Computerized comprehensive data analysis of Lung Imaging Database Consortium (LIDC)

Jun Tan,^{a)} Jiantao Pu,^{b)} Bin Zheng,^{c)} Xingwei Wang,^{d)} and Joseph K. Leader^{e)}

Department of Radiology, Imaging Research Division, University of Pittsburgh, 3362 Fifth Avenue, Pittsburgh, Pennsylvania 15213

(Received 26 February 2010; revised 28 April 2010; accepted for publication 1 June 2010; published 29 June 2010)

Purpose: Lung Image Database Consortium (LIDC) is the largest public CT image database of lung nodules. In this study, the authors present a comprehensive and the most updated analysis of this dynamically growing database under the help of a computerized tool, aiming to assist researchers to optimally use this database for lung cancer related investigations.

Methods: The authors developed a computer scheme to automatically match the nodule outlines marked manually by radiologists on CT images. A large variety of characteristics regarding the annotated nodules in the database including volume, spiculation level, elongation, interobserver variability, as well as the intersection of delineated nodule voxels and overlapping ratio between the same nodules marked by different radiologists are automatically calculated and summarized. The scheme was applied to analyze all 157 examinations with complete annotation data currently available in LIDC dataset.

Results: The scheme summarizes the statistical distributions of the abovementioned geometric and diagnosis features. Among the 391 nodules, (1) 365 (93.35%) have principal axis length ≤ 20 mm; (2) 120, 75, 76, and 120 were marked by one, two, three, and four radiologists, respectively; and (3) 122 (32.48%) have the maximum volume overlapping ratios $\geq 80\%$ for the delineations of two radiologists, while 198 (50.64%) have the maximum volume overlapping ratios $< 60\%$. The results also showed that 72.89% of the nodules were assessed with malignancy score between 2 and 4, and only 7.93% of these nodules were considered as severely malignant (malignancy ≥ 4).

Conclusions: This study demonstrates that LIDC contains examinations covering a diverse distribution of nodule characteristics and it can be a useful resource to assess the performance of the nodule detection and/or segmentation schemes. © 2010 American Association of Physicists in Medicine. [DOI: [10.1118/1.3455701](https://doi.org/10.1118/1.3455701)]

Key words: LIDC, lung nodule/cancer, CT, computer-aided detection and diagnosis

I. INTRODUCTION

Lung cancer is the leading cause of cancer related deaths in USA.¹ High resolution computed tomography (CT) has been widely accepted as an effective imaging modality in early detection of lung cancer (or nodules).² However, visual detection of lung nodules depicted on CT images is difficult and time consuming due to the large number of image slices involved in a single chest CT examination and the large interobserver variability.³ Hence, developing computer-aided detection or diagnosis (CAD) schemes for lung nodule has drawn significant research interests and efforts.⁴ To assess the performances of the CAD schemes developed by different research groups and to verify their potential clinical utility, it is important to have a common “gold standard.” In 2000, the National Cancer Institute (NCI) initiated a multi-institutional effort to build a public database named Lung Image Database Consortium (LIDC).^{5,6} The motivation was to allow researchers around the world to commonly use a large and diverse image dataset of the low-dose chest CT examinations for various purposes, especially for the development and evaluation of the related CAD schemes. The selected cases in LIDC have been evaluated by a panel of four experienced radiologists. Up to January 2010, the Na-

tional Biomedical Imaging Archive (NBIA) repository of the NCI released CT examinations of a total of 165 patients who underwent lung cancer screening at multiple medical institutions and allowed researchers to download the complete set of these cases from the NBIA server. In the past several years, several research groups used LIDC either as a training/testing dataset in developing and evaluating CAD schemes for lung nodule detection and/or classification.^{7,8}

Since CAD performance and reliability heavily depend on the quality or characteristics of the training and testing database,^{9,10} quantitatively evaluating the image characteristics of LIDC is important for the researchers to optimally use this database in different applications. However, the diversity and reliability of the diagnostic and image characteristics provided by the annotation files of LIDC have not been fully analyzed and provided to the public. This leads to potential difficulties or uncertainties for the researchers to use this database including the association of the malignancy of a nodule and its shape, the percentage of nodules with the certain characteristics queried by the researchers, and the global statistical analysis of interobserver variability. Furthermore, whenever new cases are added into LIDC, a new round of analysis has to be initiated to incorporate the intro-

FIG. 1. An example showing interobserver variability in detecting a nodule and delineating its boundary in a sequential CT slices. One radiologist (observer 2) marked that there were two individual nodules, while the other three radiologists marked as one nodule.

duced features of the newly added nodules, which is very time consuming. Hence, developing an efficient and reliable method and tool to automatically analyze and report the characteristics of LIDC is useful. In our research laboratory, we developed a fully automated computerized scheme to quantitatively analyze the characteristics of this developing database. Our goal is to assist researchers to better understand the advantages and limitations of LIDC and thus optimally use this database for lung cancer related investigations.

II. MATERIALS AND METHODS

Based on NCI-approved case selection protocol, LIDC developers have been acquiring and selecting CT examinations of the patients underwent lung cancer screening in several medical institutions. These CT examinations were acquired under x-ray tube current ranging from 40 to 260 mA and tube voltage at either 120 or 140 kVp. The CT images were reconstructed with pixel resolution ranging from 0.54 to 0.74 mm, depending on the patient body size and slice thickness ranging from 0.5 to 3.0 mm. Each selected case was evaluated by a panel of four board-certified chest radiologists using a two-phase strategy.¹¹ In the first phase (blinded), a case was independently read and interpreted by four participated radiologists to detect suspicious nodules. If the nodule cross-sectional diameter is measured ≥ 3 mm, the radiologist drew (delineated) entire nodule boundary and reported a set of diagnostic and shape features of the nodule such as malignancy, spiculation level, etc. If the size of the identified nodule is < 3 mm, radiologist only marked the voxel of the nodule center without delineating and providing the descriptions of nodule characteristics. In the second phase (unblinded), each radiologist reviewed his/her previous assessment by allowing accessing assessment results

provided by other radiologists and can decide whether to modify the original assessment. Through these two review phases, a final modified annotation file was generated for each CT examination. In summary, for each selected case, LIDC contains (1) a complete set of CT scan images with DICOM headers that store all technical factors about CT scan protocol and (2) the associated XML documentation file that contains nodule boundary contours marked by radiologists as well as related clinical information and diagnostic results (if available).

Our computerized scheme for LIDC data analyses includes the following steps. First, the scheme categorizes CT examinations according to their scan protocols by retrieving the global information embedded in the DICOM files, e.g., the distribution of the examinations in slice thickness. The scheme also detects the information completeness of each case. For example, the scheme detects whether individual patient has multiple CT examinations and whether the associated annotation files are missing or incomplete.

TABLE I. Number of cases and nodules with different slice thicknesses.

Slice thickness (mm)	No. of cases	No. of nodules
0.5	4	7
0.6236	1	1
0.625	27	68
0.7	18	69
1.25	4	12
1.8	51	138
2	13	31
3	27	65
Total	145	391

TABLE II. Distribution of nodules based on various characteristics.

Feature	Value range	No. of nodules	Percent (%)
Calcification	[3, 4)	52	13.30
	[4, 5)	11	2.81
	[5, 6)	13	3.32
	6	315	80.56
Sphericity	[1, 2)	2	0.51
	[2, 3)	32	8.18
	[3, 4)	155	39.64
	[4, 5)	161	41.18
Margin	5	41	10.49
	[1, 2)	17	4.35
	[2, 3)	35	8.95
	[3, 4)	73	18.67
Lobulation	[4, 5)	180	46.04
	5	86	21.99
	[1, 2)	256	65.47
	[2, 3)	71	18.16
Spiculation	[3, 4)	41	10.49
	[4, 5)	10	2.56
	≥ 5	13	3.32
	[1, 2)	296	75.70
Elongation	[2, 3)	52	13.30
	[3, 4)	23	5.88
	[4, 5)	10	2.56
	5	10	2.56
Malignancy	[1, 1.5)	57	14.58
	[1.5, 2.5)	282	72.12
	≥ 2.5	52	13.30
	[1, 2)	75	19.18
Principal axis length (mm)	[2, 3)	155	39.64
	[3, 4)	130	33.25
	[4, 5)	27	6.91
	5	4	1.02
	[3, 6)	58	14.83
	[6, 9)	151	38.62
	[9, 12)	90	23.02
	[12, 20)	66	16.88
	≥ 20	26	6.65

Second, the scheme retrieves each annotation file to count for the number of marked nodules and reconstructs the nodules surface in 3D volumetric space from the boundary delineation points. All delineation points are surface voxels. Voxels encircled by the boundary delineation points on the first and last slices of a nodule are surface voxels. Voxels

TABLE IV. The distribution of the number of nodules detected and delineated by different numbers of radiologists.

No. of observers	No. of nodules	Percent (%)
1	120	30.6
2	75	19.1
3	76	19.4
4	121	30.9

encircled by the boundary delineation points on other slices of that nodule are its body voxels. After 3D reconstruction, a set of image features for each marked nodule is computed including nodule size and shape features. Two features are computed to assess nodule size: (1) Principal axis length (PAL) and (2) approximated nodule volume (ANV). To compute PAL, the scheme applies a 3D frame to cover the nodule and computes the gravity center of the nodule and generates an eigenvector of the nodule volume. PAL is then defined by the distance between two intersection points of the eigenvector passing through the 3D frame boundary. To compute the ANV of a nodule, the scheme counts the total number of its surface and body voxels and then multiply this number by the volume of a single voxel. To assess nodule shape, the scheme computes elongation. Given a circular cylinder that has the same volume (Vol) of a nodule and its height is equal to the nodule's PAL, the elongation (E) of the nodule is defined as

$$E = \text{PAL} / \sqrt{4 \times \text{Vol} / (\pi \times \text{PAL})},$$

representing the actual ratio of PAL to the diameter of the bottom of the cylinder. If the elongation value of a nodule is ≥ 1.5 , this nodule is considered elongated. Since each nodule depicted on CT images can be identified and marked by multiple radiologists (i.e., 1–4), the scheme sequentially maps the different nodule boundary contours delineated by different radiologists to the CT images and computes multiple sets of the same features for the nodule. The average feature values are then computed by the scheme to represent each individual nodule.

Third, the scheme generates a series of summary tables to classify the nodules into different categories defined by a number of image features that are used by clinicians and/or CAD schemes in nodule detection and diagnosis. For example, nodule spiculation rating is an important feature to

TABLE III. The interobserver variability in different delineated nodule volume overlapping categories.

No. of observers overlapping ratio	1	2	3	4	Sum	Percent (%)
(0.0, 0.2)		0	1	3	4	1.02
[0.2, 0.4)		6	11	24	41	10.49
[0.4, 0.6)		27	50	76	153	39.13
[0.6, 0.8)		37	12	17	66	16.88
[0.8, 1.0)		3	2	0	5	1.28
1	120	2	0	0	122	31.20

TABLE V. The distribution of small nodules (with diameters less than 3 mm) marked by different numbers of radiologists.

No. of observers	No. of nodules	Percent (%)
1	376	36.5
2	262	25.5
3	239	23.2
4	153	14.8
Total	1030	100

predict the likelihood of malignancy and nodules with larger elongation ratio is more difficult to be detected by CAD schemes because it is often more ambiguous to be distinguished from the other linear lung tissue structures. Thus, CAD is likely to perform poorer in databases containing higher fraction of spiculated and elongated nodules. In summary, the scheme generates the report tables in the distribution of nodules in LIDC based on (1) CT exam slice thickness, (2) nodule size, and (3) diagnostic and shape features including calcification, sphericity, margin, lobulation, spiculation level, elongation ratio, and likelihood of malignancy. These tables aim to provide a comprehensively statistical summary of the identified nodules in LIDC.

Fourth, the scheme detects and quantifies the consistence of the nodule features. The scheme (1) identifies and analyzes the difference of the participated radiologists in nodule detection, (2) counts and summarizes the total number of individual nodules marked in the database, and (3) computes the distribution of the nodules marked by different number of radiologists and identifies whether one nodule was marked as different nodules in adjacent CT slice images. Based on the reconstructed nodule surface in 3D data, the scheme also evaluates the delineation difference made by different radiologists by computing the overlapping ratios of the segmented nodules. Figure 1 shows an example of nodule delineations marked by four radiologists on the same sets of CT image slices. In this scheme, given two nodule delineations a and b from the same nodules marked by two different radiologists, the interobserver variability (R_I) is computed as $R_I = |V_a \cap V_b| / |V_a \cup V_b|$, where V_a and V_b are the sets of voxels of the nodule defined by delineation a and b , respectively, $|\cdot|$ is the cardinality of the set, and R_I is the ratio of the number of voxels belonging to both nodules to the number of voxels belonging to either nodule. The scheme can also compute the interobserver variability for small nodules (<3 mm), in which a nodule was only marked with its center voxel. If a marked voxel from one radiologist is only N pixels (a predetermined distance threshold, which is 1 in our

TABLE VI. An example of comparison of data analysis results on four nodules selected by the users from one CT examination. These four nodules were marked by all four radiologists (the second column) with differently internal marked ID number (the third column). The fourth column shows the computed overlap ratios in which the higher ratio means lower interobserver variability.

Nodule ID	No. of radiologists	Internal marked ID	Average overlap ratio (%)
1	4	5, 25, 31, 44	50.5
2	4	6, 24, 27, 45	53.5
3	4	8, 23, 36, 47	69.7
4	4	9, 16, 33, 48	51.8

software) away along any axis (x , y , or z) in the 3D space from another voxel marked by a different radiologist, the scheme classifies that these two marked voxels represent the same nodule; otherwise, they represent different nodules. As a result, the scheme generates a set of supplemental tables to report the interobserver variability in LIDC database.

This scheme was developed using MATLAB 7.6.0 (R2008a) and implemented on a desktop based personal computer equipped with WINDOWS XP operating system, an AMD Athlon 64×2 dual core processor 5000+, and 3 Gbyte RAM memory. The program can be downloaded from the web or contact the authors.¹²

III. RESULTS

The current available LIDC database includes CT examinations of 165 patients among which one patient (with UID “13614193285030047”) had two examinations so the total number of CT examinations is 166. 9 of the examinations were recently added (after August 2009) without associated annotation files to date. Therefore, 157 CT examinations from 156 patients are available to be used and analyzed. The scheme reported 3213 nodule marks, among which 983 were delineations for nodules ≥ 3 mm and 2230 were single points for nodules <3 mm. The 983 annotated nodule markings represent 391 individual nodules. Table I shows the distribution of 391 nodules depicted on CT images with different slice thicknesses. Among them, 40.2% (157/391) were delineated in thin-slice CT images (slice thickness ≤ 1.25 mm) and the rest (59.8%) were delineated in thick-slice CT images. Table II summarizes the distributions of eight image features among 391 identified nodules. The results show that (1) 348 nodules (89%) have spiculation ratings less than 3, indicating the relatively smooth boundary; (2) 282 nodules (72%) have moderate elongation ratios, while only small fractions of nodules either have extreme

FIG. 2. Delineation is missed on slice 14 for this nodule.

FIG. 3. Correlation between shape features and malignancy.

elongation ratios (52/391) or resemble an almost perfect sphere (57/391); (3) 299 nodules (76%) have PAL between 3 and 12 mm (the computed maximum PAL is 41.12 mm and the minimum is 3.06 mm); and (4) there were five malignant nodules (malignancy rate=5) and 27 highly suspicious nodules (malignancy rate ≥ 4).

Nodule delineation is a very subjective concept depending heavily on radiologists' experience and/or preference. Table III shows the wide distribution of overlapping ratios of the nodules marked by multiple radiologists. Since one nodule typically depicts on several CT image slices, our scheme also detected that among 391 nodules, one was marked by one radiologist as two different nodules cross different CT slices, while other three radiologists marked it as a single nodule (Fig. 1). Tables IV and V summarize the interobserver variability in detecting and marking the nodules ≥ 3 and < 3 mm, respectively. The results show that 30.9% of large

nodules (≥ 3 mm) and 14.8% small nodules (< 3 mm) were marked by all four radiologists, respectively, indicating that interobserver variability may increase as the nodule size decreases.

Table VI shows one example of applying our scheme to analyze and compare a small set of four nodules selected by the users from one CT examination. It shows that although these four nodules were marked by all four radiologists, the marked (nodule counting) orders and the computed 3D nodule overlap ratios are different. In this example, the average overlap ratios for all four marked nodules are $< 70\%$.

Our scheme also detected a few inconsistencies between the images and the associated annotation files in three cases. In the first case, a marked nodule does not have "ROI" node in the DOM field of its XML file. Since diameter of this nodule is less than 3 mm, it does not affect our analysis results in this study. In the second case, the annotation file

FIG. 4. Correlation between shape features and calcification.

does not contain nodule delineation data. In the third case, one nodule extends from slices 13 to 24. However, its delineation on slice 14 is missing, so the nodule is broken into two parts along the z axis (Fig. 2).

IV. DISCUSSION

LIDC is a developing and widely used image database for developing and evaluating CAD schemes. However, the database does not include the detailed information or statistical data distribution of the nodules selected in the database. The motivation of this work is to provide investigators with an automated tool (scheme) to quantitatively and statistically analyze LIDC database. We demonstrated that this automated scheme was able to generate a series of summary tables to categorize the distributions of detected and marked nodules based on different diagnostic and shape features as well as to provide the detailed analysis results on the inter-

observer variability in both nodule detection and segmentation. All data analysis results are organized in a set of tables that show the nodule distributions under different image features or categories (Tables I and II) and interobserver variability (Tables III–V). Investigators can also use these summary results to conduct other types of data analysis tasks. For example, we analyzed the correlations between a number of image features (including 3D principal axis length, sphericity, elongation, spiculation level, and lobulation) and the likelihood of the nodules being malignancy. The five charts (Fig. 3) show that there is no clear correlation between each of the five nodule shape or boundary related features and malignancy. Thus, we concluded that based on the currently identified and marked nodules in LIDC, the likelihood of a nodule being malignancy was insensitive to size, shape, and boundary smoothness of the nodules. We also found that, in general, there were no clear correlations between shape fea-

tures and the calcification numbers associated with nodules (Fig. 4). However, as shown in Figs. 4(a)–4(c), the large (i.e., principal axis length >20 mm) or extreme elongated (i.e., sphericity <3 or elongation >3) nodules are likely to be the highly calcified.

In CAD development, researchers often pursue to have a database with reliable “ground truth.” Unfortunately, the lack of pathology verification in the majority of the cases and the substantially large interobserver variability in nodule detection are the major limitations of LIDC. Because it is very difficult to distinguish anatomically where a nodule should be separated from the surrounding non-nodule lung tissue, there is often no “ground truth” about the real nodule volume and shape. As a result, most of the previous studies used the consensus or average results provided by radiologists as the reference “standard.” Our data analysis results show that the nodule boundaries defined by radiologists vary substantially (i.e., Fig. 1). The researchers in CAD development should be aware of such limitation and use other methods to compensate the potentially negative impact of using the reference standard provided by LIDC.

In summary, we presented an automated tool (scheme) for quantitative analysis of LIDC and we also tested this scheme by conducting a comprehensive data analysis of the most updated LIDC. This scheme will be made publicly available so that other researchers can use it to analyze and review updated LIDC database at anytime and track the latest status. The comprehensive database analysis results will help researchers optimally select and use entire or partial database of LIDC for different research purpose.

ACKNOWLEDGMENT

This work was supported in part by Grant Nos. P50 HL084948, R01 HL085096, and R01 HL096613 from Na-

tional Heart, Lung, and Blood Institute, National Institutes of Health, to the University of Pittsburgh.

^{a)} Author to whom correspondence should be addressed. Electronic mail: tanj@upmc.edu; Telephone: (412)-641-2570; Fax: (412)-641-2582.

^{b)} Electronic mail: puj@upmc.edu

^{c)} Electronic mail: zhengb@upmc.edu

^{d)} Electronic mail: wangx6@upmc.edu

^{e)} Electronic mail: leaderjk@upmc.edu

¹ American Cancer Society, Cancer Facts and Figures 2009, <http://www.cancer.org/downloads/STT/500809web.pdf>.

² P. B. Bach, M. J. Kelley, R. C. Tate, and D. C. McCrory, “Screening for lung cancer: A review of the current literature,” *Chest* **123**, 72S–82S (2003).

³ J. K. Leader *et al.*, “Pulmonary nodule detection with low-dose computed tomography of the lung: Agreement among radiologists,” *AJR, Am. J. Roentgenol.* **185**, 973–978 (2005).

⁴ R. Wiemker *et al.*, “Aspects of computer-aided detection (CAD) and volumetry of pulmonary nodules using multislice CT,” *Br. J. Radiol.* **78**, S46–S56 (2005).

⁵ National Cancer Institute, LIDC, <http://imaging.cancer.gov/programsandresources/InformationSystems/LIDC/page2>.

⁶ S. G. Armato *et al.*, “The Lung Image Database Consortium (LIDC): An evaluation of radiologist variability in the identification of lung nodules on CT scans,” *Acad. Radiol.* **14**, 1409–1421 (2007).

⁷ T. W. Way *et al.*, “Computer-aided diagnosis of pulmonary nodules on CT scans: Segmentation and classification using 3D active contours,” *Med. Phys.* **33**, 2323–2337 (2006).

⁸ R. A. Ochs *et al.*, “Automated classification of lung bronchovascular anatomy in CT using AdaBoost,” *Med. Image Anal.* **11**, 315–324 (2007).

⁹ R. M. Nishikawa *et al.*, “Effect of case selection on the performance of computer-aided detection schemes,” *Med. Phys.* **21**, 265–269 (1994).

¹⁰ B. Zheng, Y. H. Chang, W. F. Good, and D. Gur, “Adequacy testing of training set sample sizes in the development of a computer-assisted diagnosis scheme,” *Acad. Radiol.* **4**, 497–502 (1997).

¹¹ M. McNitt-Gray *et al.*, “The Lung Image Database Consortium (LIDC) data collection process for nodule detection and annotation,” *Acad. Radiol.* **14**, 1464–1474 (2007).

¹² See supplementary material E-MPHYA6-37-044007 which contains the source code files developed using MABLAB and results output for user’s convenience.