A Novel Approach to Malignant-Benign Classification of Pulmonary Nodules by Using Ensemble Learning Classifiers

A. Tartar, A. Akan, *Senior Member, IEEE*, N. Kilic

Abstract— **Computer-aided detection systems can help radiologists to detect pulmonary nodules at an early stage. In this paper, a novel Computer-Aided Diagnosis system (CAD) is proposed for the classification of pulmonary nodules as malignant and benign. The proposed CAD system using ensemble learning classifiers, provides an important support to radiologists at the diagnosis process of the disease, achieves high classification performance. The proposed approach with bagging classifier results in 94.7 %, 90.0 % and 77.8 % classification sensitivities for benign, malignant and undetermined classes (89.5 % accuracy), respectively.**

I. Introduction

Lung cancer is considered as one of the main causes of cancer deaths worldwide [1, 2]. In recent years, the cases of lung cancer have increased rapidly in the developed industrial countries such as United State of America, Japan and China [3, 4]. Today, the early detection of cancer is still not achieved at the desired level. Indeed, there is no change in the rate of cancer survivors in the past twenty years [2]. With improving technology, the development of the systems which may help physicians is extremely important at the diagnosis and treatment of cancer.

Early detection of potentially pulmonary nodules increases undoubtedly the patient's chance of survival in lung cancer. Today, the early detection of the disease is provided by using advanced medical imaging technologies such as computed tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography. Especially by using computed tomography systems, the monitoring of pulmonary nodules which have small size patterns is an important contribution to the early diagnosis. However, the use of advanced imaging systems causes many complexities. For each patient, the evaluation of CT scans containing numerous sectional images takes quite a long time and causes a tiresome process for radiologists. In this regard, the reduction of potential human error in the diagnosis step, and the development of computer-aided detection/diagnosis systems which make quick decisions to diagnosis and treatment of patient are highly important. Thus, computer-aided detection/diagnosis systems are a compact tool providing radiologists with a second opinion to improve the sensitivity of their diagnosis decision-making process [5].

In previous studies, proposed CAD systems have been focused on detection/classification of pulmonary nodules as nodule or non-nodule patterns [4, 6, 7]. In order to classify pulmonary nodule patterns, artificial neural network (ANN) approaches [8], random forest (RF) classifier [9], support vector machines (SVM) and ensemble learning classifiers such as bagging and adaboost algorithms are utilized. In this paper, different from above studies we also aim at classifying pulmonary nodule patterns as malignant or benign.

The remainder of this paper is organized as follows. The proposed CAD system and its sub-blocks are described in Section 2. This section gives a background on our previous work, patient database/imaging protocol information, feature sets and classifier algorithms of ensemble learning as well. Overall performance results are presented in Section 3. Conclusions are drawn in Section 4.

II. Material and Methods

A. Previous Work

In our previous study, a new classification approach for pulmonary nodules from CT imagery is presented by using hybrid features [19, 20]. In the study, the overall detection performance is evaluated using various classifiers such as artificial neural network, random forest, bagging and adaboost algorithms. The results are compared to similar techniques in the literature by using standard measures. The proposed approach with the hybrid features results in 90.7 % classification accuracy, 89.6 % sensitivity and 87.5 % specificity values for ANN classifier [19].

B. Patient Dataset and Imaging Protocol

In this section, a brief description of the patient dataset used is given. The proposed study was performed by using the CT images of 35 different patients diagnosed with lung cancer between 2010-2012 years at department of radiology, Cerrahpasa Faculty of Medicine, University of Istanbul. In lung cancer, the malignant-benign differentiation of pulmonary nodules requires a difficult clinical process. This distinction/classification can be made by radiologists as a result of monitoring of the disease for a long time and the evaluation of both pathology and radiology patient reports. In the study, the gender and pattern distributions according to the diagnosis type of patient used in dataset are given in

*Research supported by Scientific Research Projects Coordination Unit of Istanbul University.

A. Tartar is with the Department of Engineering Sciences, Istanbul University, 34320 Avcilar Istanbul, Turkey (phone: +90(212)4737070; fax: +90(212)4737180; e-mail: atartar@istanbul.edu.tr).

N. Kılıç is with the Department of Electrical and Electronics Engineering, University of Istanbul, Avcilar Campus, 34320 Istanbul, Turkey (e-mail: niyazik@istanbul.edu.tr).

A. Akan is with the Department of Electrical and Electronics Engineering, Istanbul University, 34320 Avcilar Istanbul, Turkey (e-mail: akan@istanbul.edu.tr).

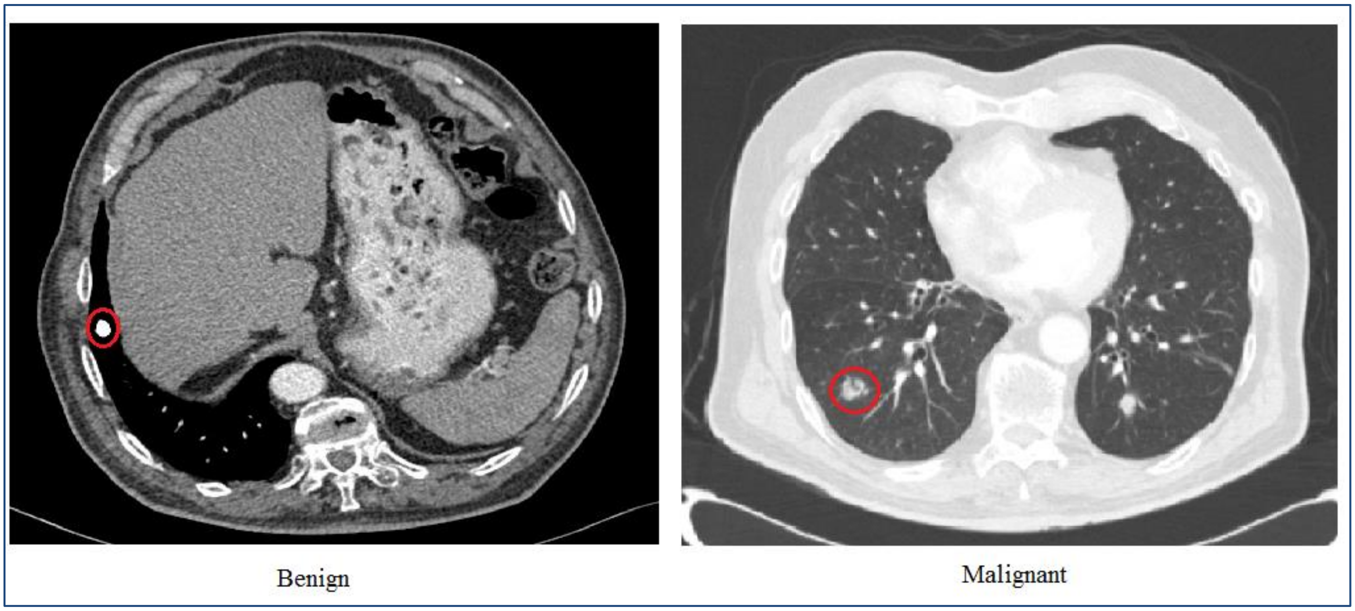

Figure 1. The malignant and benign pattern samples of pulmonary nodules.

Table 1. The malignant and benign pattern samples of pulmonary nodules are given in Fig.1 as well. Dataset consists of a total of 38 pulmonary nodule patterns collected from 24 male and 11 female patients whose ages ranging from 19 to 75 years. The mean age of the patients is 56.2±15.1 years. The number of pulmonary nodule patterns detected in the right and left lung parenchyma, as illustrated in Fig.2, is a total of 30 (13 in the upper part, 9 on the bottom part, 8 pleural case) and a total of 8 (4 in the upper part, 3 on the bottom part, 1 pleural case), respectively.

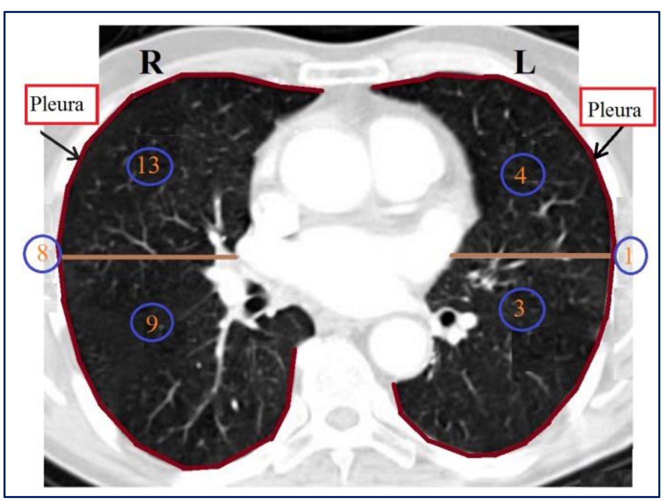

Figure 2. The regional distributions of pulmonary nodules.

The chest CT images of patients used in the study were obtained by using "Sensation 16" CT scanner (*Siemens Medical Systems*) at Department of Radiology, Cerrahpasa School of Medicine, University of Istanbul. All CT scans were acquired at a tube potential voltage of 120 kVp by utilizing the international imaging protocols. CT images are in size of 512x512 pixels and stored as DICOM (*Digital Imaging and Communications in Medicine*) format files, directly from the CT modality. The block diagram of the proposed computer-aided diagnosis (CAD) system for the malignant-benign classification of pulmonary nodules is shown in Fig.3.

Table 1. Dataset information and the number of patterns.

Diagnosis	The number of Patterns	Male	Female
Benign	19	13 Patient (15 Patterns)	4 Patient (4 Patterns)
Malignant	10	6 Patient (7 Patterns)	3 Patient (3 Patterns)
Undetermined	9	5 Patient (5 Patterns)	4 Patient (4 Patterns)

C. Feature Extraction

In order to classify a two-dimensional (2D) pattern by using image processing algorithms, the features are obtained for each 2D pattern. In this study, the morphological features are computed by the regional descriptors of the 2D pulmonary nodule patterns based on the basic morphological shape information. In addition, the demographic information of patient (age, gender) and nodule diameter, contrast, localization values computed from the CT images of the patients are identified as risk factors. The proposed feature set in this study for malignant-benign classification of pulmonary nodule patterns is given in Table 2. *P, A* and *L* features denote the perimeter, area and maximum diameter of the 2D pattern, respectively. Thus, a total of 10 features are evaluated for extracting features of the 2D pulmonary nodule patterns.

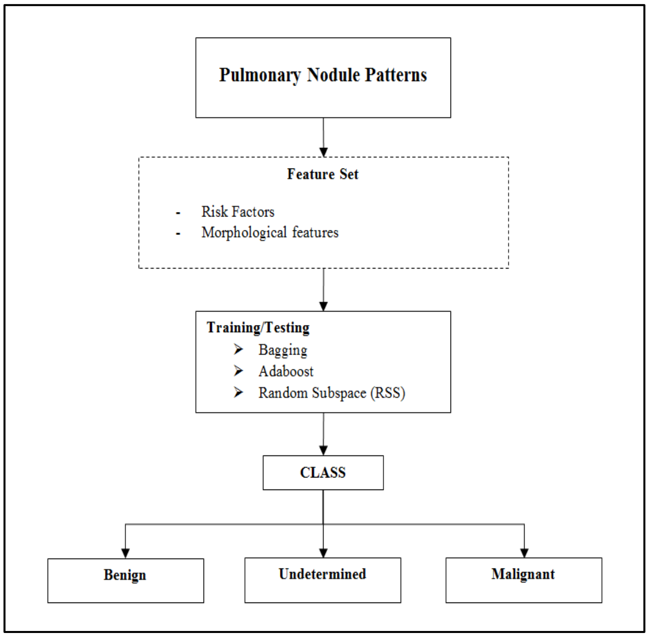

Figure 3. Block diagram of the proposed CAD system for the malignant-benign classification.

Table 2. Feature set extracted for malignant-benign classification.

Order	Quantity	Feature	Definition
1	Risk Factors	Age	-
2		Gender	-
3		Nodule Localization	-
4		Contrast	-
5		Nodule Diameter	-
6	Morphological Features	Eccentricity	-
7		Compactness	$P^2/4\pi A$
8		Roundness	$4A/\pi L^2$
9		Circularity	$4\pi A/P^2$
10		Ellipticity	$\pi L^2/2A$

D. Ensemble Learning Classifiers

1) Bagging: Ensemble learning classifiers have recently gained a very broad attention in terms of the minimization of test error of single classifiers and presenting faster classification algorithms [10, 12]. Bagging is an unstable learning algorithm for small data set if small changes in the training data generate very diverse classifiers. The use of bagging to improve the performance by taking advantage of this effect was proposed by Breiman [13]. A single classifier could have a higher test error. The combination of classifiers may produce a lower test error than that of the single classifier because the diversity of classifiers usually compensates for errors of any single classifier [11]. A learning algorithm combination in those small changes in the training set leads to relatively large changes in accuracy.

2) Adaboost: Adaboost is one of the powerful methods for pattern recognition [14]. Adaboost classifier first introduced by Freud and Schapire [14, 15] is an ensemble classifier composed of many weak classifiers for the two-class classification problem. It generates a strong classifier with weak classifiers. Adaboost makes a committee of member weak classifiers by adaptively adjusting the weights at each loop. While the weights of the training patterns classified correctly by a weak classifier are decreased, the weights of the training patterns misclassified by the weak classifier are increased.

Adaboost algorithm shows good performance effect because of the ability to generate expanding diversity. In order to improve the performance of the final ensemble, adaboost algorithms consist of diverse weak classifiers. Especially, the boosting algorithm Adaboost.M1 –the first directly- extends the original Adaboost algorithm to the multiclass case without reducing it to multiple two-class problems [15].

3) Random SubSpace (RSS): Random subspace method is composed of several classifiers [16]. RSS utilizes a generalized algorithm of random forest classifier. While random forest algorithm is generated from decision trees, RSS may be consisted from any classifier according to the structure of problem. RSS may also be applied in classification problems with single class [17].

In the study, random forest classifier is used as core classifier in the ensemble learning algorithms [18]. The features of the 2D pulmonary nodule patterns are obtained by using Matlab (*The Mathworks, Inc.*). All classification processes are provided by using data mining software called the Weka tool version 3.7.10 which is available from http://www.cs.waikato.ac.nz/~ml/weka. Computer simulations are performed on a PC with Intel Core i7, 1.90 GHz CPU and 4.00 GB RAM.

III. Results

Image processing algorithms are used in the classification of medical images. Recently, different CAD systems were proposed in the literature for detecting pulmonary nodules from CT images [4-11, 19-20].

The performance of the proposed system is tested with bagging, adaboost and random subspace classifiers by using the feature set extracted for 2D pulmonary nodule patterns in the dataset. All performance results of the proposed CAD system approach are given in Table 3 for each classifier. Confusion matrix, sensitivity, accuracy, kappa statistics, RMSE and AUROC metrics are measured.

$$Sensitivity = \frac{TP}{TP+FN} \quad (1)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum(y'-y)^2}{n}} \quad (3)$$

Here *TP, TN, FP* and *FN* denote the number of pulmonary nodule patterns classified as true positive, true negative, false positive, and false negative, respectively. In the root mean squared error (*RMSE*), *y* and *y'* depict actual and predicted values. *n* is the number of pulmonary nodule patterns. Sensitivity denotes the number of correctly predicted positives divided by the total number of positive cases.

Table 3. Overall performance results of the proposed system.

	Actual	Predict		
BAGGING		*Benign*	*Malignant*	*Undetermined*
	Benign	18	1	0
	Malignant	0	9	1
	Undetermined	1	1	7
	Sensitivity	**94.7 %**	**90.0 %**	**77.8 %**
	AUROC	0.968	0.871	0.931
	KAPPA	**0.831**		
	RMSE	**0.338**		
	Accuracy	**89.5 %**		
	Actual	**Predict**		
ADABOOST		*Benign*	*Malignant*	*Undetermined*
	Benign	18	1	0
	Malignant	0	8	2
	Undetermined	0	3	6
	Sensitivity	94.7 %	80.0 %	66.7 %
	AUROC	0.970	0.884	0.939
	KAPPA	0.750		
	RMSE	0.313		
	Accuracy	84.2 %		
	Actual	**Predict**		
RSS		*Benign*	*Malignant*	*Undetermined*
	Benign	17	2	0
	Malignant	0	9	1
	Undetermined	2	2	5
	Sensitivity	89.5 %	90.0 %	55.6 %
	AUROC	0.947	0.846	0.828
	KAPPA	0.704		
	RMSE	0.364		
	Accuracy	81.6 %		

In Table 3, AUROC depicts the area under the receiver operating characteristic curve. Kappa statistics is a chance-corrected measure of agreement between the classifications and the true classes. It is in the range of [1, -1]. If Kappa is equal to 1, it indicates perfect agreement. If Kappa is equal to 0, it represents chance agreement. In the study, the performance measurements are obtained by using 5-fold cross validation in the classifier algorithms. When evaluating the measurements, only taking into account the accuracy or sensitivity metrics cannot give a reliable result [20]. In addition to these, kappa statistics and error values should also be taken into consideration.

IV. CONCLUSION

This study, which is the first in the literature, provides a classification of pulmonary nodule patterns as malignant and benign. In this context, this classification effort is a valuable task in the clinical diagnosis. In our approach, we include a new class called "*undetermined*" indicating that the nodule should be further monitored, as well as malignant and benign classes in the literature.

Considering the results in Table 3, ensemble learning algorithms yield acceptable performance in the malignant-benign classification of pulmonary nodules. In our tests, the use of bagging classifier appears to be more appropriate in the classification.

In conclusion, our proposed CAD system is a highly promising method, providing high performance in the malignant-benign classification of pulmonary nodules which may be used in the diagnosis of lung cancer.

ACKNOWLEDGMENT

This work was supported by Scientific Research Projects Coordination Unit of Istanbul University, Project Numbers: 44942, 38262.

REFERENCES

[1] Cancer Facts And Figures, "American Cancer Society", Atlanta. 2010.

[2] Health Statistics Annual, Ministry of Health, Republic of Turkey, Ankara, 2009.

[3] Kaizoe, T., "Figures on Cancer in Japan, Foundation for Promotion of Cancer Research", 1995.

[4] T. W. Way, B. Sahiner, H-P Chan, L. Hadjiiski, P.N. Cascade, A. Chughtai, N. Bogot, E. Kazerooni, "Computer-aided diagnosis of pulmonary nodules on CT scans: Improvement of classification performance with nodule surface features", *Medical Physics,* 36(7), 3086-3098, 2009.

[5] Doi K., "Computer-aided diagnosis in medical imaging: historical review, current status and future potential", *Computerized Medical Imaging and Graphics*, 31(4–5), 198–211, 2007.

[6] Cascio D., Magro R., Fauci F., Iacomi M., Raso G., "Automatic detection of lung nodules in CT datasets based on stable 3D mass-spring models", *Computers in Biology and Medicine*, 42, 1098-1109, 2012.

[7] Choi W-J., Choi T-S., "Automated pulmonary nodule detection based on three-dimensional shape-based feature descriptor", Computer Methods and Programs in Biomedicine, 113, 37-54, 2014.

[8] Retico, A., Delogu, P., Fantacci, M.E., Gori, I., Martinez, A.P., "Lung nodule detection in low-dose and thin-slice computed tomography", *Computers in Biology and Medicine*, 38, 525-534, 2008.

[9] Lee S.L.A., Kouzani A.Z., Hu E.J., "Random forest based lung nodule classification aided by clustering", *Computerized Medical Imaging and Graphics*, 34, 535-542, 2010.

[10] Tartar, A., Kılıç, N., Akan, A., "Bagging support vector machine approaches for pulmonary nodule detection", *IEEE International Conference on Control, Decision and Information Technologies*, s.047-050, Tunisia, 2013.

[11] D. Opitz and R. Mach, "Popular Ensemble Methods: An Empirical Study", *Journal of Artificial Intelligence Research*, 11,169-198, 1999.

[12] Lee M.C., Boroczky L., Sungur S-K., Cann A.D., Borczuk A.C., Kawut S.M., Powell C.A., "Computer-aided diagnosis of pulmonary nodules using a two-step approach for feature selection and classifier ensemble construction", *Artificial Intelligence in Medicine*, 50, 43-53, 2010.

[13] L. Breiman, "Bagging predictors," Mach. Learn., vol. 24, pp. 123–140, 1996.

[14] Freund, Y., Schapire, R. E., "Experiments with a new boosting algorithm", *Proceedings of the 13th Conference on Machine Learning*, 148-156, 1996.

[15] Freund, Y., Schapire, R. E., "A decision-theoretic generalization of on-line learning and an application to boosting", *Journal of computer and system science*, 55, 119-139, 1997.

[16] Bryll, R., "Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets", *Pattern Recognition*, 20(6), 1291–1302, 2003.

[17] Ho Tin, "The Random Subspace Method for Constructing Decision Forests", IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(8), 832–844, 1998.

[18] Breiman, L., "Random forests", *Machine Learning*, 45, 5-32, 2001.

[19] Tartar A., Kılıc N., Akan A., "Classification of pulmonary nodules by using hybrid features", *Computational and Mathematical Methods in Medicine*, Volume 2013, s.1-11, doi:10.1155/2013/148363, 2013.

[20] Tartar A., Kılıc N., Akan A., "A new method for pulmonary nodule detection using decision tress", *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC13)*, s. 7355-7359, Osaka (Japan), 2013.