Determining Malignancy of Small Pulmonary Nodules by Nodule Characteristics

Aydın Kaya and Ahmet Burak Can

Hacettepe University, Computer Engineering Department,
Ankara, Turkey

Abstract— **Classification of small pulmonary nodules is an important task for lung cancer diagnosis. Studies on classification of these nodules generally concentrate on determining nodule malignancy using image features. In the recent years, publicly available databases offer researchers various types of data other than image features. LIDC database includes such information about radiologists' annotations on nodule characteristics. In this paper, a cascaded classification method is studied to classify malignancy of small pulmonary nodules using nodule characteristics and image features. Results are compared with single classifiers based on nodule characteristics and image features separately.**

Keywords— **Lung nodules, nodule characteristics, unbalanced dataset, cascaded classifier.**

I. INTRODUCTION

Detection and classification of small nodules are challenging processes for the diagnosis of lung cancer. Small size and unclear boundaries of nodules are the main reasons of this difficulty. For this reason, many studies are conducted on nodule segmentation and classification in the literature. While most studies focused on determining the degree of malignancy, some studies tried to classify nodule characteristics (radiographic descriptors) separately. Especially with the deployment of NELSON and Lung Image Database Consortium databases [1], the studies on nodule characteristics have increased in the recent years. Nodule characteristics are the features defined by radiologists to evaluate nodules based on nodule appearance and tissues around nodules. These characteristics are important to decide about nodule malignancy.

In this study, a classification approach is proposed to determine malignancy of nodules using nodule characteristics. A cascaded classifier is built using linear discriminant classifiers and support vector machines (SVM). First of all, nodule characteristics are determined using image features obtained from nodule images. Then, nodule characteristics and image features are used to classify nodule malignancy. Classification results are compared with the results of single classifiers trained on annotations of radiologists on nodule characteristics and image features separately. The cascaded classifier produces better results than single classifiers. The proposed method produces ratings for nodule characteristics in addition to malignancy rating. When classifying nodules, this extra information may help both radiologists and computer aided classification systems for a better understanding of analyzed nodules.

II. RELATED WORK

Most studies in the literature use image features to classify small pulmonary nodules. Xu et al. [3] classified nodules based on size, shape and margin characteristics that included by NELSON screening trial. Li et al. [4] used feature-based, a pixel-value-difference based, a cross-correlation-based and neural-network-based techniques for determination of similarity measure among radiologist's evaluation over nodule types. Samuel et al. [5] proposed a Mamdani-type fuzzy logic system for recognizing nodules based on wavelet transform, bi-histogram equalization, and morphological transform. Aoyama et al. [6] segmented nodules with dynamic programming and classified them by using linear discriminant analysis. Different combinations of features examined for classification performance. Lo et al. [7] extracted nodule's 3D area and used back-propagation neural network to classify nodules as benign and malignant. Way et al. [8] segmented nodules by three dimensional active contour method and classified using linear discriminant classifier. Kawata et al. [9] extracted features from nodule and surrounding structures and used stepwise linear discriminant classifier.

Some studies use image features to classify nodule characteristics. Zinovev et al. [10] proposed an ensemble classifier and active learning method to predict nodule characteristics. They stated that highly imbalanced data like LIDC can be expressed better by ensemble methods than single classifiers. Zinovev et al. [14] also proposed a system for predicting nodule characteristics by ensemble of probabilistic classifiers based on belief decision trees and ADABoost learning. Li et al. [15] proposed a method for predicting malignancy using four nodule characteristics and image based features by neural networks.

Giuca et al. [11] developed a content based image retrieval method to annotate large unlabeled data with small amount of labeled data. Kim et al. [12] proposed a semantic and content based image retrieval model to determine the relationship between nodule characteristics and image features. In this model, related images are found using linear regression and similarity measures.

I. Lacković and D. Vasić (eds.), *6th European Conference of the International Federation for Medical and Biological Engineering*,
IFMBE Proceedings 45, DOI: 10.1007/978-3-319-11128-5_37

III. Method

A. Dataset

National Cancer Institute has formed a demand in 2001 for a lung CT image database which can be accessed via the Internet under the title of "*Lung Image Database Resource for Imaging Research*" [1]. Data collection and evaluation phase has started after preparation steps like compliance check of CT images, definition of nodule evaluation criteria, choosing suitable database model for expansion, defining statistical framework for user guidance.

For the evaluation of the images, radiologists are assigned from four different institutions. Evaluation phase is divided into two steps "blinded" (first evaluation of case) and "unblinded" (evaluation after taking consideration of other readers) reads. After "blinded" reading session is finished, all information of different institutions are gathered and distributed again. In the unblinded evaluation, each radiologist takes into consideration the evaluation of other radiologists and then can edit his/her decisions. Final XML data contains only unblinded evaluation results. LIDC database does not only contain malignancy evaluation of nodules but evaluations of eight other nodule characteristics (radiographic descriptors). Short descriptions of these characteristics and their rating range are given in Table 1.

Table 1 Nodule characteristics and ratings [12]

Nodule Characteristic	Description	Rating
Calcification	Calcification appearance in the nodule.	1-6
Internal Structure	Expected internal composition of the nodule.	1-4
Lobulation	Whether lobular shape is apparent from margin or not.	1-5
Malignancy	Likelihood of malignancy.	1-5
Margin	How well defined the margins are.	1-5
Sphericity	Dimensional shape in terms of roundness.	1-5
Spiculation	Degree of exhibition of spicules.	1-5
Sublety	Contrast between nodule and surroundings.	1-5
Texture	Internal density of nodule.	1-5

LIDC database contains evaluations of radiologists from different institutions. Radiologists are not expected to agree on each nodule characteristics. This situation makes it difficult to determine the ground truth or golden ratio on nodule characteristics. Nevertheless, Zinovev et al. [14] stated, lack of ground truth and radiologist anonymity over dataset is a challenging situation; however it gives opportunity to develop different computer aided diagnosis methods.

In consideration of information from LIDC wiki page [16], if an agreement reached on a characteristic by at least three radiologist (3/3, 3/4, 4/4), this information is considered as the ground truth and added to datasets.

B. Features

155 image features are calculated for each nodule sample. There are shape, size, and texture-based features. Some features are extracted from the largest nodule slice, some from all nodule slices and nodules surrounding structures (obtained by dilating nodule area with a six pixel diameter disk structure element). High and low degree Zernike moments [13] are obtained from the largest nodule area. Eccentricity, solidity, circularity, aspect ratio, area of bounding box, standard deviation, and Haralick texture features [19] are extracted from the largest area of a nodule, average of all nodule slices, and nodule surrounding areas.

C. Preprocessing

Individual datasets are created for each nodule characteristic. Agreement of at least three radiologists is expected on ratings of each characteristic. This provides us separate datasets for each characteristic with a different amount of samples. Let S_{chr} be such a set where *chr* is a nodule characteristic. General description of LIDC nodule dataset S is given as in (1). In the equation, x_i is i^{th} image feature, m is the number of features, y_c is the rating for c characteristic, and n is the number of nodules.

$$S_{chr} \subset (S = \{x_1, \dots, x_m, y_1, \dots y_c\}^{1..n} \in \mathbb{R}^m) \qquad (1)$$

In datasets, some samples have missing values in consequence of small nodule size. We use k-nearest neighbor imputation [17] method for missing value completion. Different methods like expectation maximization, regression, single value decomposition, etc. can also be used for this problem [17].

Most of the nodule characteristic datasets are highly unbalanced. For an example, in subtlety dataset, 71% of all nodules are marked 5, 19% are marked 4, 9% are marked 3, and remaining nodules are marked 1 or 2. Similar situation is also observed in the other nodule characteristic datasets. To get rid of inadequate expression of small sample class problem, we applied sample balancing methods.

If a rating value is under a certain ratio (10%) among all samples, data are oversampled with Synthetic Minority Over-sampling Technique (SMOTE) [18]. In brief, SMOTE method artificially generates synthetic samples that were not in the dataset, rather than by over-sampling with replacement. Depending on the amount of over-sampling required, minority class is over-sampled by taking each minority class sample of the k nearest neighbors.

If a rating is above a ratio (above thirty percent), data are under sampled by class-specific feature space convex hull points. Nodules that have feature values on convex hull are chosen for under sampled dataset, others are discarded.

For the malignancy dataset, the nodules that at least three radiologists agreed are selected. This dataset contains both image features and nodule characteristics for the selected nodules.

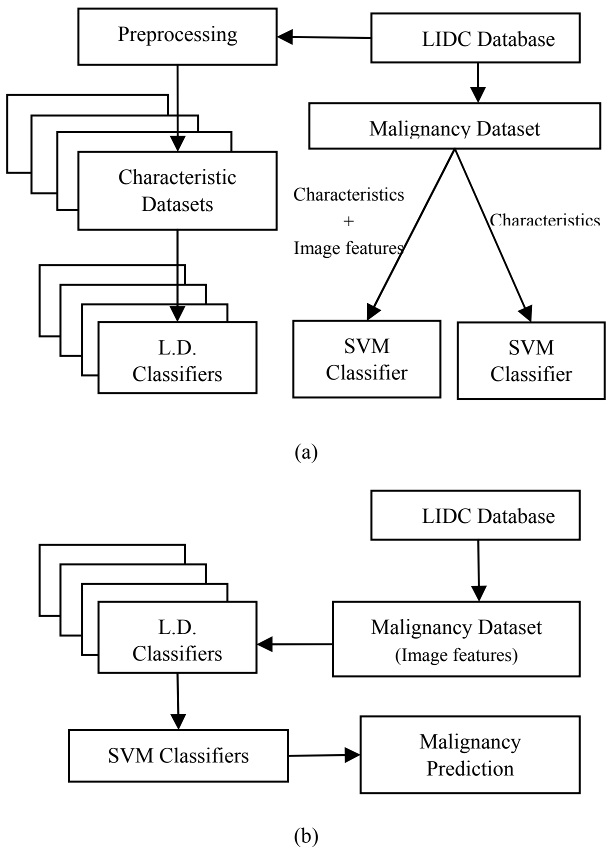

Fig. 1 General schema for proposed method. (a) Training phase (b) Testing phase.

D. Classification

General schema of the proposed method is given in Figure 1. We used separate linear discriminant classifiers for each nodule characteristic in the first step of the cascaded classifier. Dimension reduction with feature selection is applied to all characteristic datasets before constructing first-step classifiers. Relieff [2] method is used for feature selection. The most important six features are selected. Each classifier is trained with a different dataset using different feature sets. In our experiments, selecting six image features was optimal for acceptable computational complexity. Feature set size can be changed arbitrarily or according to performance requirements.

In the second classification step, two SVM classifiers are trained as shown in Figure 1(a). The first one is trained with characteristic ratings of nodules in the malignancy dataset. The second one is trained with image features (the most important six for malignancy dataset) and characteristic ratings.

In the testing step, leave one out generalization method is used for validation of the cascaded classifier. In each step of this testing method, a sample from the malignancy dataset is used as the testing data, and the remaining samples as the training data. In the first step of testing, a sample from malignancy dataset is sent to the L.D. classifiers. Then, SVM classifiers produce a malignancy rating using the classification results of the first step classifiers. This is repeated until each sample in the malignancy dataset is used once as a validation sample.

E. Results

Before testing the cascaded classifier, we trained single classifiers to predict malignancy using our image features. Classification accuracy (CA) and area under curve (AUC) results with Naive Bayes, Adaboost, kNN, SVM, and Random Forest (RF) classifiers are given in Table 2. The best classification result with image features is %79.89 for the RF classifier.

Table 2 CA and AUC results for single classifiers on image features.

Method	CA	AUC
N. Bayes	0.7220	0.9582
AdaBoost	0.7226	0.9138
kNN	0.7623	0.9412
SVM	0.7916	0.9509
RF	0.7989	0.9545

Table 3 CA and AUC results for single classifiers on nodule characteristics and our method.

Method	CA	AUC
N. Bayes	0.7904	0.9364
AdaBoost	0.7904	0.9029
kNN	0.7113	0.9014
SVM	0.7969	0.9368
RF	0.6273	0.8843
Our Method (w/o image features)	0.7791	0.9051
Our Method (with image features)	0.8159	0.9451

Then, experiments are made on the malignancy dataset to classify malignancy using radiologist's annotations on nodule characteristics, which is considered as the ground truth data. Ratings of radiologists for nodule characteristics are used as features to train classifiers. The trained models are tested on the malignancy dataset with leave-one-out cross validation. Results of single classifiers and our approach are given in Table 3. According to the results, highest classification accuracy obtained with single classifiers on radiologist opinions (ground truth) is 79.69%, which is a close result obtained with image features.

As shown in Table 3, classification accuracy of our cascaded classifier without image features is 77.91% and with image features is 81.59%. The cascaded classifier has better performance than the single classifiers which are trained with image features or characteristics. Thus, results showed that using nodule characteristics and image features together can be beneficial for predicting malignancy.

IV. CONCLUSION

In the literature, there are many studies on predicting nodule malignancy using low level image features. In this paper a cascaded classifier method is proposed to predict malignancy by using nodule characteristics. In the first step, different datasets are created and different linear discriminant classifiers are trained for each nodule characteristic. Results of the first step - ratings for nodule characteristics and image features- are classified with SVM classifiers in the second step to predict malignancy. Experiments showed that using nodule characteristics facilitates malignancy prediction and preliminary results are promising for the further development.

For the future work, we are going to expand our studies with ensemble classifiers and fuzzy inference methods. We plan to overcome the imbalanced data problem with these methods in addition to data balancing.

ACKNOWLEDGMENT

The authors would like to thank Dr. Murat Dündar at IUPUI for his valuable opinions and contributions.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

1. Armato III, S.G. et al. (2004). Lung image database consortium: developing a resource for the medical imaging research community. Radiology, 232, 739-748.
2. Sun, Y., Li, J. (2006). Iterative RELIEF for feature weighting. In Proceedings of the 23rd international conference on Machine learning (pp. 913-920). ACM.
3. Xu, D. M., et al. (2008). Limited value of shape, margin and CT density in the discrimination between benign and malignant screen detected solid pulmonary nodules of the NELSON trial. European journal of radiology, 68, 347–352.
4. Li, Q., Li, F., Shiraishi, J., Katsuragawa, S., Sone, S., Doi, K. (2003). Investigation of new psychophysical measures for evaluation of similar images on thoracic computed tomography for distinction between benign and malignant nodules. Medical physics, 30, 2584–2593.
5. Samuel, C. C., Saravanan, V., Vimala Devi, M. R. (2007). Lung nodule diagnosis from CT images using fuzzy logic. In Conference on Computational Intelligence and Multimedia Applications. International Conference on (Vol. 3, pp. 159-163). IEEE.
6. Aoyama, M., Li, Q., Katsuragawa, S., Li, F., Sone, S., Doi, K. (2003). Computerized scheme for determination of the likelihood measure of malignancy for pulmonary nodules on low-dose CT images. Medical physics, 30, 387–394.
7. Lo, S. B., (2003).Classification of lung nodules in diagnostic CT: an approach based on 3D vascular features, nodule density distribution, and shape features. Proceedings of SPIE 5032: 183-189.
8. Way, T. W., et al. (2006). Computer-aided diagnosis of pulmonary nodules on CT scans: segmentation and classification using 3D active contours. Medical physics, 33, 2323–2337.
9. Kawata, Y., et al. (2001). Computerized analysis of 3-D pulmonary nodule images in surrounding and internal structure feature spaces. Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205), 2.
10. Zinovev, D., Raicu, D., Furst, J., Armato III, S. G. (2009). Predicting Radiological Panel Opinions Using a Panel of Machine Learning Classifiers. Algorithms.
11. Giuca, A.M., Seitz, K.A., Furst, J., Raicu, D. (2012). Expanding diagnostically labeled datasets using content-based image retrieval. In 19th IEEE International Conference on Image Processing (ICIP) (pp. 2397–2400).
12. Kim, R., Dasovich, G., Bhaumik, R., Brock, R., Furst, J. D., Raicu, D. S. (2010). An investigation into the relationship between semantic and content based similarity using LIDC. Proceedings of the international conference on Multimedia information retrieval - MIR ’10, 185.
13. Zernike, F. (1934). Diffraction theory of the cut procedure and its improved form, the phase contrast method, Physica, 1:pp. 689-704.
14. Zinovev, D., Furst, J., & Raicu, D. (2011). Building an Ensemble of Probabilistic Classifiers for Lung Nodule Interpretation. 2011 10th International Conference on Machine Learning and Applications and Workshops, 2, 155–161.
15. Li, Guangxu, et al. (2013) Semantic characteristics prediction of pulmonary nodule using Artificial Neural Networks., Engineering in Medicine and Biology Society (EMBC), 35th Annual International Conference of the IEEE.
16. The Cancer Imaging Archive Wiki Site: https://wiki.cancerimagingarchive.net/display/Public/Wiki, last access: 1 April 2014.
17. Troyanskaya, O., et al., (2001). Missing value estimation methods for DNA microarrays. Bioinformatics 17(6), 520–525.
18. Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P. (2011). SMOTE: synthetic minority over-sampling technique. arXiv preprint arXiv:1106.1813.
19. Haralick, R.M., Shapiro, L.G. (1992) Computer and Robot Vision: Vol. 1, Addison-Wesley p. 459.