World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:11, No:6, 2017

An Approach Based on Statistics and Multi-Resolution Representation to Classify Mammograms

Nebi Gedik

***Abstract*—**One of the significant and continual public health problems in the world is breast cancer. Early detection is very important to fight the disease, and mammography has been one of the most common and reliable methods to detect the disease in the early stages. However, it is a difficult task, and computer-aided diagnosis (CAD) systems are needed to assist radiologists in providing both accurate and uniform evaluation for mass in mammograms. In this study, a multiresolution statistical method to classify mammograms as normal and abnormal in digitized mammograms is used to construct a CAD system. The mammogram images are represented by wave atom transform, and this representation is made by certain groups of coefficients, independently. The CAD system is designed by calculating some statistical features using each group of coefficients. The classification is performed by using support vector machine (SVM).

***Keywords*—**Wave atom transform, statistical features, multi-resolution representation, mammogram.

I. Introduction

BREAST cancer is one of the most common types of cancer threatening the lives of women around the world. It is considered as the second main cause of cancerous death [1]. Saving lives and extending survival time can be possible through the early detection of the cancer. The early detection can made possible through the periodic screening and examination of breasts. In this regard, the most reliable and common method is the mammogram assisting radiologists for the early detection and treatment planning [2], [3]. For all that, interpretation of a mammogram is not an easy task [4]; it requires accurate analysis to avoid misinterpretation of breast cancer cases. Therefore, automated detection systems, so-called CAD, are being developed to achieve more accurate diagnosis [5]-[7].

The generation of a CAD system has several steps, of which, feature extraction is one of the most effective among them. In literature, some of the studies have focused on feature extraction employing multiresolution techniques. Khan et al. [8] studied six different approaches to classify mass in mammograms using directional feature extraction. In those approaches, directional textural features were extracted via a Gabor filters bank. The directional textural features provide information about structural properties of ROIs. Classification was performed using Successive Enhancement Learning based on weighted Support Vector Machine (SELwSVM) to distinguish the ROIs as normal, malignant and benign. Gedik [9] proposed a new feature extraction method to classify mammograms. The method used fast finite shearlet transform (FFST) to extract the features. In the study, feature vectors for every ROI were created using coefficients of the FFST, and feature matrix was built using them. To obtain the most effective features, a feature selection process was carried out based on t-test statistics and dynamic thresholding. Result of the classification was obtained using 5-fold cross validation, SVM and effective feature set. Guo et al. [10] represented a hybrid method to detect micro-calcification in digital mammograms. The method is conducted in several stages: first stage is removing label and pectoral muscles using the largest connected region marking and region growing method, and application of the combination of double top-hat transform and grayscale-adjustment function to enhance micro-calcification; second stage is to retain the significant information by modifying the contourlet coefficients using nonlinear function; third stage is classification by using the non-linking simplified pulse-coupled neural network. Gedik et al. [11] represented another CAD system which has a feature extraction approach based on wave atom transform. After decomposing the ROIs in wave atom domain, feature vectors were composed using coefficients of the transform for every ROI. SVM classifier was used to classify the ROIs as normal, benign and malignant with and without feature selection process. The feature selection was performed using principle component analysis (PCA) method. Francis et al. [12] proposed a method to detect abnormality in breast thermograms. Curvelet transform based statistical and texture features were extracted from thermograms, and used to feed SVM for classification.

In this study, a method was presented to classify mammograms as normal and abnormal using statistical features obtained from wave atom sub-bands. Initially, wave atom transform was applied to ROIs, and coefficients of the transform are obtained. Subsequently, some statistical features (mean, energy, standard deviation, entropy, variance and maximum value) were calculated from the coefficients of each band and each scale of the transform. Using these statistical features, feature matrix was constituted. Every feature in the matrix is ranked to determine the effectiveness, and feature selection was performed applying dynamic thresholding over the ranking values. The feature matrix was rebuilt for every

Nebi Gedik is with the Department of Maritime Management, Faculty of Marine Sciences, Karadeniz Technical University, Surmene, 61530 Trabzon, Turkey (e-mail: ngedik@ktu.edu.tr).

thresholding using the statistical features that remains after thresholding. Classification was repeated for every rebuilt feature matrix by using SVM classifier. The feature set that gives the best classification performance is chosen as optimal feature set and the classification is performed again using optimal features and 5-fold cross validation to validate the result.

II. Wave Atom Transform

The wave atom transform was introduced by Demanet and Ying [13] in 2007 as a new member in the family of oriented, multiscale transforms. The aim of the construction of the transform is to overcome the weakness of wavelet transform when representing high dimensional signals, whereat it is a variation of wavelet transform.

To define the 2D wave atoms, $\varphi_\mu(x)$ is used with subscript $\mu = (j, m, n) = (j, m_1, m_2, n_1, n_2)$ [13]. The subscripts (j, m, n) index to scale, direction and location of the corresponding wave atom, respectively. Those quantities (j, m_1, m_2, n_1, n_2) index a point (x_μ, ω_μ) in phase-space providing;

$$x_\mu = 2^{-j} n, \omega_\mu = \pi 2^j m, \tag{1}$$

$$C_1 2^j \leq \max_{i=1,2} |m_i| \leq C_2 2^j \tag{2}$$

where C_1 and C_2 are positive constants. x_μ is the position vector and the center of $\varphi_\mu(x)$, and ω_μ is the wave vector which determines the centers of both bumps [13]. When considering the wave atom and wavelet transform, there are two differences between them. First, while wavelet transform obey a linear scaling wavelength, wave atom transform obey the parabolic scaling wavelength. Second, the basis function of wave atom transform is different from wavelet transform.

III. Proposed Method

Initially, coefficients of the wave atom transform were obtained by applying to ROIs. Then, to produce the feature vectors, six statistical features (mean, energy, standard deviation, entropy, maximum value and variance) were extracted from the coefficients for each ROI. The statistical features were obtained from the coefficients of every band (packet of coefficients) of wave atom transform. Wave atom transform bring forth the coefficients in four scales and two bands in every scale. So, 48 features were extracted from wave atom sub-bands for each ROI, and feature matrix was built using them. In order to detect the most significant features, a feature ranking and thresholding techniques were performed. The ranking was the process that every feature was ranked based on the statistical *t*-test technique using (3) [14] to determine the capability of the feature in terms of class discrimination.

$$v\_r = \frac{|\mu_{c1} - \mu_{c2}|}{\sqrt{\frac{(s_{c1})^2}{N_{c1}} + \frac{(s_{c2})^2}{N_{c2}}}} \tag{3}$$

where v_r denotes value of ranking and μ and s are means and standard deviations of the class respectively, and N is numbers of ROIs in each class. The thresholding was the process that was applied over the values of ranking for changing the feature set to determine the most effective one. After thresholding, the feature matrix was rebuilt by using remained features to present to SVM classifier. The classification is repeated dividing the data set into a training group (70% of the dataset) and a testing group (30% of the dataset) for all threshold values that are 48 in this study to obtain the optimal point (point of effective feature set) that gives the best classification performance with the minimum number of coefficients. Finally, the optimal point is chosen and the classification is performed via 5-fold cross validation using the feature set at that point to validate the result. The flowchart of the present method is illustrated in Fig. 1.

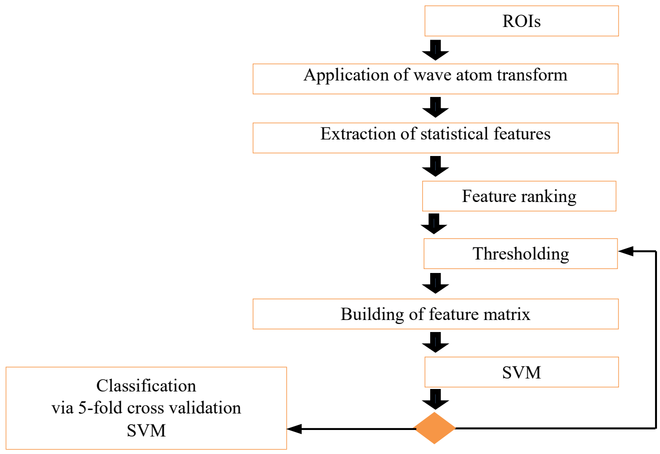


Fig. 1 Flowchart of the present system

TABLE I
DEMONSTRATION OF THE ROIS ACCORDING TO THE TYPE OF ABNORMALITY

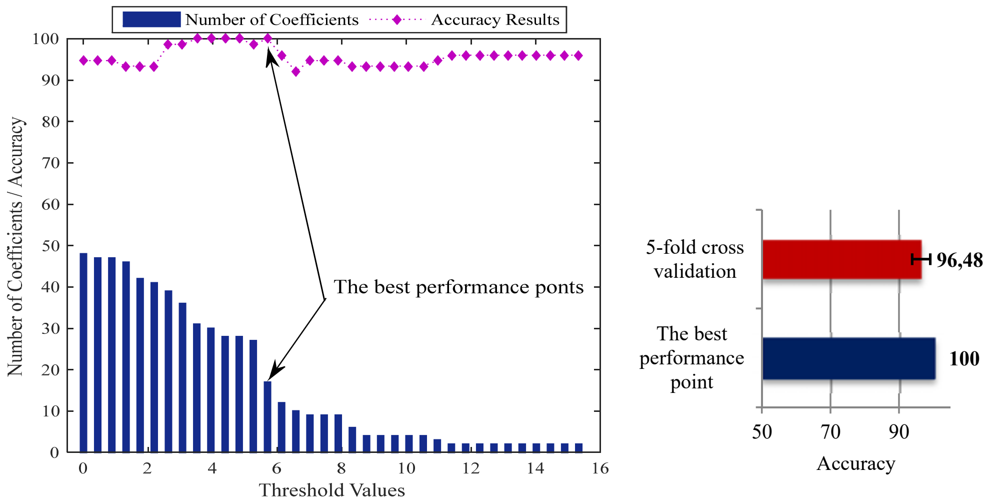
Abnormality	Benign	Malignant	Total
Microcalcification	12	12	24
Circumscribed masses	19	4	23
Ill-defined masses	7	7	14
Spiculated masses	11	8	19
Architectural distortion	9	10	19
Asymmetry lesion	6	9	15
Normal tissue	–	–	114
Total	64	50	228

IV. Experimental Work

ROIs that were used in this study were obtained from DDSM database [15], which consists of 2548 mammograms (including 914 malignant, 870 benign and 764 normal). The mammograms in the database were previously labeled by radiologists. To test the proposed system, 228 mammogram images (114 normal, 50 malignant, 64 benign) were selected to construct ROI set. ROIs were obtained at 128 × 128 pixels size by using manual cropping operation from original mammograms. Manual cropping for abnormal images was performed considering that the center of ROI corresponded to center of abnormality which is determined by radiologists.

Manual cropping for normal images was randomly performed including all tissue types equally (fatty, fatty–glandular, and dense–glandular).

In the present method, the initial step was application of wave atom transform to ROIs and obtaining the coefficients of it. Subsequently, statistical features were extracted from the coefficients of every band (8 bands with 4 scales). Feature matrix was created using these statistical features. Each feature in the matrix was ranked based on t-test statistics. Using rank values and dynamic thresholding, a feature selection was performed to determine the most significant features. Dynamic thresholding was carried out over values of v_r, and the feature matrix was recreated with each changing threshold using the features remained after thresholding. Classification was performed using SVM classifier dividing ROIs into a training group (70% of the ROIs) and a testing group (30% of the ROIs. In this study, classification was performed 48 times for 48 threshold values. Among the classification results, the threshold point that provides highest accuracy with the minimum number of features was chosen as the most significant features point. To validate the result at optimal point, the classification is performed via 5-fold cross validation using the feature set at that threshold point.



(a) (b)

Fig. 2 Normal-abnormal classification for (a) 70-30% ratio of all threshold values (b) 5-fold cross validation

V. Results and Discussions

The ROIs, in the present study with the proposed system, was classified as normal and abnormal. The classification performance of the system with SVM classifier is illustrated in Fig. 2 (a) corresponding to the number of extracted features with different threshold values. The best performance was obtained with 17 features as 100%. There are several points that reach the same accuracy. However, that point was chosen because it had the minimum number of features. Fig. 2 (b) illustrates the classification results of 5-fold cross validation and 70-30% ratio of the dataset using optimal feature set. The result (96.48%) of 5-fold cross validation was represented with error-bar that present the standard deviation between different folds. The results show that the present system has good classification performance decreasing the number of features from 48 to 17. Hence, the system provides a capability to classify mammograms as normal and abnormal.

VI. Conclusions

This paper addresses a method classified mammograms as normal and abnormal by using statistical features based on wave atom transform, t-test statistics, and dynamic thresholding. Considering a CAD system, feature extraction is a key issue in terms of classification because more effective features provide more accurate results. The proposed method represents a good classification performance with selected significant features.

References

[1] Malvezzi, M., Bertuccio, P., Levi, F., La Vecchia, C., & Negri, E. (2014). European cancer mortality predictions for the year 2014. *Annals of Oncology, 25*(8), 1650–1656.

[2] Pisano, E. D., Cole, E. B., Hemminger, B. M., Yaffe, M. J., Aylward, S. R., Maidment, A. D., ... & Fajardo, L. L. (2000). Image Processing Algorithms for Digital Mammography: A Pictorial Essay 1. *Radiographics, 20*(5), 1479–1491.

[3] Bozek, J., Mustra, M., Delac, K., & Grgic, M. (2009). A survey of image processing algorithms in digital mammography. In *Recent advances in multimedia signal processing and communications* (pp. 631-657). Springer Berlin Heidelberg.

[4] Moayedi, F., Azimifar, Z., Boostani, R., & Katebi, S. (2010). Contourlet-based mammography mass classification using the SVM family. *Computers in biology and medicine*, *40*(4), 373-383.

[5] Jiang, J., Yao, B., & Wason, A. M. (2007). A genetic algorithm design for microcalcification detection and classification in digital mammograms. *Computerized Medical Imaging and Graphics*, *31*(1), 49-61.

[6] Verma, B., McLeod, P., & Klevansky, A. (2009). A novel soft cluster neural network for the classification of suspicious areas in digital mammograms. *Pattern Recognition*, *42*(9), 1845-1852.

[7] Marrocco, C., Molinara, M., D'Elia, C., & Tortorella, F. (2010). A computer-aided detection system for clustered microcalcifications. *Artificial intelligence in medicine*, *50*(1), 23-32.

[8] Khan, S., Hussain, M., Aboalsamh, H., & Bebis, G. (2015). A comparison of different Gabor feature extraction approaches for mass classification in mammography. *Multimedia Tools and Applications*, 1-25.

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

[10] Guo, Y. N., Dong, M., Yang, Z., Gao, X., Wang, K., Luo, C., ... & Zhang, J. (2016). A new method of detecting micro-calcification clusters in mammograms using contourlet transform and non-linking simplified PCNN. *Computer Methods and Programs in Biomedicine*, *130*, 31-45.

[11] Gedik, N., & Atasoy, A. (2014). Performance evaluation of the wave atom algorithm to classify mammographic images. *Turkish Journal of Electrical Engineering & Computer Sciences*, *22*(4), 957-969.

[12] Francis, S. V., Sasikala, M., & Saranya, S. (2014). Detection of breast abnormality from thermograms using curvelet transform based feature extraction. *Journal of medical systems*, *38*(4), 1-9.

[13] Demanet, L., & Ying, L. (2007). Wave atoms and sparsity of oscillatory patterns. *Applied and Computational Harmonic Analysis*, *23*(3), 368-387.

[14] Liu, H., Li, J., & Wong, L. (2002). A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns. *Genome informatics*, *13*, 51-60.

[15] DDSM database. http://marathon.csee.usf.edu/Mammography/Database.html, (accessed 05.03.2016).