Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/110310

Mammographic Mass Detection Using Curvelet Moments

Mohamed Meselhy Eltoukhy[*]

Computer Science Department, Faculty of Computers and Informatics, Suez Canal University, Ismailia, Egypt

Received: 20 Feb. 2016, Revised: 18 Apr. 2016, Accepted: 27 Apr. 2016
Published online: 1 May 2017

Abstract: The aim of this paper is to introduce a robust CAD system that is able to increase the accuracy rate and reduce the false positive detection rate. This paper presents a system based on calculating the second order moment (variance) for the task of mass detection in digital mammogram. The goal is to develop a feature vector which is able to provide an accurate discrimination between the mass and normal tissues. The feature vectors are investigated in terms of their capability to achieve the classification task using Random Forests with 10-fold cross validation. The proposed system has been tested using 1515 images from Image Retrieval in Medical Applications (IRMA) dataset and 265 images from Mammographic Image Analysis Society (MIAS) dataset. The study shows that the second order moment can be used efficiently for mammographic mass detection with accuracy of 100%.

Keywords: Breast cancer, Early detection, Mass detection, Curvelet, Second order moment.

1 Introduction

Breast cancer is one of the leading causes of death among women worldwide. Early detection is the key to reduce the mortality rate. Mammography screening has proven to be one of the effective tools for diagnosis of breast cancer. Computer aided diagnosis (CAD) system is a fast, reliable, and cost-effective tool in assisting the radiologists/physicians for diagnosis of breast cancer. Computer-Aided Diagnosis (CAD) systems play an increasingly important role in improving the accuracy of breast cancer detection. CAD systems help radiologists to detect abnormal regions presented in breast. Developing a CAD system with high capability to pinpoint the abnormal regions in the mammogram is very important and challenging task. It needs to be accurate and precise. Abnormal cells appear in mammograms as highly-bright regions. However, breast has different types of tissues, each of which has different textural variation in intensity. This makes the naked eye examination for the diagnosis of breast abnormality difficult. To reduce the potential high miss detection rate, an objective method to identify and classify mammograms is needed. The main purpose of this research is to develop a system that acts as a second opinion for radiologists. The proposed system can explore the breast tissue types in order to find out the abnormal mammographic mass cells in mammogram. A mass is defined as a space occupying lesion seen in at least two different projections [1]. It is always hidden inside the breast tissue. Hence, several techniques have been developed for the detection and classification of breast masses in mammograms [1]. The proposed systems for breast mass detection are consisting of four steps namely: segmentation, feature extraction, feature selection, and classification. The segmentation step aims to find regions of interest (ROI) that contain the mass region. In the feature extraction step, each ROI is characterized with a set of features to produce the feature vector. Feature extraction is then followed by feature selection step, which identifies the best set of features that can be used to distinguish between the normal and abnormal tissues. The classification step heavily relies on the accuracy of feature extraction step. The greatest difficulty lies in finding some properties of the image from which such features may be extracted. Generally, the extracted features should satisfy the following conditions [2]. Discriminability: The feature reflects the variations between the different classes. Robustness: the feature is able to perform without failure under a wide range of conditions. Invariance: the feature is not influenced by variations. Independence: any of the features cannot be formulated using only the other features from the same set of features. Multi-resolution analysis has been proven to be successful in image analysis. Therefore, the texture features are not affected by the size of the pixel neighborhood. The curvelet transform as a

[*] Corresponding author e-mail: Mohamed_eltokhy@ci.suez.edu.eg

multi-resolution tool have the capabilities of providing better texture discrimination than wavelet [3]. This work focuses on using multi-resolution texture analysis for the mammographic mass detection. In literature, multi-resolution techniques are successfully utilized in breast cancer detection [4] - [10]. The multi-resolution representation has the advantage of representing the edges discontinuities and curves in images efficiently. Ferreira and Borges [5] proposed an algorithm to classify mammogram images into normal and masses using wavelet bases. Their algorithm achieved 94.85% classification rate. Rashed et al. [6] used the multi-resolution wavelet for mammogram analysis by extracting a set of the biggest coefficients. In their study, they used Daubechies-4,-8,-16 wavelet functions with four decomposition levels. The Euclidian distance classifier achieved 87.06% classification rate. Eltoukhy et al. [4] presented an algorithm for mammogram classification using a percentage of biggest coefficients at each decomposition level of the curvelet. Their results show that multi-resolution representations can significantly improve the classification rate with accuracy 98.59%. Moayedi et al. [8] combined between the contourlet coefficients, co-occurrence matrix features and geometrical features to produce the feature vector that represent the ROI. They employed the genetic algorithms to select the most prominent features. Finally, their results concluded that the contourlet features offer an improvement to the classification step. Eltoukhy et al. [9] proposed an optimized feature selection method from the wavelet and curvelet features to find most discriminative features that have high capability to classify normal and abnormal mammogram. Zyout et al. [10] extracted the texture features from wavelet transform and gray level co-occurrence matrix. They applied particle swarm optimization to find the features that have the ability to differentiate between the normal and abnormal regions. Their results show the promising performance of the textural features that are based on co-occurrence matrix of wavelet representation. Jardezi and Faye [11] combined the completed local binary patterns (CLBP) with texture features of curvelet sub-bands to distinguish between mass and normal images selected from MIAS and IRMA datasets. The nearest neighbor classifier was used to evaluate each feature set individually as well as after combining CLBP and curvelet texture features. The classification accuracy rate of 96.68% was achieved. Dhahbi et al. [12] extracted curvelet moments as a feature vector. They investigated two techniques: curvelet moments from each level and from each sub-band, a statistical feature ranking method is used to find the most discriminative features [13].

Recently, Gedik [14] used the features of fast finite shearlet transform as a feature vector to classify digital mammograms. The features of fast finite shearlet transform were ranked according to their capabilities to distinguish between different classes. A thresholding process was implemented to maximize differences between class representatives and classifications were calculated over the optimal feature set using 5-fold cross validation and a support vector machine (SVM) classifier. The present results suggested that the proposed method provides credible classification of mammogram. The contribution in this paper is to combine the advantages of curvelet transform as multi-resolution representation with moment's theory; the normalized central moments have the capability to be invariant to translation, rotation and scale. It is dimensionless quantities that able to represents an independent distribution of any linear changes of sales. Hence, the normalized central moments of curvelet sub-bands have been tested and evaluated as a feature vector to identify the mammographic mass from normal tissues. The remaining of this paper is organized as follows. Section 2 presents the methodology and experimental works. It gives a brief introduction to curvelet and datasets used in the study. Section 3 introduces the results and discussion. Finally, the paper is concluded in Section 4.

2 METHODOLOGY AND EXPERIMENTAL WORK

In the experimental level, the main objective is to combine the advantages of curvelet representation with moment theory to solve the problem of mass identification from the normal tissues. The following subsections present a little description of curvelet and the used datasets. In this study, two different datasets will be used to evaluate the proposed feature vector.

2.1 The curvelet transform

The curvelet transform is a multi-scale decomposition representation method. It has been developed to naturally represent objects in 2D to improve the wavelet limitations for representing geometrical information [3]. Discrete curvelet transform is an image representation approach that codes image edges more efficiently than wavelet transform. Figure 1, illustrates the curvelet tiling.

The curvelet transform is an effective tool for curve finding at multiple resolution levels. As seen in Figure 2, the left side is the original image. The right side shows the distribution of the ROI in different wedges using four decomposition levels with 16 angles. However, the moment features may provide the most discriminating features, for each of curvelet wedge, the second, third and fourth order moments namely variance, skewness, and kurtosis are computed, respectively.

2.2 Dateset

In this study, two different datasets are used. The first one is the Image Retrieval in Medical Applications (IRMA)

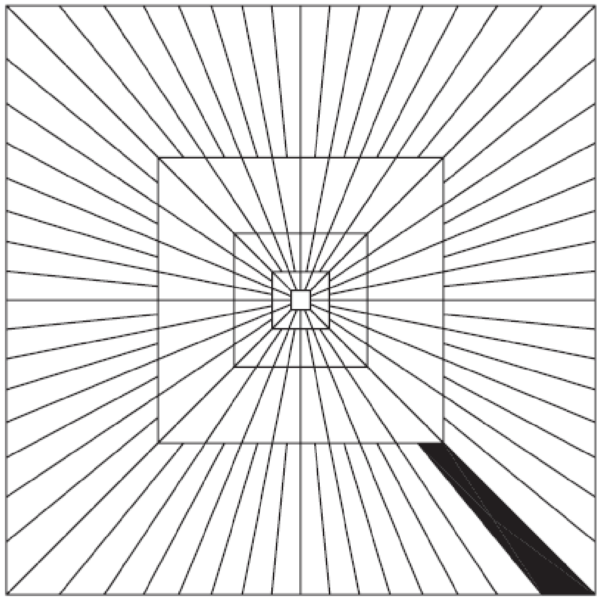

Fig. 1: Curvelet basic digital tiling in two dimensions. The shaded region represents one such typical wedge.

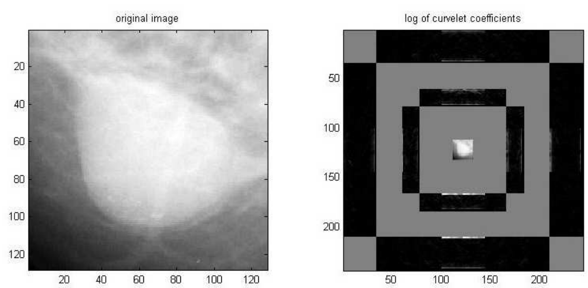

Fig. 2: The curvelet transform of mass ROI. Left: original ROI image, Right: The different wedges representation.

dataset. IRMA is the union of the Mammographic Image Analysis Society (MIAS), Digital Database for Screening Mammography (DDSM), the Lawrence Livermore National Laboratory (LLNL), and routine images from the Rheinisch-Westflische Technische Hochschule (RWTH) Aachen [15]. It consists of 931 normal regions and 584 abnormal (mass) regions. Each region has a size of 128x128. Second is MIAS dataset, we used 207 normal ROI combined with 58 mass regions [16].The original dataset images are 1024x1024 pixels, therefore a cropping operation is needed, the cropping process was performed manually, where the given center of the abnormality area is selected to be the center of ROI. Regions of interest (ROIs) 128x128 are cropped. The details of the used datasets are explained as shown in Table 1.

Table 1: The distribution of IRMA and MIAS dataset

Dataset	Normal tissue	Abnormal mass
IRMA	931	584
MIAS	207	58

2.3 Feature extraction and capability evaluation

The curvelet transform is applied on ROI with scale 4 and 16 angles, i.e. the ROI is decomposed into 81 wedges (the first level gives the low frequency level i.e considered one wedge. The second produces 16 wedges. The third and fourth levels produce 32 in each level). The central moment features (variance, kurtosis and skewness) are calculated for each wedge, so that a total of 243 features are calculated to form a features vector. The moment features are evaluated based on their ability to differentiate between normal and abnormal regions. The mean for each corresponding feature is calculated for each class separately in order to determine the distribution of the feature in the space. The following equations calculate the mean of the feature number i for class a and class b, respectively, where m is the number of images in class a, and n is the number of features.

$$\mu_a = \frac{\sum_{i=1}^{n} f_i(a)}{m_a} \tag{1}$$

$$\mu_b = \frac{\sum_{i=1}^{n} f_i(b)}{m_b} \tag{2}$$

Then, the capability for each feature $C_{(a,b)}(f_i)$ is calculated as the difference between the mean of two classes.

$$C_{(a,b)}(f_i) = \mu_a - \mu_b \tag{3}$$

Figure 3 and Figure 4 present the comparison of the different normalized central moments with IRMA and MIAS datasets, respectively. For simplicity the Figure presents only the difference for each different curvelet level. The Y-axis presents the capability of the different features, while X-axis presents the kurtosis, skewness and variance for four curvelet levels sub-bands. The given Figure shows how the second central moment is able to distinguish between normal and mass classes.

In order to support our previous observation that the variance feature vector have the capability to distinguish between the mass and normal tissues and for comparison purposes; ten different regions of interest (ROI) are selected from the normal class against another ten ROI from mammographic masses. The values of central moment features are compared as illustrated in Figures 5, 6 and 7. The trend from the Figures indicates the significant inherent variations to distinguish between the normal and mass regions. Figure 5 shows the kurtosis of the curvelet sub-bands of the normal and mass ROIs. It is noted that the values of curves are close to each other. This means that this feature is not that efficient to achieve

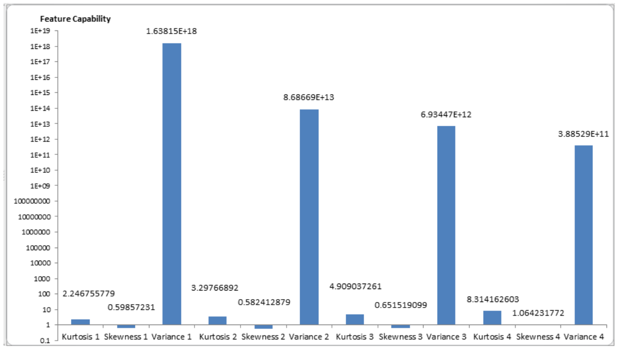

Fig. 3: The capability distribution of different curvelet moment features in IRMA dataset.

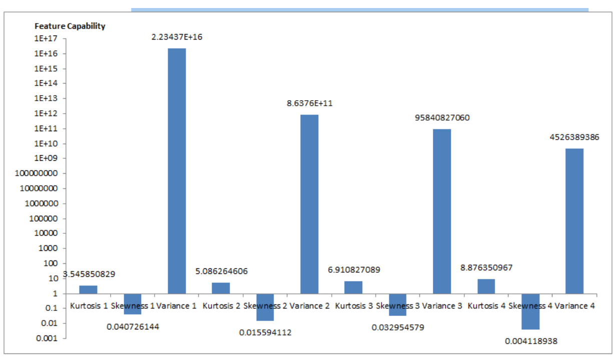

Fig. 4: The capability distribution of different curvelet moment feature in MIAS dataset.

the classification task. Skewness curves of curvelet sub-bands are illustrated in Figure 6. The normal and mass regions are showing overlapping regions, which mean that it is not significant to differentiate between them. In Figure 7, variance shows the maximum difference and best separation between the values of the normal and mass regions. This indicates the robustness of that feature. In conclusion, the computed variance feature vector is the most prominent feature that is able to produce an accurate and reliable classification rate. These observations come parallel along the previous observation obtained from Figures 3 and 4.

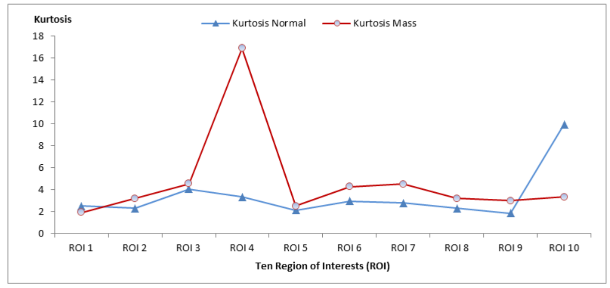

Fig. 5: Comparison of values of kurtosis feature between ten different ROIs from normal and mass classes.

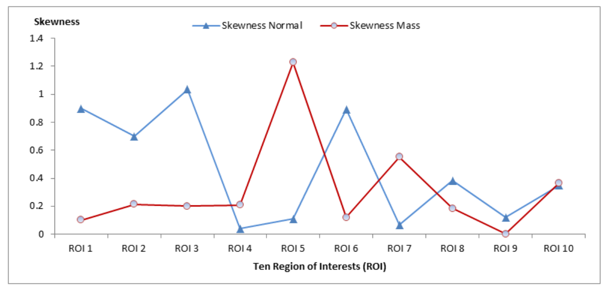

Fig. 6: Comparison of values of skewness feature between ten different ROIs from normal and mass classes.

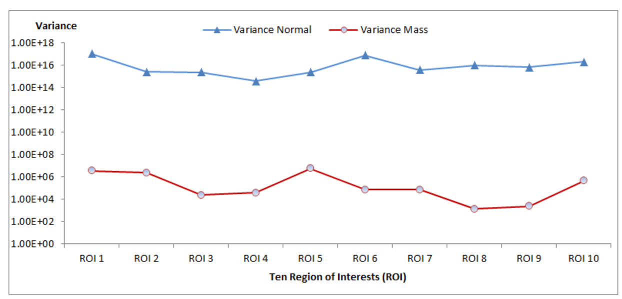

Fig. 7: Comparison of values of variance feature between ten different ROIs from normal and mass classes.

3 RESULTS AND DISCUSSION

In order to evaluate the proposed system, it has been tested using 1515 images from IRMA dataset and 265 images from MIAS dataset. The obtained features are presented to the random forests classifier with 10-fold cross validation. The Evaluation has been measured in relative to the correspondence between two classes in terms of their true positive, true negative, false positive and false negative parts as shown in Figure 8. In addition, receiver operating characteristic (ROC) illustrates the performance of the random forests classifier.

Table 2, presents the classification accuracy rates achieved to distinguish between the mass and normal tissues using IRMA dataset. It is noticed that using all the feature vectors combined for all curvelet sub-bands, the random forests classifier achieved 100% classification accuracy rate. The features were further investigated to

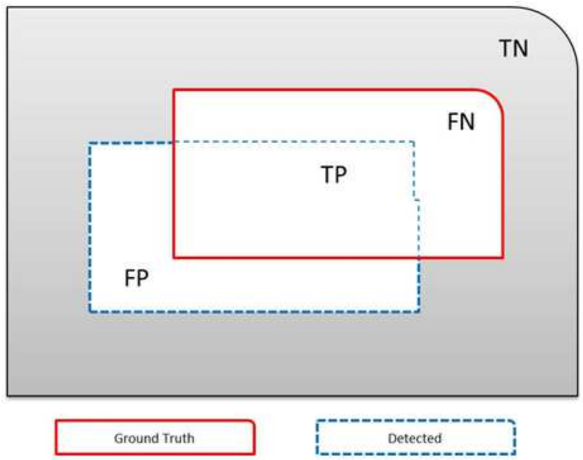

Fig. 8: The true positive (TP), false positive (FP), true negative (TN), and false negative (FN) are presented.

Table 2: The results obtained for IRMA dataset classification with different features using random forests classifier.

	Classes	Correct classified	Accuracy	TP Rate	FP Rate	ROC
All features	Normal	931	100%	1	0	1
	Mass	584		1	0	1
Skewness	Normal	603	54.52%	0.648	0.618	0.526
	Mass	223		0.382	0.352	0.526
Kurtosis	Normal	614	57.62%	0.662	0.557	0.557
	Mass	259		0.443	0.34	0.557
Variance	Normal	931	100%	1	0	1
	Mass	584		1	0	1

Table 3: The results obtained for MIAS dataset classification with different features using random forests classifier.

	Classes	Correct classified	Accuracy	TP Rate	FP Rate	ROC
All features	Normal	207	100%	1	0	1
	Mass	58		1	0	1
Skewness	Normal	174	71.32%	0.841	0.741	0.546
	Mass	15		0.259	0.159	0.546
Kurtosis	Normal	166	66.79%	0.802	0.810	0.49
	Mass	11		0.190	0.198	0.49
Variance	Normal	206	99.62%	1	0	1
	Mass	58		1	0	1

test the claim that variance feature vector have the highest capability to distinguish between the different classes. The accuracy was 54.52% for skewness, 57.62% for kurtosis and 100% for variance. It is apparent that the variance feature vectors distributions are the most discriminative among all vectors, i.e. variance outperforms kurtosis and skewness which confirms our initial observation about the robustness of the variance feature.

On the other hand, for the MIAS dataset, classification accuracy rates are presented in Table 3. It shows that the random forests classifier successfully achieved 100% classification accuracy rate using all central moments features for each curvelet sub-bands. However, the variance feature vector is able to achieve 99.62% classification rate, the skewness and kurtosis were able to achieve only 71.32% and 66.79% respectively. These results prove the effectiveness of curvelet based moments in mammogram analysis and curvature representation. This was expected since the curvelet transform is able to capture the multi-dimensional features in wedges as opposed to points in wavelet transform. These results strongly suggest that the Second order moments (variances) outperform the third (skewness) and fourth (kurtosis) moments.

4 CONCLUSIONS

Mammogram images are critical resources for breast cancer detection if robust features are used. Reliable and accurate features may greatly impact the patient life. This paper proposed a CAD based on finding a robust feature vector that takes the advantages of both curvelet transform and moment theory. The obtained feature vector is then presented to random forests classifier using 10 fold cross validation. Experiments are applied on real labelled data and results show promising use of this approach with mammographic mass detection. An accuracy rate of 100% is achieved by using all features for each ROI sub-bands to classify between the normal and mammographic mass. In order to simplify the proposed method, the variance feature for the first curvelet level alone is tested and produced 100% classification rate as well. The obtained results support our previous notation that the variance feature vector can be used efficiently for mammographic mass detection with accuracy of 100%. We believe that the high successful classification rate achieved is a result of combining the curvelet representation with moment theory.

Acknowledgment

The author is grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper. In addition, he thanks Department of Medical Informatics, RWTH Aachen, Germany for providing IRMA dataset.

References

[1] H.D. Cheng, X.J. Shi, R. Min, L.M. Hu, X.P. Cai, and H.N. Du, Approaches for automated detection and classification of masses in mammograms, Pattern Recognition, vol. 39, Issue 4, , pp. 646-668, 2006.
[2] Jan Flusser, Tom Suk and Barbara Zitov , Moments And Moment Invariants In Pattern Recognition, John Wiley & Sons, Ltd. ISBN: 978-0-470-69987-4, 2009.
[3] E. Candes, L. Demanet, D. Donoho, and L. Ying, Fast discrete curvelet transforms, Multiscale Model. Simul., vol. 5, pp. 861-899, 2006.

[4] M. M. Eltoukhy, I. Faye, and B. B. Samir. Breast cancer diagnosis in digital mammogram using multiscale curvelet transform. Computerized medical imaging and graphics, 34(4):269276, 2010.

[5] C. B. R. Ferreira, D. L. Borges. Analyses of mammogram classification using a wavelet transform decomposition. Pattern Recognition Letters 24, 2003, 973982.

[6] E.A. Rashed, I.A. Ismail, S.I. Zaki., Multiresolution mammogram analysis in multilevel decomposition. Pattern Recognition Letters 28, 2007, 286292.

[7] M. M. Eltoukhy, I. Faye, and B. B. Samir. A comparison of wavelet and curvelet for breast cancer diagnosis in digital mammogram. Computers in Biology and Medicine, 40(4):384391, 2010.

[8] F. Moayedi, Z. Azimifar, R. Boostani, and S. Katebi. Contourlet-based mammography mass classification using the svm family. Computers in Biology and Medicine, 40(4):373383, 2010.

[9] M. M. Eltoukhy, and I. Faye, An optimized feature selection method for breast cancer diagnosis in digital mammogram using multiresolution representation. Appl. Math, 8(6):29212928, 2014.

[10] I. Zyout, J. Czajkowska, and M. Grzegorzek. Multi-scale textural feature extraction and particle swarm optimization based model selection for false positive reduction in mammography. Computerized Medical Imaging and Graphics, 46, Part 2:95 107, 2015.

[11] S.J.S. Gardezi, I. Faye, Fusion of Completed Local Binary Pattern Features with Curvelet Features for Mammogram Classification. Appl. Math 9 (6), 3037-3048, 2015.

[12] S. Dhahbi, W. Barhoumi, and E. Zagrouba. Breast cancer diagnosis in digitized mammograms using curvelet moments. Computers in biology and medicine, 64:7990, 2015.

[13] M. M. Eltoukhy, I. Faye, and B. B. Samir, A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation. Computers in biology and medicine, 42(1):123128, 2012.

[14] N. Gedik, A new feature extraction method based on multi-resolution representations of mammograms, Applied Soft Computing, 44: 128-133, 2016.

[15] Jlia E. E. Oliveira, Mark O. Gueld, Arnaldo de A. Arajo, Bastian Ott, and Thomas M. Deserno. Toward a standard reference database for computer-aided mammography, Proceedings of SPIE - The International Society for Optical Engineering, 2008.

[16] MIAS Available at: http://www.wiau.man.ac.uk/services/MIAS/MIASweb.html.

Mohamed Meselhy Eltoukhy is an Associate professor at Computer Science Department, Faculty of Computers and Informatics, Suez Canal University, Egypt. He received his PhD in Electrical and Electronic Engineering, Universiti Teknologi PETRONAS, Malaysia in 2011. In November 2012 he joined Center for Intelligent Signal and Imaging Research, Universiti Teknologi PETRONAS as a postdoctoral fellow. He has published in several peer reviewed reputed journals and conference proceedings. He is a reviewer for several international journals such as IEEE Journal of Biomedical and Health Informatics, Biomedical Signal Processing and Control, Chaos, Solitons and Fractals, Computers and Electrical Engineering, Computers in Biology and Medicine, Digital Signal Processing, Elsevier, Journal of Visualized Experiments Expert Systems with Applications and Expert Systems. His main research interests are: Artificial Intelligence, Medical Image Analysis and Multiresolution Representation.