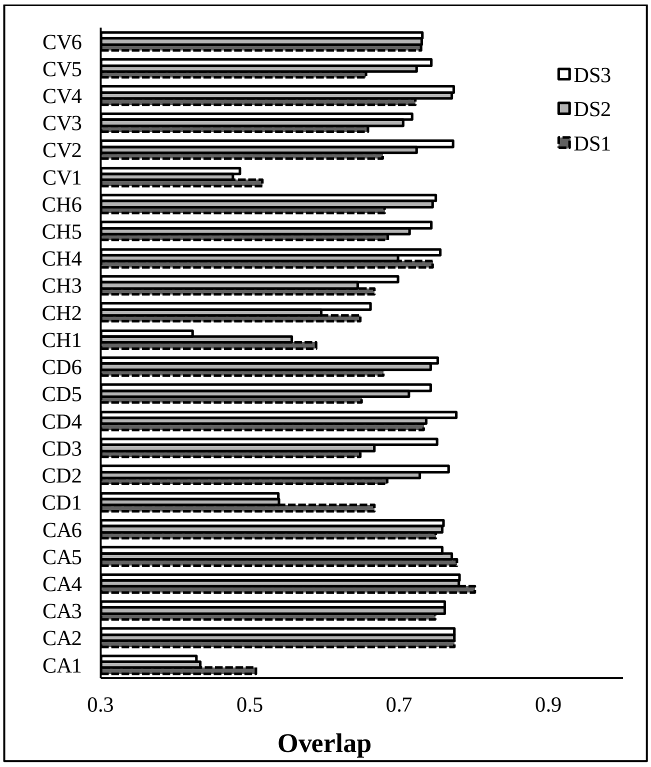

Fig.3. Overlap measure for Daubechies Wavelet filter (six taps, four sub-bands)

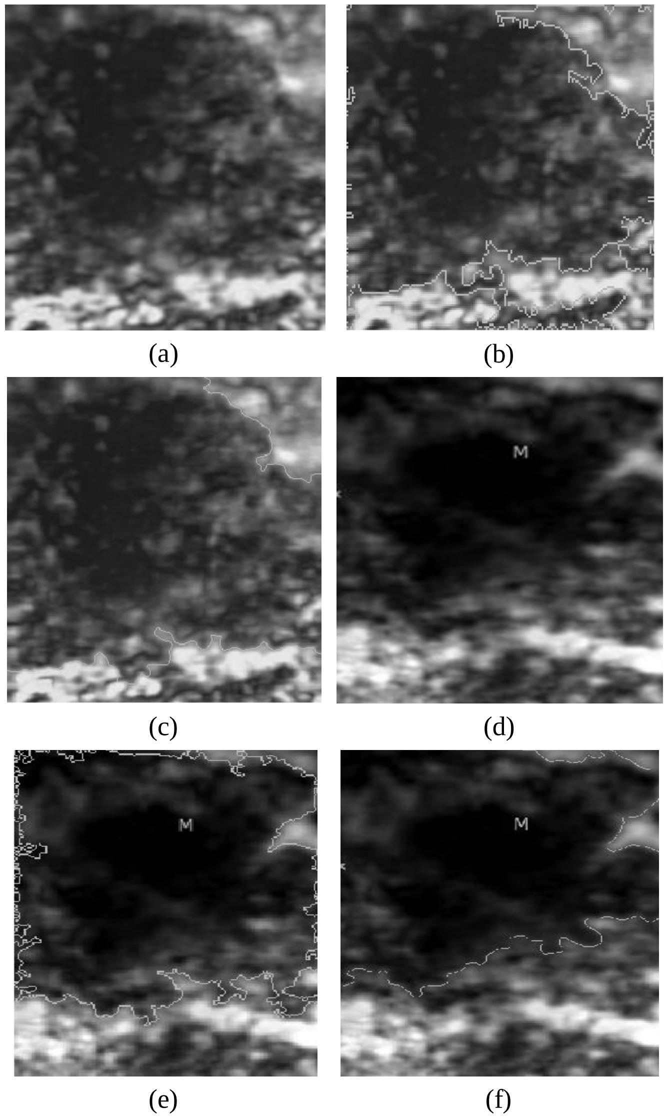

Fig.4. (a, d) – Input Images, (b, e) – Segmented Images, (c, f) – Ground truth images

6. CONCLUSION

Breast cancer is the second worldwide cancer that causes death among women. Sonoelastography is the emerging imaging modality when combined with sonography detects cancer that reduces the unnecessary biopsies compared to other techniques. For the last few decades wavelet and texture analysis plays an important role in medical image analysis. The proposed work uses the texture (GLCM and LBP), edge features extracted from the wavelet decomposed image for segmentation of the tumor from the image. The Daubechies family of order one to order six is considered for decomposition. Three set of feature vectors obtained from the extracted features are taken for experimentation. From the evaluation, the approximation coefficients obtained from the first level decomposition of the fourth order of Daubechies wavelet family segments well than the other coefficients among the six wavelet family. The optimal feature set is the first 20 features selected by the SFSS technique. Features other than texture such as shape can be extracted from both the segmented results of the sonoelastographic and sonographic breast images are used for classification of the segmented tumor.

REFERENCES

[1] J. Alison Noble and Djamal Boukerroui, "Ultrasound Image Segmentation: A Survey", *IEEE Transactions on Medical Imaging*, Vol. 25, No. 8, pp. 987 -1010, 2006.

[2] André Victor Alvarenga, Wagner C.A. Pereira, Antonio Fernando C. Infantosi and Carolina M. Azevedo, "Complexity curve and grey level co-occurrence matrix in the texture evaluation of breast tumor on ultrasound images", *International Journal of Medical Physics Research and Practice*, Vol. 34, No. 2, pp. 379-387, 2007.

[3] Anupa Maria Sabu, D. Narain Ponraj and Poongodi, "Textural Features Based Breast Cancer Detection: A Survey", *Journal of Emerging Trends in Computing and Information Sciences*, Vol. 3, No. 9, pp.1329-1334, 2012.

[4] Anuradha Phadke, "*Brain MRI Segmentation using Texture Features: Use of GLCM texture features*", Lambert Academic Publishing, 2012.

[5] Baopu Li and Max Q.H. Meng, "Texture analysis for ulcer detection in capsule endoscopy images", *Image and Vision Computing*, Vol. 27, No. 9, pp. 1336-1342, 2009.

[6] M.V. Boland, "Quantitative Description and Automated Classification of Cellular Protein Localization Patterns in Fluorescence Microscope Images of Mammalian Cells", Ph.D. Dissertation, Department of Biomedical Engineering, CMU, 1999.

[7] G. Castellano, L. Bonilha, L.M. Li and F. Cendes, "Texture analysis of medical images", *Clinical Radiology*, Vol. 59, No. 12, pp.1061-1069, 2004.

[8] D.R. Chen, R.F. Chang, W.J. Kuo, M.C. Chen and Y.L. Huang, "Diagnosis of breast tumors with sonographic texture analysis using wavelet transform and neural networks", *Ultrasound in Medicine Biology*, Vol. 28, No. 10, pp. 1301-1310, 2002.

[9] H.D. Cheng, Juan Shan, Wen Ju, Yanhui Guo and Ling Zhang, "Automated breast cancer detection and classification using ultrasound images: A survey", *Pattern Recognition*, Vol. 43, No. 1, pp. 299-317, 2010.

[10] R.N. Czerwinski, D.L. Jones and W.D. O'Brien, Jr., "An approach to boundary detection in ultrasound imaging", *Proceedings of the IEEE Ultrasonics Symposium*, 1993.

[11] I. Daubechies, "*Ten Lectures on Wavelets*", SIAM, 1992.

[12] R.O. Duda, P.E. Hart and D.G. Stork, "*Pattern Classification*", 2nd edition, John Wiley & Sons, 2000.

[13] A. Itoh, et. al., "Breast disease: clinical application of US elastography for diagnosis", *Radiology*, Vol. 239, No. 2, pp. 341–350, 2006.

[14] A. Jain and D. Zongker, "Feature selection: evaluation, application, and small sample performance", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, pp. 153-158, 1997.

[15] Kevin W. Bowyer, Karen Hollingsworth and Patrick J. Flynn, "Image understanding for iris biometrics: A survey", *Computer Vision and Image Understanding*, Vol. 110, No. 2, pp. 281-307, 2008.

[16] T.Y. Kim, J. Son and K.G. Kim, "The recent progress in quantitative medical image analysis for computer aided diagnosis systems", *Healthcare Informatics Research*, Vol. 17, No. 3, pp. 143-149, 2011.

[17] K. Konig, Scheipers U, Pesavento A. Lorenz A, Ermert H and Senge T, "Initial experiences with real-time elastography guided biopsies of the prostate", *The Journal of Urology*, Vol. 174, No. 1, pp. 115-117, 2005.

[18] S. Livens, P. Scheunders, Van de Wouwer G and Van Dyck D, "Wavelets for texture analysis, an overview", *Sixth International Conference on Image Processing and Its Applications*, Vol. 2, pp. 581-585, 1997.

[19] A. Lyshchik, et. al., "Thyroid gland tumor diagnosis at US elastography", *Radiology*, Vol. 237, No. 1, pp. 202-211, 2005.

[20] S. Mallat, "*A Wavelet Tour of Signal Processing: The sparse way*", Academic Press, 1999.

[21] Marek Kowal, et. al., "Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images", *Computers in Biology and Medicine*, Vol. 43, No. 10, pp. 1563-1572, 2013.

[22] L. Nanni, A. Lumini and S. Brahnam, "Local binary patterns variants as texture descriptors for medical image analysis", *Artificial Intelligence in Medicine*, Vol. 49, No. 2, pp. 117-25, 2010.

[23] T. Ojala, M. Topi and M. Pietikainen, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, pp. 971-987, 2002.

[24] J. Ophir, I. Cespedes, H. Ponnekanti, Y. Yazdi and X. Li, "Elastography: a quantitative method for imaging the elasticity of biological tissues", *Ultrasonic Imaging*, Vol. 13, No. 2, pp. 111-134, 1991.

[25] P. Pudil, J. Novovicova and J. Kittler, "Floating search methods in feature selection", *Pattern Recognition Lett*ers, Vol. 15, No. 11, pp. 1119-1125, 1994.

[26] Robert M Haralick, K. Shanmugam and Its'hak Dinstein, "Textural Features for Image Classification", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-3, No. 6, pp. 610-621, 1973.

[27] H.K. Sawant and Vaishali D. Shinde, "Breast tumor analysis using texture features and wavelet transform with dynamic neural network based training", *Journal of Information, Knowledge and Research in Computer Engineering*, Vol. 1, No. 2, pp. 46-51, 2011.

[28] J.A. Schaar, et. al., "Characterizing vulnerable plague features with intravascular elastography", *Circulation*, Vol. 108, No. 21, pp. 2636-2641, 2003.

[29] Wan-Ting Lin, Chuen-Horng Lin, Tsung-Ho Wu and Yung-Kuan Chan, "Image Segmentation Using the K-means Algorithm for Texture Features", *World Academy of Science, Engineering and Technology*, Vol. 41, No. 97, pp. 612-615, 2010.

[30] H. Zhi et. al., "Comparison of Ultrasound Elastography, Mammography, and Sonography in the Diagnosis of Solid Breast Lesions", *Journal of Ultrasound in Medicine*, Vol. 26, No. 6, pp. 807-815, 2007.

[31] http://www.komen.org/breastcancer/statistics.html

[32] http://www.worldwidebreastcancer.com/learn/breast-cancer-statistics-worldwide.