A Texture Feature Analysis for Diagnosis of Pulmonary Nodules Using LIDC-IDRI Database

Fangfang Han[1,2], Guopeng Zhang[3], Huafeng Wang[2], Bowen Song[2], Hongbing Lu[3], Dazhe Zhao[1], Hong Zhao[1], and Zhengrong Liang[2]

[1] Northeastern University, Shenyang, Liaoning 110819, China
E-mail: neuhanfangfang@gmail.com
[2] Dept. of Radiology, State University of New York at Stony Brook, Stony Brook, NY 11794, USA
E-mail: jerome.liang@sunysb.edu

[3] Dept. of BME, The Fourth Military Medical University, Xi'an, Shanxi 710032, China

Abstract: This paper evaluated the performance of two-dimensional (2D) and 3D texture features from CT images on pulmonary nodules diagnosis using the large database LIDC-IDRI. Total of 905 nodules (422 malignant and 483 benign) with certain expert observer ratings of malignancy were extracted from the database based on the radiologists' painting boundaries. Feature analysis on the extracted nodules was not only based on the popular texture analysis method, e.g., the 2D Haralick texture feature model, we also explored a 3D Haralick feature model with variable directions in space. The relationships of more neighbour voxels on more directions were included for texture feature analysis. The well-established Support Vector Machine (SVM) classifier was used for the malignancy classification based on the 2D and 3D Haralick texture features. Half of the benign and malignant nodules were extracted randomly for training, and the left half nodules for testing. This operation was implemented for 100 iterations. Then the 100 classification results were shown based on the area under the curve (AUC) of the Receiver Operating Characteristics (ROC). The distinguishing results on the nodule malignancy based on the 3D Haralick texture features (Az = 0.9441) is noticeably more consistent with the expert observer ratings than that on the 2D features (Az = 0.9372).

Key Words: 3D Haralick Texture Features, LIDC-IDRI Database, Pulmonary Nodules, Malignancy Diagnosis, SVM, ROC.

1. INTRODUCTION

In the past decades, because of the widely use of computer tomography (CT) for lung cancer screening [1], a large and increasing number of pulmonary nodules have been detected each year. The task of evaluating a large number of detected nodules by the experts or radiologists can be very burdensome. Therefore, computer-aided diagnosis (CADx) is expected to play an important role in the evaluating task, where the extraction of the features from the detected nodules and classification of the features are the basic research interests [2-4].

The computer-aided diagnosis of pulmonary nodules on CT images is expected to take all available patients information into account, such as sizes, shapes, volumetric growing rate, intensity, texture, etc. [5-7]. In these features, sizes, shapes and volumetric growing rate have been considered as most reliable features for malignant diagnosis of nodules. However, most of the above features need more accurate segmentation of nodules except texture. Therefore, we take more attention into the texture features for assisting the malignancy diagnosis of nodules in this paper. Two-dimensional (2D) Haralick model is a popular method for texture feature analysis. And we used it to evaluate the performance of texture feature. On the other hand, we improved it to a new 3D model to include more space information based on our prior work [8]. In this paper, we analyzed more types of direction combination in details.

To achieve more stable and confident results, we used the largest database built by the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI), which contains 1012 chest cases and 1356 nodules (the diameters are between 3 and 30 mm) at the time of this study. Although there is no golden standard of malignancy diagnosis in the database, it contains a large quantity of nodule cases with multiple types from many different institutions [9]. Nearly each nodule case contains the painting boundary coordinates and vision features from four radiologists. And the malignancy ratings were set by the visual decision of the four radiologists from one to five. The bigger value represents more malignant suspicious. The nodules with malignancy rating '3' were not contained in this study because they could not be distinguished. Therefore, 905 nodules (422 malignant and 483 benign) with certain expert observer ratings of malignancy were studied in our experiments.

The purpose of this paper is to evaluate the performance of texture features on lung CADx and the consistency with expert diagnosis based on visual features from CT images. In our future work, to address the lack of gold standard of LIDC-IDRI database, we will study the texture features on the nodule data with biopsy reports.

The remainder of this paper is organized as follows. In Section 2, the LIDC-IDRI database and the nodule extraction method are introduced. Then the theory of 2D

978-1-4673-5887-3/13/$31.00 ©2013 IEEE

Haralick features and the 3D feature model are described for details in Section 3. In Section 4, the experiments and results are shown. Finally, the discussion of the results and conclusions are given in Section 5.

2. LIDC-IDRI DATABASE AND NODULE EXTRACTION METHOD

In this section, the source of the nodule cases and the extraction method are described in details. For the analysis of nodule features, it is important to get the nodule area from chest CT images. To obtain more stable results, we selected the largest lung nodule database which contains more than one thousand nodule cases. In addition, each nodule in the database are attached by an annotation file which includes the boundary coordinates and some image features recognized by almost four radiologists. This provides a more convenient condition for malignancy diagnosis directly.

2.1. LIDC-IDRI Database

LIDC-IDRI database is a web-accessible international open resource for development, training, and evaluation of CAD methods for lung cancer detection and diagnosis [10]. The database contains 1012 chest CT image cases, each case includes images from a clinical thoracic CT scan and an associated XML file that records the locations of the image pixels on the nodule boundary in each image slice and nine characteristics of the nodule malignancy given by up to four experienced thoracic radiologists. The nine characters are malignancy, texture, spiculation, lobulation, margin, sphericity, calcification, internal structure, and subtlety. Each character was set an evaluation value by each radiologist after a visual diagnosis based on the information from the CT images. The slice thickness (collimation) of the CT scans varies between 1.25 – 3 mm, the slice spacing (reconstruction interval) varies between 0.625 – 3.0 mm, and the tube current between 40 – 500 mA [9].

Based on the above parameters, we selected 905 nodules which have certain malignancy ratings (‘1’, ‘2’, ‘4’, and ‘5’) by averaging the four radiologists’ decisions. If the average malignancy rating is ‘1’ or ‘2’, the nodule is seemed as benign, and the label is zero for classification. On the opposite, if the average malignancy rating is ‘4’ or ‘5’, the nodule is identified by malignant, and the label is one for classification.

2.2. Nodule Extraction Method

As described in the subsection 2.1, each nodule (which diameter is bigger than 3mm) in the LIDC-IDRI database has the boundary coordinates painted by the radiologists. Therefore, we made a combination boundary for the nodule by a intersection rule. The main process is to extract all image pixels inside the painting boundaries and select the intersection area of at least three radiologists’ decisions as the nodule’s region in an image slice. The algorithm of extracting the intersection region on a single 2D image slice is outlined as follows.

- Read all the pixels on the boundary of the nodule in a single image slice.
- Fill the inner pixels inside each of the radiologists’ painting boundaries.
- Calculate the times (t_n) of each pixel being contained by all the boundaries.
- Label the inner pixels with $t_n >= 3$.
- Combine the labeled pixels as the final region data in one 2D slice of the candidate nodule.

As an example, the combination processing of a single slice of one nodule is shown in Figure 1.

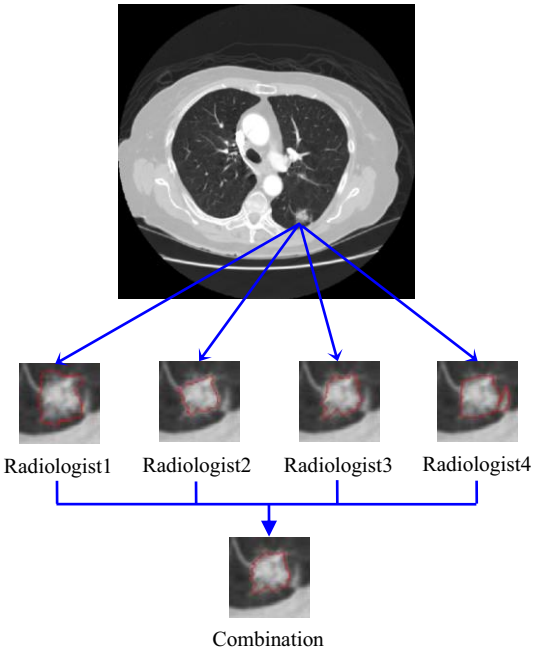

Fig.1. The combination processing of one nodule slice based on the paintings from four radiologists.

After the intersection regions on all of the slices were determined, the 3D volume data of the candidate nodule can be obtained by superimposing all the 2D combination areas.

3. 2D AND 3D HARALICK TEXTURE FEATURE MODELS

Haralick feature analysis algorithm was first proposed by Robert M. Haralick in 1973 [11]. It provided a calculation method of texture features in 2D images for classification. And this method has been widely used in digital image processing.

In this paper, we calculated the texture features by Haralick method on 2D nodules images to evaluate the performance of texture features on nodule CADx. Then we also improved the 2D calculation method to 3D space for containing more information. The foundations and principles of the methods are described as follows.

3.1. 2D Haralick Features

This texture feature extraction method believes that the texture information can be adequately specified by a set of gray-tone spatial-dependence matrices (GTSDMs)

which are computed from various angular relationships and distances between neighboring resolution cell pairs (or image pixel pairs) on the image slice.

For one pixel on the 2D image except the edges, there are 8 neighborhoods with 1 pixel distance in four directions shown in Figure 2.

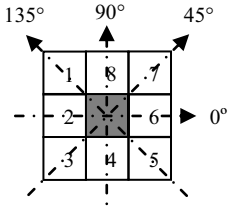

Fig. 2. Pixels 2 and 6 are 0° neighbors to the central pixel *; pixels 3 and 7 are 45° neighbors; pixels 4 and 8 are 90° neighbors; and pixels 1 and 5 are 135° neighbors.

Supposing the gray value of each image pixel belonging to the candidate nodule is quantized into N_g levels. Then the GTSDM can be calculated according to the gray value combinations of the neighbor and the central pixels in each of the four directions, which are shown in Table 1, where # (i, j) $(i, j = 0, 1, 2, \ldots, N_g)$ stands for the number of times gray tones i and j have been neighbors. Therefore, the resolution of each GTSDM is $N_g \times N_g$.

Table1. General form of any GTSDM for any direction.

Gray level	**0**	**1**	**2**	**…**	N_g
0	#(0, 0)	#(0, 1)	#(0, 2)	…	#(0, N_g)
1	#(1, 0)	#(1, 1)	#(1, 2)	…	#(1, N_g)
2	#(2, 0)	#(2, 1)	#(2, 2)	…	#(2, N_g)
…	…	…	…	…	…
N_g	#(N_g, 0)	#(N_g, 1)	#(N_g, 2)	…	#(N_g, N_g)

Then the GTSDMs of four directions can be calculated as the above description. From each GTSDM, 14 statistical measures can be calculated. The 14 statistical measures are Angular Second Moment, Contrast, Correlation, Sum of Squares, Inverse Difference Moment, Sum Average, Sum Variance, Sum Entropy, Entropy, Difference Variance, Difference Entropy, two Information Measures of Correlation, and Maximal Correlation Coefficient respectively. The calculation formulas can be found in paper [11].

Hence we can obtain a set of four values for each of the preceding 14 measures from four directions' matrices. Finally, a set of 28 Haralick texture features comprised by the mean and range of each of these 14 measures, averaged over the four values from the four directions, which can be used as inputs to the classifier.

3.2. 3D Haralick Features

According to the principle of 2D Haralick texture features calculation, the relationships of neighbor pixels on only four directions are considered. However, the pulmonary nodules are spheroidal, the information from 3D space may contain more beneficial information. Therefore, we analyzed more texture features on more different directions in 3D space.

The calculation of GTSDMs of 3D features is similar to that of the 2D Haralick features. Only more matrices of more 3D directions should be calculated from the volume data of the candidate nodule. In order to ensure the same distance between the voxels on the three coordinate axises, we interpolated the volume data to be isotropic. After this preprocessing, we only need to select the distance parameter of Haralick features calculation as one voxel. The schematic diagram of 3D directions contained in this paper is shown in Figure 3.

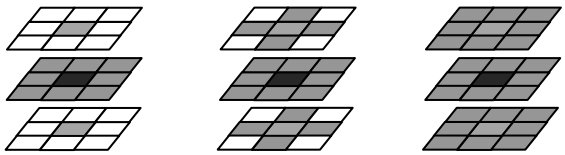

Fig. 3. Three combination modes of the neighbor voxels on 3D directions. (a) Combination of five directions; (b) Combination of nine directions; (c) Combination of thirteen directions.

From Figure 3, three different combinations of the neighbor voxels on 3D directions are shown. The central voxel is filled in black. And the voxels contained in each combination are filled in gray color. In each combination diagram, the three slices are adjacencies.

Therefore, we can get 13 GTSDMs from 13 directions in 3D space. And the 3D Haralick texture features are the average and range values of the 14 statistical measures from these directions. The numbers of directions are 5, 9, and 13 according to different combination modes.

Supposing H_{ij} is the jth (j=1, 2, …, 14) texture feature value in the ith direction (i=1, 2, …, n, n=5, 9, 13) calculated from the GTCMs proposed by Haralick, then the final Haralick features should be 2×14=28 which contain the mean and range of the texture feature values on the directions of different combination modes. The formulas can be shown as follows.

$$Mean_{jn} = \frac{1}{n}\sum_{i=1}^{n} H_{ij} \tag{1}$$

$$Range_{jn} = \max\{H_{ij}\} - \min\{H_{ij}\} \tag{2}$$

Then the 3D Haralick texture features can be obtained from the above formulas for the input of classification process.

4. EXPERIMENTS AND RESULTS

The purpose of this study is to evaluate the performance of texture features from CT images for pulmonary nodules' malignancy diagnosis and the consistency of the texture features and the expert visual diagnosis. For that purpose, all the nodule cases (905 nodules) with explicit benign or malignant labels from LIDC database were utilized which is so far the largest international database available. For the whole datasets, the nodules were labeled by five levels for malignancy with number 1 to 5. It means more suspicious of malignant with bigger

values. Therefore, we set the nodules with label '1' and '2' as benign nodules, '4' and '5' as malignant nodules, '3' as uncertain nodules which were not included in this study. But we will take attention to the uncertain nodules in future work.

To analysis the texture features of CT images, we adopt the most classical algorithm, namely 2D Haralick texture features. Then we also improved it into 3D space based on our prior work. To study the effectiveness of different kinds of 3D directions for diagnosis, we used three different combination modes of 2D and 3D directions according to the distances between the neighbor and the central voxels, which are called 5 directions mode, 9 directions mode, and 13 directions mode.

The 2D Haralick texture feature analysis is based on the single CT slice which contains the largest region from all the CT images containing the candidate nodule. And for 3D texture feature analysis, the volume data of the candidate nodule used in this paper is isotropic. To obtain the isotropic volume data, we used a cubic interpolation algorithm mentioned in [12].

By inputting the extracted texture features from the 905 nodules and the labels of two classes of malignance and benign into the well-known Support Vector Machine (SVM) with the widely-used kernel of Radial Basis Function (RBF) [13] and randomizing the training and testing process for 100 times, the corresponding 100 classification outcomes of sensitivity and specificity (i.e., the Az value) were obtained. From each set of 100 classification outcomes by different kinds of texture features, the mean and standard deviation measures were computed. Figure 4 shows the Receiver Operating Characteristics (ROC) curves of the four types of 2D and 3D Haralick texture features' performances. The measures of the area under the ROC curve Az are shown in Table 2.

Table 2. The AUC information of classifications based on 2D and 3D Haralick texture features.

Az Information	Mean	Standard D
2D Haralick Features (Single slice)	0.9373	0.0103
3D Haralick Features (5 Directions)	0.9426	0.0088
3D Haralick Features (9 Directions)	0.9440	0.0100
3D Haralick Features (13 Directions)	0.9441	0.0088

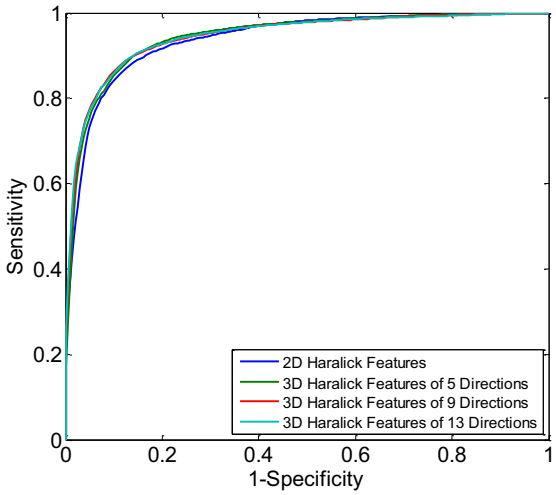

Fig 4. The ROC curves of the 2D and three types of 3D combination texture features' classification performances.

5. DISCUSSION AND CONCLUSION

To extract efficient features from CT images for lung CADx has always been a popular topic in these years. Some morphological features have been seemed as recognized standards for distinguishing the nodules from benign to malignant. For example, bigger nodules with rough and sharp edges are more suspicious of malignant ones. However, no matter volume measurement or shape recognition, accurately segmentation is a necessary component for the automatic diagnosis from medical images. Only texture features have lower requirement on the segmentation algorithm. Therefore, in this paper, we focus on evaluating the performance of texture features on malignancy diagnosis of nodules, and the study on the consistency of texture features and expert observer ratings based on morphology features.

LIDC-IDRI database is the largest open source of chest CT images with identified nodules, which collected multiple kinds of nodules from many institutions in the whole world. Although it lacks biopsy reports, it contains four radiologists' visual diagnosis reports of each nodule case. This is enough for our study to get a preliminary result of the effectiveness of texture features on lung CADx.

For the texture feature analysis, we adopt Haralick features, which is a well-known algorithm in digital image processing. But the former researchers all focus on the 2D Haralick feature analysis. Hence we improved this model into 3D space to add more information based on our prior work. And we also compared the contributions of texture features from different 3D directions. After extraction of texture features, the SVM classifier was used for training and testing. To achieve more stable and reliable results, 905 nodules (422 malignant and 483 benign) were divided into two half parts randomly for 100 times. On each time, half of the benign and malignant nodules were training in the SVM classifier, and the left half nodules were testing. Then we painted the ROC curves based on the 100 classifications.

And the average value of the best classification for the 100 times was calculated as the Az measure.

From the experimental results, we can make a conclusion that the 3D Haralick texture features can add more effectiveness to the malignancy diagnosis of pulmonary nodules than only the 2D Haralick texture features. And CADx based on texture features has an absolutely consistency with the expert observer distinguishing performance. In future work, we will focus on the nodule cases containing the biopsy reports for the effectiveness study on the 3D texture features for lung CADx.

ACKNOWLEDGEMENT

This work was partly supported by the NIH/NCI under grants #CA143111 and #CA082402. This work was also supported by the Chinese National Natural Science Foundation under grant #61071213 and #61172002, the National Science Foundation of China under grant #81071220, the Fundamental Research Funds for the Central Universities under grant N120518001 and Liaoning Natural Foundation 2013020021.

REFERENCES

[1] Heber MacMahon, John H. M. Austin, Gordon Gamsu, Christian J. Herold, James R. Jett, David P. Naidich, et al., Guidelines for Management of Small Pulmonary Nodules Detected on CT Scans: A Statement from the Fleischner Society, Radiology, Vol. 237, 395-400, 2005.

[2] Bram van Ginneken, Bart M. ter Haar Romeny, and Max A. Viergever, Computer-Aided Diagnosis in Chest Radiography: A Survey, IEEE Transactions on Medical Imaging, Vol. 20, No. 12, 1228-1241, 2001.

[3] K DOI, Current Status and Future Potential of Computer-Aided Diagnosis in Medical Imaging, The British Journal of Radiology, Vol. 78, S3-S19, 2005.

[4] Ayman El-Baz, Garth M. Beache, Georgy Gimel' farb, Kenji Suzuki, Kazunori Okada, Ahmed Elnakib, et al., Computer-Aided Diagnosis Systems for Lung Cancer: Challenges and Methodologies, International Journal of Biomedical Imaging, 2013.

[5] Ingrid C. Sluimer, Paul F. van Waes, Max A. Viergever, and Bram van Ginneken, Computer-Aided Diagnosis in High Resolution CT of the Lungs, Medical Physics, Vol. 30, No.12, 3081-3090, 2003.

[6] Michael K. Gould, James Fletcher, Mark D. Iannettoni, William R. Lynch, David E. Midthun, David P. Naidich, and David E. Ost, Evaluation of Patients with Pulmonary Nodules: When is It Lung Cancer? ACCP Evidence-Based Clinical Practice Guidelines, Chest, Vol. 132, 108S-130S, 2007.

[7] Dmitriy Zinovev, Daniela Raicu, Jacob Furst, and Samuel G. Armato III, Predicting Radiological Panel Opinions Using a Panel of Machine Learning Classifiers, Algorithms, Vol. 2, 1473-1502, 2009.

[8] Fangfang Han, Huafeng Wang, Bowen Song, Guopeng Zhang, Hongbing Lu, William Moore, Zhengrong Liang and Hong Zhao, A New 3D Texture Feature Based Computer-Aided Diagnosis Approach to Differentiate Pulmonary Nodules, Proc. SPIE 8670, Medical Imaging 2013: Computer-Aided Diagnosis, 86702Z (February 28, 2013).

[9] Rafael Wiemker, Martin Bergtholdt, Ekta Dharaiya, Sven Kabus and Michael C. Lee, Agreement of CAD features with expert observer ratings for characterization of pulmonary nodules in CT using the LIDC-IDRI database, Proc. SPIE 7260, Medical Imaging 2009: Computer-Aided Diagnosis, 72600H (February 27, 2009).

[10] Samuel G. Armato III, Geoffrey Mclennan, Luc Bidaut, Michael F. McNitt-Gray, Charles R. Meyer, Anthony P. Reeves, et al., The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans, Medical Physics, Vol. 38, 915--931, 2011.

[11] Robert M. Haralick, K. Shanmugam, and Its'Hak Dinstein, Textural Features for Image Classification, IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-3, No. 6, 610-621, November 1973.

[12] Frederick N. Fritsch and Robert E. Carlson, Monotone Piecewise Cubic Interpolation, Society for Industrial and Applied Mathematics, Vol. 17, No. 2, 238-246, 1980.

[13] Johan Suykens, Joos Vandewalle, Least Squares Support Vector Machine Classifiers, Neural Processing Letters, Vol. 9, No. 3, 293-300, June 1999.