Artificial Neural Network-Based Classification System for Lung Nodules on Computed Tomography Scans

Emre Dandıl[1,3], Murat Çakıroğlu[2], Ziya Ekşi[3], Murat Özkan[3,4], Özlem Kar Kurt[5], Arzu Canan[6]

[1]Bilecik Vocational High School, Bilecik Şeyh Edebali University, Bilecik, Turkey, emre.dandil@bilecik.edu.tr
[2]Faculty of Technology, Mechatronics Engineering, Sakarya University, Sakarya, Turkey, muratc@sakarya.edu.tr
[3]Faculty of Technology, Department of Comp. Eng., Sakarya University, Sakarya, Turkey, ziyae@sakarya.edu.tr
[4]Bolu Vocational High School, Abant Izzet Baysal University, Bolu, Turkey, muratozkan@ibu.edu.tr
[5]Faculty of Medicine, Department of Chest Diseases, Abant İzzet Baysal University, Bolu, Turkey, aghhozlem@yahoo.com
[6]Faculty of Medicine, Department of Radiology, Abant İzzet Baysal University, Bolu, Turkey, arzuolcun@gmail.com

Abstract—**Lung cancer is the most common type of cancer among various cancers with the highest mortality rate. The fact that nodules that form on the lungs are in different shapes such as round or spiral in some cases makes their detection difficult. Early diagnosis facilitates identification of treatment phases and increases success rates in treatment. In this study, a holistic Computer Aided Diagnosis (CAD) system has been developed by using Computed-Tomography (CT) images to ensure early diagnosis of lung cancer and differentiation between benign and malignant tumors. The designed CAD system provides segmentation of nodules on the lobes with neural networks model of Self-Organizing Maps (SOM) and ensures classification between benign and malignant nodules with the help of ANN (Artificial Neural Network). Performance values of 90.63% accuracy, 92.30% sensitivity and 89.47% specificity were acquired in the CAD system which utilized a total of 128 CT images obtained from 47 patients.**

Keywords-lung cancer, lung nodule, CAD, CT images, ANN classification

I. INTRODUCTION

Nowadays, lung cancer is one of the most deadly types of cancer. [1]. Various treatment options are used for lung cancer patients such as surgery, radiotherapy and chemotherapy. Despite these methods, 5 year survival rate for lung cancer patients is as low as 14 %. However, as in other cancer cases, survival rate may go up to 49 % if identified at an early stage [2].

Computerized tomography (CT) is the most frequently used imaging technique in the diagnosis of lung cancer [3]. Nodules and pathological residues with varied diameter can be comfortably viewed by CT [3]. Nodules on the lung are classified as benign or malignant. During diagnosis, malignant nodules that are solid and atypical can be assessed as benign in some cases. However, in most cases, a solid nodule is usually classified as malignant [4]. It is crucial to diagnose nodules at early stages in order to accelerate the treatment process.

CAD systems designed for the medical application provide various benefits for successful detection of pulmonary nodules. It is possible to start treatment process early with the help of these systems and they facilitate decision making process of physicians. In the literature, there are some studies regarding early diagnosis of lung cancer and identification of nodules. Okumura et al. [5] detected lung cancer with filtering techniques by using X-Ray CT images. Campadelli et al. [6] used image processing techniques to segment the lung on X-ray images for nodule identification. Lee et al. [7] developed a new approach regarding automatic detection of benign nodules. They used genetic algorithm based template matching technique on CT images. Kanazawa et al. [8] suggested a fuzzy cluster based CAD system for the identification of pulmonary nodules. Biradar and Patil [9] designed a CAD system to detect benign lung nodules by using CT images. They used the extraction of regions of interest and basic image processing techniques. Choi and Choi [10] proposed a CAD system to automatically classify lung nodules. Furthermore, there are various ANN-based CAD in literature. Suzuki et al. [11] proposed a pattern-recognition technique based on ANN using low-dose CT images for reduction of false positives in computerized detection of lung nodules. In another paper, Coppini et al. [12] presented a neural-network-based system for the computer aided detection of lung nodules in chest radiogram. Kuruvilla and Gunavathi [13] described a computer-aided classification method in CT images of lungs developed using ANN. However, in these studies, the true positive and false positive rates are not enough to meet the requirements of clinical use. Moreover, since these studies don't focus on early-detection of lung nodule. They don't include any suggestion for the detection of small size nodules.

This study proposes an ANN based CAD system for automatic classification of benign/malign pulmonary nodules at early stages. In this paper, Self-Organizing Maps (SOM) [14] has been used for nodule segmentation to enable the smallest nodules in the lungs. GLCM (gray-level co-occurrence matrix) [15] method has been utilized for the feature extraction of benign or malignant nodules. ANN, which is an effective classification technique, has been employed for classification.

Rest of the paper is organized as follows: Section 2 provides details of the designed CAD system. Section 3 includes results of the experimental processes and analysis. Performance evaluation of the proposed CAD and Discussions are explained in the last section.

II. MATERIAL AND METHOD

A. CT Image Dataset

An image database was created for the designed CAD system by collecting a total of 128 CT images from 47 different patients. There are 128 benign/malignant nodule in dataset. Based on pathological results, 52 of these nodules were malignant and 76 were benign. Images in

978-1-4799-5934-1/14/$31.00 ©2014 IEEE

the database were obtained from 35 male and 12 female volunteer patients between ages of 30 and 79. The database includes in variety of nodules with different sizes from 4 to 58 mm. CT images were acquired from the CT scanners at Abant Izzet Baysal University Medical Faculty in DICOM format and were then saved as 2 dimensional jpeg format with 256x256 resolution. Distribution of nodules in dataset is shown in Table I.

Table I. DISTRIBUTION OF NODULES IN DATASET

Nodule size(mm)	**Percentage of data (%)**	**Number of nodule**
<5 mm	31.25	40
5-20 mm	56.25	72
>20 mm	12.50	16

50% of the data in the database were defined as training cluster whereas the other 50% were defined as the test cluster.

B. Designed CAD System

The designed CAD system is composed of four main phases: (i) image pre-processing and selection of lung lobes, (ii) segmentation of the region of interest (ROI), (iii) feature extraction and feature selection and (iv) classification of benign and malign nodules. Block design of the CAD system is presented in Figure 1.

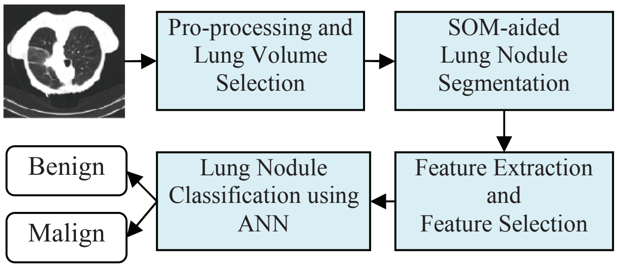

Figure 1. Architecture of the designed CAD system to detect and classify pulmonary nodules

Pro-processing and Lung Volume Selection

Image pre-processing is performed to enhance image quality and remove noise in the first step of the designed CAD system. 3x3 median filter was applied to remove noise and enhance the images. In this manner, regions with nodules and other regions become more distinct on CT images and noises are removed. Histogram equalization was used to balance the distribution of pixel value on the images. In lobe stripping process, lung lobes were extracted from among pre-processed CT images with the help of morphological operations. After this process, the remaining piece on the sides and edges were removed with Double thresholding method. So, lung region was successfully acquired.

SOM-aided Lung Nodule Segmentation

The various segmentation methods such as FCM, K-means, Otsu, Watershed, Region Growing and Graph Cuts can be used for the nodule segmentation [16]. In this study, SOM (Self-Organizing Maps) method was used to segment lung nodules [14] since it can organize large quantities of complex data sets and can design data maps that are easily interpreted. Based on these advantages, SOM enables easy segmentation of even the smallest nodules in the lungs.

Feature Extraction and Feature Selection

In the feature extraction and selection step, the features of Region of Interest were extracted to differentiate benign/malign tumors in lung CT images. The differentiation of tumors can be performed by the help of statistical and shape features of tumors. For example, malign nodules tend to be more complex and irregular whereas benign tend to be rounder with well-defined borders. The malign nodules, however, showed relatively higher variance values, indicating irregular shapes as shown Figure 2.

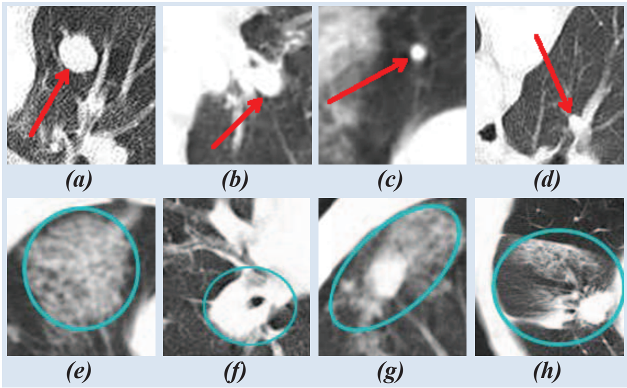

Figure 2. Examples of pulmonary nodule: (a, b, c, d) benign lung nodule, (e, f, g, h) malign lung nodule

Since GLCM (gray-level co-occurrence matrix) [14,15] is a statistical based texture feature extraction method, it is very useful for the classification of the benign or malignant tumor. So, in this study, GLCM method was used to extract the lung nodule features. The features extracted by GLCM are following: *(1)Angular Second Moment, (2)Entropy, (3)Dissimilarity, (4)Contrast, (5)Inverse Difference, (6)Correlation, (7)Homogeneity, (8)Autocorrelation, (9)Cluster Shade, (10)Cluster Prominence, (11)Maximum probability, (12)Sum of Squares, (13)Sum Average, (14)Sum Variance, (15)Sum Entropy, (16)Difference Variance, (17)Difference Entropy, (18)Information measures of correlation1, (19)Information measures of correlation2, (20)Maximal correlation co-efficient, (21)Inverse difference normalized, (22)Inverse difference moment normalized.* A total of 88 features were extracted with the help of GLCM from 0^0,45^0,90^0 and 135^0 angle directions in d=2 distance

After feature extraction process, the most appropriate features were selected through the Principal Component Analysis (PCA). PCA is a statistical based feature reduction method used to reduce dimensionality of complex data entries composed of large pieces of information [18,19]. The most appropriate 6 features, which provide best performance according to the experiments, were selected between the 88 features by PCA.

Lung Nodule Classification using ANN

In the classification step of proposed CAD systems, the tumors are classified such as benign or malignant. ANN is one of the artificial intelligence approaches that aim to generate a new system inspired by the operation of the human brain [20,21]. It is one of the most preferred methods in classification problems. So, in this paper, ANN was used for the classification of malign and benign lung nodules.

Multi-layer feed-forward perceptron model was used in ANN. Back-propagation algorithm was also utilized to train the network. Levenberg-Marquardt method was used as learning method. In addition, performance of network was calculated according to mean square error (MSE) rule.

There are three layers in the developed model: input, hidden and output. Only one hidden layer with 22 neurons was utilized in the study. This number of neuron has been decided in this way because performance of network has been obtained for best value. Input layer is composed of 6 digital inputs obtained from lung CT images through feature extraction. Output layer is composed of two outputs called the benign and malignant.

Figure 3 shows the architecture of developed ANN model.

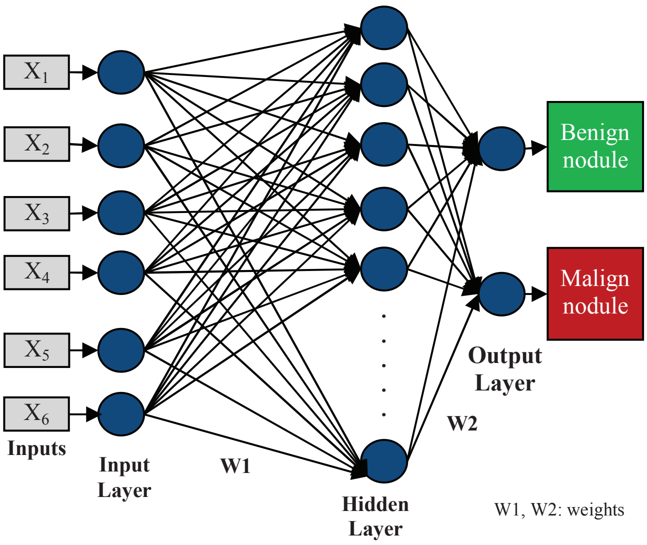

Figure 3. ANN architecture of proposed CAD system

III. EXPERIMENTAL RESULTS

The performance evaluation of the proposed CAD system was performed with MATLAB software. All experiments were performed by using a PC with 3.4 GHz i7 processor, 8 GB memory and Windows 7 operating system.

Figure 4 presents CT images for the outputs of procedures in the proposed system. Figure 4a displays the original lung image, Figure 4b is the image following image enhancement and image pre-processing steps and Figure 4c shows the lung stripping by using morphological operations. Figure 4d presents the lung nodules segmented with SOM method. Thus, Region of Interest (ROI) to be used in classification was obtained.

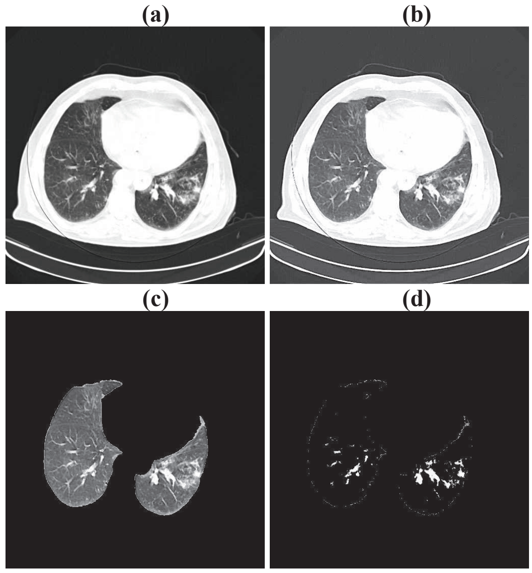

Figure 4. Segmentation of lung nodules, (a) original raw and unprocessed image (b) pre-processed image (c) stripped lung image with side remains removed (d) segmented image of the lung nodule

Following the identification of nodules on the CT image with the help of SOM method as shown in Figure 3d, feature extraction of ROI was performed. GLCM, a feature extraction method was used in this study. A total of 88 features were selected with this method from gray level image textures. These features were reduced with PCA. ANN was first trained by training data and then tested with test data. Training cluster consisted of 6 selected features and one class element (benign or malignant). Test cluster included only the features to be used in classification. At the end of the test phase, a confusion matrix was obtained through comparison of actual and predicted cases. Confusion matrix shows the true and false rates between the actual cases and predicted cases. Table II presents the confusion matrix for obtained results.

TABLE II. CONFUSION MATRIX OF OBTAINED RESULTS FROM TEST SET

CAD System	Positive	Negative
Positive (malign)	24	2
Negative(benign)	4	34

As shown Table 1, 34 of the 38 benign lung nodules were identified as benign (TN), only 4 nodules were misclassified as malignant (FN). On the other hand, 24 of the 26 malignant lung nodules were identified as malignant (TP) and only 2 nodules were misclassified as benign (FN). Table III shows the performance results based on the accuracy, sensitivity and specificity criteria of the designed CAD system.

TABLE III. EVALUATION OF PERFORMANCE MEASUREMENT CRITERIA

Performance criteria	Result
Accuracy	90.63
Sensitivity	92.30
Specifity	89.47

Figure 5 presents the ROC curve of the system obtained for classification accuracy in the designed CAD system. ROC curve is a preferred method to identify accuracy of diagnosis tests and to undertake safe comparisons. Large areas under the ROC curve point to high test performance [22]. Examination of the area under the ROC curve shows high system performance.

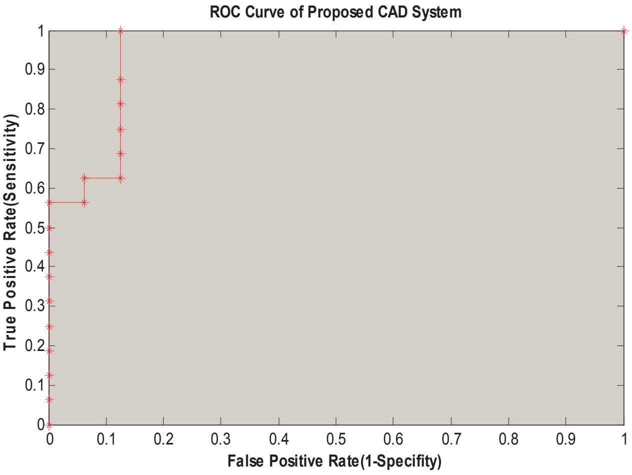

Figure 5. ROC graphic for the classification accuracy in CAD system

IV. DISCUSSION AND CONCLUSIONS

This study proposes an automatic CAD system that successfully differentiates the lung nodules as benign or malignant on CT images. The proposed CAD system is an integrated structure since it includes pre-processing, segmentation, feature extraction, feature selection and classification steps. SOM method included in CAD system allows successful detection of lung nodules in early stages. ANN was preferred in this study based on high accuracy rates (90.63 % accuracy, 92.30 % sensitivity and 89.47 % specifity) in classification.

Table IV shows the comparison of the proposed CAD system with state of the art CAD systems. The different CAD systems have shown reasonable sensitivity values in lung nodule detection. As a result, our proposed method shows significantly high sensitivity with very large CT image database.

Table IV. PERFORMANCE COMPARISON OF THE CAD SYSTEMS

CAD system	**Num. of case**	**Sensitivity (%)**
Opfer and Wiemker [24]	93	74.0
Rubin et al.[23]	20	76.0
Park et al.[27]	38	80.0
Messay et al.[25]	84	82.66
Suzuki et al. [11]	101	80.3
Dehmenski et al. [26]	70	90.0
Proposed method	**128**	**92.30**

One of the important contributions of this study is the early detection of lung cancer by identifying small sized lung nodules with the help of SOM method during segmentation. In our dataset, 31% of nodules are so small (<5mm), and 56% of nodules are medium size. So, the proposed CAD can detect the lung nodules at the early stage.

Comparison of this study with the literature shows that the proposed CAD system is more successful in the classification of malignant/benign nodules in terms of sensitivity and specificity criteria.

ACKNOWLEDGMENT

The authors would like to thank authorities of Medical Faculty of Abant Izzet Baysal University due to providing lung CT images. This work was funded by Sakarya University BAPK (No: 2014-50-02-015)

REFERENCES

[1] American Cancer Society, ACS cancer facts and figures 2002, American Cancer Society, Atlanta, GA, 2003.

[2] L. Ries et al., SEER Cancer Statistics Review 1973-1996, National Cancer Institution, Bethesda, MD, 1999.

[3] M. Dolejsi, Detection of Pulmonary Nodules from CT Scans, Czech Technical University, Faculty of Electrical Engineering, Center of Machine Perception, Prag, 2007.

[4] The international early lung cancer action program investigators, Survival of patients with stage I lung cancer detected on CT screening, N Engl J Med., 355, pp. 1763-1771, 2006.

[5] T. Okumura, T. Miwa, J. Kato, S. Yamamoto, M. Matsumoto, Y. Tateno, T. Iinuma ve T. Matsumoto, Variable N-Quoit filter applied for automatic detection of lung cancer by X-ray CT, Proc. CAR'98,, Tokyo, Japan, 1998.

[6] P. Campadelli, E. Casiraghi, S. Columbano, Lung Segmentation and Nodule Detection in Postero-Anterior Chest Radiographs, 2004.

[7] Y. Lee, T. Hara, H. Fujita, S. Itoh, T. Ishigaki, "Automated Detection of Pulmonary Nodules in Helical CT Images Based on an Improved Template-Matching Technique, IEEE Transactions on Medical Imaging, 20(7), July 2001.

[8] K. Kanazawa, Y. Kawata, N. Niki, H. Satoh, H. Ohmatsu, R. Kakinuma, M. Kaneko, N. Moriyama ve K. Eguchi, Computer-aided diagnosis for pulmonary nodules based on helical CT images, Comput. Med. Imag. Graph., 22(2), pp. 157-167, 1998.

[9] V. Biradar, U. Patil, Computer Aided Detection (CAD) System for Automatic Pulmonary Nodule Detection in Lungs in CT Scans, The International Journal of Engineering and Science (IJES), 2(1),pp. 18-21, 2013.

[10] W.-J. Choi ve T.-S. Choi, Automated Pulmonary Nodule Detection System in Computed Tomography Images: A Hierarchical Block Classification Approach, entropy, 15, pp. 507-523, 2013.

[11] K. Suzuki, S. G. Armato, F. Li, S. Sone, K. Doi, Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography, Med. Phys., 30, 1602, 2003.

[12] G. Coppini, S. Diciotti, M. Falchini, N. Villari, G. Valli, Neural Networks for Computer-Aided Diagnosis: Detection of Lung Nodules in Chest Radiograms, IEEE Transactions on Information Technology in Biomedicine, vol. 7(4), pp. 344-357, 2003.

[13] J. Kuruvilla, K. Gunavathi, Lung cancer classification using neural networksfor CT images, Computer Methods and Programs in Biomedicine, 113, pp. 202–209, 2014.

[14] T. Kohonen, Self-Organizing Maps, 3rd Edition, Springer, 2001.

[15] R. M. Haralick, K. Shanmugam ve I. Dinstein, Texture features for image classification, IEEE Trans. Syst.Man Cybern, 3(6), pp. 610-621, 1973.

[16] Z. Ekşi, E. Dandıl, M. Çakıroğlu, Bilgisayar Destekli Kırık Kemik Tespiti, 20.IEEE Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SIU'12), Fethiye, Türkiye, 18-20 Nisan, 2012.

[17] D. A. Clausi, An analysis of co-occurrence texture statistics as a function of grey level quantization, Can. J. Remote Sensing, 28(1), pp. 45-62, 2002.

[18] H. Camdevyren, A. Kanik, S. Keskyn, Use of principal components cores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs, Ecological Modelling, 181, pp. 581-589, 2005.

[19] L. H. Chen, S. Chang, An adaptive learning algorithm for principal component analysis, IEEE Transactions on Neural Networks, 6(5), pp. 1255-1263, 1995.

[20] K. K. Çevik, E. Dandıl, Development of visual educational software for artificial neural networks on .Net Platform, International Journal of Informatics Technologies, 5(1), pp. 19-28, 2012.

[21] H. Karamanli, N. Allahverdi, Design of a hybrid system for the diabetes and heart diseases, Expert System and Applications, 35, pp. 82-89, 2008.

[22] E. A. Kanık, S. Erden, Tanı Testlerinin değerlendirilmesinde ROC (Receive Operating Characteristics) Eğrisinin Kullanımı, Mersin Üniversitesi Tıp Fakültesi Dergisi, vol. 3, pp. 260-264, 2003.

[23] G. Rubin, J. Lyo, D. Paik, A. Sherbondy et. Al, Pulmonary nodules on multi-detector row CT scans: performance comparison of radiologists and computer-aided detection, Radiology 234: 274, 2005.

[24] R. Opfer, R. Wiemker, Performance Analysis for Computer-Aided Lung Nodule Detection on LIDC Data, In Proceedings of SPIE Medical Imaging, San Diego, CA, USA,Volume 6515, p. 65151C, 2007.

[25] T. Messay, R. Hardie, S. Rogers, A new computationally efficient CAD system for pulmonary nodule detection in CT imagery, Med. Image Anal, 14, pp. 390–406, 2010.

[26] J. Dehmeshki, X. Ye, X. Lin, M. Valdivieso, H. Amin, Automated detection of lung nodules in CT images using shape-based genetic algorithm, Comput. Med. Imaging Gr, 31, pp. 408-417, 2007.

[27] S.C. Park, J. Tan, X. Wang, D. Lederman, J.K. Leader, S.H. Kim, B. Zheng, Computer-aided detection of early interstitial lung diseases using low-dose CT images, Phys. Med. Biol., 56, pp. 1139-1153. 2011.