

The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Reference Database of Lung Nodules on CT Scans

Samuel G. Armato III^{a)}

Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, MC 2026, Chicago, Illinois 60637

Geoffrey McLennan

Department of Internal Medicine, Pulmonary Division, University of Iowa Carver College of Medicine, 200 Hawkins Drive, Iowa City, Iowa 52242

Luc Bidaut^{b)}

University of Texas, MD Anderson Cancer Center, Houston, Texas 77030

Michael F. McNitt-Gray

Department of Radiological Sciences, David Geffen School of Medicine at UCLA, 924 Westwood Boulevard, Los Angeles, California 90024

Charles R. Meyer

Department of Radiology, University of Michigan Medical School, 109 Zina Pitcher Place, A522, Ann Arbor, Michigan 48109

Anthony P. Reeves

School of Electrical and Computer Engineering, Cornell University, 392 Rhodes Hall, Ithaca, New York 14853

Binsheng Zhao^{c)}

Department of Medical Physics, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065

Denise R. Aberle

Department of Radiological Sciences, David Geffen School of Medicine at UCLA, 924 Westwood Boulevard, Los Angeles, California 90024

Claudia I. Henschke^{d)}

Department of Radiology, Mount Sinai School of Medicine, 1 Gustave Levy Place, New York, New York 10029

Eric A. Hoffman

Department of Radiology, University of Iowa Carver College of Medicine, 200 Hawkins Drive, Iowa City, Iowa 52242

Ella A. Kazerooni

Department of Radiology, University of Michigan Health System, Cardiovascular Center Number 5482, 1500 East Medical Center Drive, Ann Arbor, Michigan 48109

Heber MacMahon

Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, MC 2026, Chicago, Illinois 60637

Edwin J. R. van Beek^{e)}

Department of Radiology, University of Iowa Carver College of Medicine, 200 Hawkins Drive, Iowa City, Iowa 52242

David Yankelevitz^{d)}

Department of Radiology, Mount Sinai School of Medicine, 1 Gustave Levy Place, New York, New York 10029

Alberto M. Biancardi

School of Electrical and Computer Engineering, Cornell University, 392 Rhodes Hall, Ithaca, New York 14853

Peyton H. Bland

Department of Radiology, University of Michigan, A502 BSRB, 109 Zina Pitcher Place, Ann Arbor, Michigan 48109

Matthew S. Brown

Department of Radiological Sciences, David Geffen School of Medicine at UCLA, 924 Westwood Boulevard, Los Angeles, California 90024

Roger M. Engelmann

Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, MC 2026, Chicago, Illinois 60637

Gary E. Laderach

Department of Radiology, University of Michigan, A502 BSRB, 109 Zina Pitcher Place, Ann Arbor, Michigan 48109

Daniel Max^{d)}

Department of Radiology, Mount Sinai School of Medicine, 1 Gustave Levy Place, New York, New York 10029

Richard C. Pais and David P.-Y. Qing^{f)}

Department of Radiological Sciences, David Geffen School of Medicine at UCLA, 924 Westwood Boulevard, Los Angeles, California 90024

Rachael Y. Roberts^{g)}

Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, MC 2026, Chicago, Illinois 60637

Amanda R. Smith

Department of Radiology, University of Iowa Carver College of Medicine, 200 Hawkins Drive, Iowa City, Iowa 52242

Adam Starkey

Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, MC 2026, Chicago, Illinois 60637

Poonam Batra^{h)}

Department of Radiological Sciences, David Geffen School of Medicine at UCLA, 924 Westwood Boulevard, Los Angeles, California 90024

Philip Caligiuriⁱ⁾

Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, MC 2026, Chicago, Illinois 60637

Ali Farooqi^{d)}

Department of Radiology, Mount Sinai School of Medicine, 1 Gustave Levy Place, New York, New York 10029

Gregory W. Gladish

Department of Diagnostic Radiology, MD Anderson Cancer Center, Unit 1478, 1515 Holcombe Boulevard, Houston, Texas 77030

C. Matilda Jude

Department of Radiological Sciences, David Geffen School of Medicine at UCLA, 924 Westwood Boulevard, Los Angeles, California 90024

Reginald F. Munden

Department of Diagnostic Imaging, MD Anderson Cancer Center, Unit 1478, 1515 Holcombe Boulevard, Houston, Texas 77030

Iva Petkovska^{j)}

Department of Radiological Sciences, David Geffen School of Medicine at UCLA, 924 Westwood Boulevard, Los Angeles, California 90024

Leslie E. Quint

Department of Radiology, University of Michigan Health Systems, 1500 East Medical Center Drive, Ann Arbor, Michigan 48109

Lawrence H. Schwartz^{k)}

Department of Radiology, Box 29, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065

Baskaran Sundaram

Department of Radiology, University of Michigan Health Systems, CVC 5481, 1500 East Medical Center Drive, Ann Arbor, Michigan 48109

Lori E. Dodd^{l)}

National Cancer Institute, 6130 Executive Boulevard, Bethesda, Maryland 20892

Charles Fenimore

Information Access Division, National Institute of Standards and Technology, MS 8940, 100 Bureau Drive, Gaithersburg, Maryland 20899

David Gur

Department of Radiology, University of Pittsburgh, 3632 Fifth Avenue, Pittsburgh, Pennsylvania 15213

Nicholas Petrick

United States Food and Drug Administration, WO62-4118, 10903 New Hampshire Avenue, Silver Spring, Maryland 20993

John Freymann and Justin Kirby

SAIC-Frederick, Inc., 6130 Executive Boulevard, Bethesda, Maryland 20892

Brian Hughes

TerpSys, 2115 East Jefferson Street, Suite 6000, Rockville, Maryland 20852

Alessi Vande Casteele

Agfa HealthCare NV, Septestraat 27, 2640 Mortsels, Belgium

Sangeeta Gupta

FUJIFILM Medical Systems USA, Inc., 419 West Avenue, Stamford, Connecticut 06902

Maha Sallam^{m)}

iCAD, Inc., Global Headquarters, 98 Spit Brook Road, Suite 100, Nashua, New Hampshire 03062

Michael D. Heath

Carestream Health Inc., 1049 Ridge Road West, Rochester, New York 14615

Michael H. Kuhn

Philips Medical Systems DMC GmbH, Roentgenstrasse 24, D-22315 Hamburg, Germany

Eka Dharaiya

Philips Healthcare, 595 Miner Road, Highland Heights, Ohio 44143

Richard Burns and David S. Fryd

Riverain Medical, 3020 South Tech Boulevard, Miamisburg, Ohio 45342

Marcos Salganicoff and Vikram Anand

Siemens Medical Solutions USA, Inc., 51 Valley Stream Parkway, Malvern, Pennsylvania 19355

Uri Shreterⁿ⁾

GE Healthcare, 3000 North Grandview Boulevard, W1120, Waukesha, Wisconsin 53188

Stephen Vastagh

Medical Imaging and Technology Alliance (MITA), 1300 North 17th Street, Suite 1752, Arlington, Virginia 22209

Barbara Y. Croft and Laurence P. Clarke

Cancer Imaging Program, National Cancer Institute, 6130 Executive Boulevard, Bethesda, Maryland 20892

(Received 8 September 2010; revised 16 November 2010; accepted for publication 20 November 2010; published 24 January 2011)

Purpose: The development of computer-aided diagnostic (CAD) methods for lung nodule detection, classification, and quantitative assessment can be facilitated through a well-characterized repository of computed tomography (CT) scans. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI) completed such a database, establishing a publicly available reference for the medical imaging research community. Initiated by the National Cancer Institute (NCI), further advanced by the Foundation for the National Institutes of Health (FNIH), and accompanied by the Food and Drug Administration (FDA) through active participation, this public-private partnership demonstrates the success of a consortium founded on a consensus-based process.

Methods: Seven academic centers and eight medical imaging companies collaborated to identify, address, and resolve challenging organizational, technical, and clinical issues to provide a solid foundation for a robust database. The LIDC/IDRI Database contains 1018 cases, each of which includes images from a clinical thoracic CT scan and an associated XML file that records the results of a two-phase image annotation process performed by four experienced thoracic radiologists. In the initial blinded-read phase, each radiologist independently reviewed each CT scan and marked lesions belonging to one of three categories (“nodule ≥ 3 mm,” “nodule < 3 mm,” and “non-nodule ≥ 3 mm”). In the subsequent unblinded-read phase, each radiologist independently reviewed their own marks along with the anonymized marks of the three other radiologists to render

a final opinion. The goal of this process was to identify as completely as possible all lung nodules in each CT scan without requiring forced consensus.

Results: The Database contains 7371 lesions marked “nodule” by at least one radiologist. 2669 of these lesions were marked “nodule ≥ 3 mm” by at least one radiologist, of which 928 (34.7%) received such marks from all four radiologists. These 2669 lesions include nodule outlines and subjective nodule characteristic ratings.

Conclusions: The LIDC/IDRI Database is expected to provide an essential medical imaging research resource to spur CAD development, validation, and dissemination in clinical practice. [DOI: [10.1118/1.3528204](https://doi.org/10.1118/1.3528204)]

Key words: lung nodule, computed tomography (CT), thoracic imaging, interobserver variability, computer-aided diagnosis (CAD)

I. INTRODUCTION

Publicly available medical image databases for the development and evaluation of computerized image analysis paradigms have been anticipated for nearly two decades.¹ Although the development of computer-aided diagnostic (CAD) methods has accelerated, access to well-characterized image data remains a common limitation as the task of identifying and collecting appropriate images for any specific research activity is a laborious and expensive process. An organized collection of anonymized clinical images alone would provide a valuable resource² and would eliminate database composition as a source of variability that hinders the appropriate comparison of different CAD methods.^{3,4} The utility of an image database would be greatly enhanced through the inclusion of task-specific “truth.” Investigators developing automated detection methods, for example, require the opinion of an experienced radiologist or, more appropriately, a panel of radiologists regarding the location of lesions within the images. Truth for other CAD tasks requires data such as follow-up images to evaluate change over time, pathology reports, or radiologist-drawn lesion outlines. The increasing need for CAD in the clinical practice of radiology lends urgency to the creation of common image databases with established truth to foster the development of CAD methods and enable the direct comparison of different systems.

Publicly available image databases designed to facilitate computerized image analysis research were first introduced in mammography. The most notable of these databases is the Digital Database for Screening Mammography (DDSM),^{5–7} which contains 2620 digitized four-view screening mammograms. Lesions have been annotated by an experienced radiologist to include an American College of Radiology (ACR) keyword description, BI-RADS (Breast Imaging-Reporting and Data System) rating, subtlety score, and a manual outline.

Chest radiography is the most commonly performed radiologic study, and the detection of lung nodules is one of the most important diagnostic challenges in chest radiography. This detection task became an early focus of CAD research in thoracic imaging⁸ and the motivation for the Japanese Society of Radiological Technology (JSRT) to create a publicly available database of chest radiographs for education, training,

and research. The JSRT database contains 247 digitized posteroanterior chest radiographs with either a solitary pulmonary nodule ($n=154$) or no nodule ($n=93$), as confirmed by CT and reviewed by three experienced thoracic radiologists.⁹ Each case includes patient information such as age and gender along with nodule size, malignancy status, subtlety rating, coarse anatomic location, and coordinates of the nodule center.

Cornell University in conjunction with the National Cancer Institute (NCI) and funding from the Prevent Cancer Foundation has made publicly available a growing research database of serial CT scans with nodule outlines provided by radiologists.¹⁰ The intent of the Public Image Database is to facilitate the development of computerized methods for the assessment of tumor response to therapy. A set of interactive image viewing tools is provided along with lesion measurements and growth analysis. Databases that allow for the quantitative analysis of serial CT scans are becoming more relevant to radiologic and oncologic research.^{11,12}

The collections of images acquired during comprehensive lung cancer screening trials have the potential to become valuable database resources. One of the first such trials, the Early Lung Cancer Action Program (ELCAP), made available in 2003 the ELCAP Public Lung Image Database. This database consists of 50 documented low-dose CT scans for the performance evaluation of computer-aided detection systems. The National Lung Screening Trial (NLST) randomized 26 724 subjects to the CT screening arm of its two-arm study. From among the 75 133 low-dose thoracic CT scans acquired at the 33 participating institutions according to a strict image-acquisition protocol, 48 547 scans were archived in the CT Image Library (CTIL).^{13,14} Deidentified images were transferred to a central site, which performed quality assurance on the images through confirmation of select digital imaging and communications in medicine (DICOM) fields to ensure accurate transmittal of the correct scan and through visual inspection to ensure image quality. Although the images were not annotated with lesion attributes, demographic and clinical data were maintained for eventual use by researchers once the library becomes publicly available.

The NELSON (Nederlands Leuvens Longkanker Screeningsonderzoek) trial, a Dutch acronym for “Dutch-Belgian lung cancer screening trial,” has accrued 15 523 participants across four institutions since 2003.^{15,16} Annual CT screening

studies were interpreted first at the local institution and then again at a central site. CT scans from the NELSON study have been used by investigators associated with the project to investigate, for example, interobserver variability of semi-automated lung nodule volume measurements,¹⁷ the discrimination between benign and malignant nodules,^{18,19} automated lung nodule detection,^{20,21} and automated lung segmentation.²² The research value of image databases acquired during clinical studies has been realized in other anatomic sites as well, such as CT colonography.^{23,24}

The development of CAD methods for lung nodule detection, classification, and quantitative assessment can be facilitated and stimulated through the creation of a well-characterized repository of thoracic CT scans. A true reference database, however, would provide an even greater benefit to investigators but would require an even greater commitment of time and resources to create the standards and infrastructure required to capture metadata, such as image annotations and pathologic diagnosis. To this end, the NCI issued a request for applications (RFAs) entitled “Lung Image Database Resource for Imaging Research” in April 2000 to convene a consortium of institutions to develop consensus guidelines for the creation of a CT-based lung nodule reference database.²⁵ Five institutions (Weill Cornell Medical College, University of California, Los Angeles, University of Chicago, University of Iowa, and University of Michigan) were selected to form the Lung Image Database Consortium (LIDC), which has been working since 2001 to develop a web-accessible research resource for the development, training, and evaluation of CAD methods for lung nodules to include (1) an image repository of screening and diagnostic thoracic CT scans, (2) associated metadata such as technical scan parameters (e.g., slice thickness, tube current, and reconstruction algorithm) and patient information (e.g., age, gender, and pathologic diagnosis), and (3) nodule truth information²⁶ based on the subjective assessments of multiple experienced radiologists (e.g., lesion category, nodule outlines, and subtlety ratings).²⁷

Guided by the premise that “public-private partnerships are essential to accelerating scientific discovery for human health” and their successes in this realm,²⁸ the Foundation for the National Institutes of Health (FNIH) created the Image Database Resource Initiative (IDRI) in 2004 to further advance the efforts of the LIDC. The IDRI joined the five LIDC institutions with two additional academic centers (MD Anderson Cancer Center and Memorial Sloan-Kettering Cancer Center) and eight medical imaging companies (AGFA Healthcare, Carestream Health, Inc., Fuji Photo Film Co., GE Healthcare, iCAD, Inc., Philips Healthcare, Riverain Medical, and Siemens Medical Solutions). Through the IDRI, these companies provided additional resources to expand substantially the LIDC database to a targeted 1000 CT scans and to create a complementary database of almost 300 digital chest radiographic images associated with a subset of these CT scans. The experience with chest radiographs will be the subject of a future publication. The IDRI merged the expertise of the academic centers with that of the medical imaging companies. Since the process of database collection,

annotation, and curation was exactly the same for the LIDC database and the CT component of the IDRI database, the combined database of thoracic CT scans will be referred to as the LIDC/IDRI Database.

The creation of a reference database through a consensus-based process required careful planning and the proper consideration of fundamental issues such as a governing mission statement, CT scan inclusion criteria, an appropriate definition of target lesions and associated truth requirements, a process model to guide population of the Database, and a framework to direct the application of assessment methodologies by end users. The details of these issues and the evolution of the decisions implemented by the LIDC/IDRI have been reported previously.^{27,29} The purpose of this paper is to describe the now-completed, publicly available LIDC/IDRI Database of 1018 thoracic CT scans and associated radiologist annotations. A solid understanding of the process through which the Database was created, along with important caveats on its use, is required to ensure that investigators conduct studies that are compatible with valid uses of the Database, while at the same time allowing investigators to take full advantage of the available information. Imparting this knowledge transfers the responsibility for valid use of the Database to individual investigators and to the scientific community so that the peer-review process for grants and publications can function appropriately. Ultimately, the success of the LIDC/IDRI effort will be judged by its impact on the community through the quality of grants awarded, the relevance of derivative publications, and the dissemination of CAD for thoracic CT into clinical practice after successful routing through the regulatory approval process.

II. MATERIALS AND METHODS

II.A. Patient image data

The LIDC/IDRI Database contains a total of 1018 helical thoracic CT scans collected retrospectively, with appropriate local IRB approval, from the picture archiving and communications systems (PACS) of the seven participating academic institutions. Anonymization software was applied to remove all protected health information (PHI) contained within the DICOM headers of the images in accordance with Health Insurance Portability and Accountability Act (HIPAA) guidelines.³⁰ No scan was performed specifically for the purpose of the Database so that a heterogeneous range of scanner models and technical parameters was intentionally represented. The intent was to include only a single scan from any one patient so that scans in the Database would not be correlated. As a result, the LIDC/IDRI Database is not amenable to temporal change analysis research; other publicly available databases, however, such as the NCI’s Reference Image Database to Evaluate Response to therapy in lung cancer¹² (RIDER) and Cornell University’s database provide such resources.

Certain inclusion criteria were imposed to ensure relevance of the scans to the development of state-of-the-art CAD systems.²⁷ These criteria evolved from a consensus-based process conducted over numerous telephone confer-

ences and meetings of the twelve-member LIDC Steering Committee, which included radiologists and CAD researchers. Both standard-dose diagnostic CT scans and lower-dose CT scans from lung cancer screening examinations were acceptable. Each scan selected for the Database was required to have a collimation and reconstruction interval no greater than 3 mm (advances in technology forced a reduction from the 5 mm threshold initially published by the LIDC); no requirements with regard to scanner pitch, exposure, tube voltage, or reconstruction algorithm were imposed. Scans were limited to approximately six lung nodules with longest dimension less than 30 mm (consistent with the accepted upper limit of nodule size³¹) and greater than or equal to 3 mm (a lower limit imposed for practical considerations²⁷), as determined by a cursory (and nonrecorded) review during case selection at the originating institution; the identification of a greater number of nodules during the subsequent image annotation process, however, was not grounds for case exclusion, and the image annotation process allowed for independent assessments of nodule size. The presence of other pathology, high levels of noise, and streak, motion, or metal artifacts was allowed unless these features compromised nodule interpretation, which was a judgment made by the LIDC radiologist at the originating institution during case selection. A nodule could be primary lung cancer, metastatic disease, a noncancerous process, or indeterminate in nature.

The 1018 CT scans had been acquired from 1010 different patients; it was retrospectively determined that two distinct scans from each of eight patients inadvertently had been included among the 1018 scans. These scans nevertheless were retained in the Database since the effort for image annotation already had been invested; users of the Database may identify these cases by the common patient ID in the respective image headers. A range of scanner manufacturers and models was represented (670 scans from seven different GE Medical Systems LightSpeed scanner models, 74 scans from four different Philips Brilliance scanner models, 205 scans from five different Siemens Definition, Emotion, and Sensation scanner models, and 69 scans from Toshiba Aquilion scanners). (The mention of commercial equipment is intended to specify the conditions of the present study and is not an endorsement by the LIDC/IDRI Research Group of this equipment.) The tube peak potential energies used for scan acquisition were as follows: 120 kV ($n=818$), 130 kV ($n=31$), 135 kV ($n=69$), and 140 kV ($n=100$). Tube current ranged from 40 to 627 mA (mean: 222.1 mA). Slice thicknesses were 0.6 mm ($n=7$), 0.75 mm ($n=30$), 0.9 mm ($n=2$), 1.0 mm ($n=58$), 1.25 mm ($n=349$), 1.5 mm ($n=5$), 2.0 mm ($n=124$), 2.5 mm ($n=322$), 3.0 mm ($n=117$), 4.0 mm ($n=1$), and 5.0 mm ($n=3$). Reconstruction interval ranged from 0.45 to 5.0 mm (mean: 1.74 mm). The in-plane pixel size ranged from 0.461 to 0.977 mm (mean: 0.688 mm). While the convolution kernels used for image reconstruction differ among manufacturers, these convolution kernels may be classified broadly as “soft” ($n=67$), “standard/

nonenhancing” ($n=560$), “slightly enhancing” ($n=264$), and “overenhancing” ($n=127$) (in order of increasing spatial frequencies accentuated by each class).

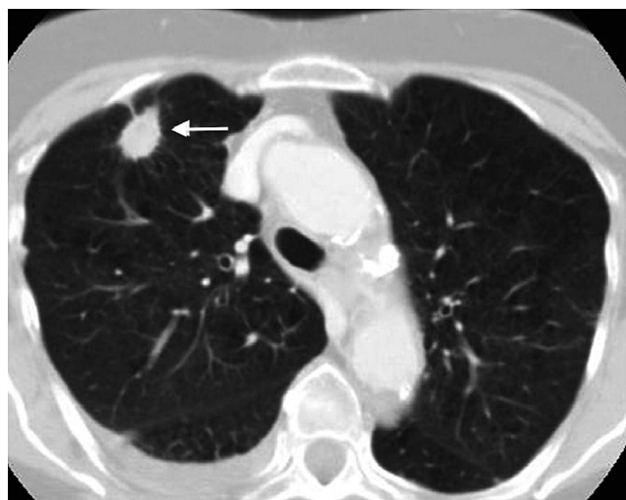
II.B. Image annotation process

To identify as completely as possible all lung nodules in a scan without requiring forced consensus, a two-phase process was developed for the asynchronous interpretation of CT scans by a thoracic radiologist at each of four different LIDC/IDRI institutions (although five of the seven academic institutions participated in the interpretation process overall, only four institutions contributed to the interpretation of any one scan), as previously reported.²⁹ A total of 12 radiologists participated in the image annotation process across all five sites over the course of the project. A comprehensive set of written instructions was available to each participating radiologist. These instructions evolved from a consensus-based process conducted over numerous telephone conferences and meetings of the twelve-member LIDC Steering Committee. In summary, the initial “blinded read phase” required each of the four radiologists to independently review a scan using a computer interface and mark lesions they identified as

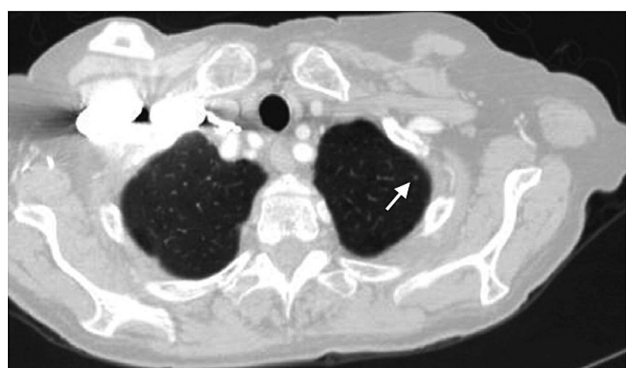
- (1) “nodule ≥ 3 mm” (defined as any lesion considered to be a nodule with greatest in-plane dimension in the range 3–30 mm regardless of presumed histology) [Fig. 1(a)],
- (2) “nodule < 3 mm” (defined as any lesion considered to be a nodule with greatest in-plane dimension less than 3 mm that is not clearly benign) [Fig. 1(b)] or,
- (3) “non-nodule ≥ 3 mm” (any other pulmonary lesion, such as an apical scar, with greatest in-plane dimension greater than or equal to 3 mm that does not possess features consistent with those of a nodule) [Fig. 1(c)].^{29,32}

Inherent in the definitions of all three lesion categories is the concept of a “nodule,” which was deliberately not defined by the LIDC/IDRI Research Group. In an earlier publication,²⁷ we recognized that the notion of nodule may not represent a single entity capable of verbal definition, and we suggested that the term nodule is more appropriately applied to a spectrum of abnormalities, which is itself a subset of a broader spectrum of abnormalities that we termed “focal abnormality.” Based on this conceptualization, all nodules are focal abnormalities, but not all focal abnormalities are nodules. The two spectra span a multidimensional space that comprises lesion characteristics such as shape, texture, and margin sharpness. Within this context, each radiologist provided their own interpretation of the “noduleness” of each observed lesion during the image annotation process.

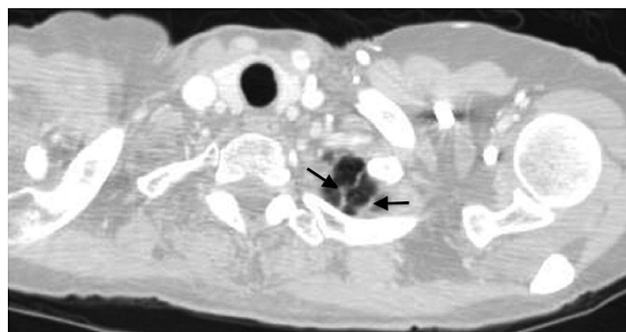
For each “nodule ≥ 3 mm” identified by a radiologist, that radiologist used the computer interface to construct outlines around the nodule in each CT section in which it appeared; for each lesion in one of the other two lesion categories identified by a radiologist, that radiologist used the computer interface to mark the approximate three-dimensional center-of-mass location. Electronic measure-



(a)



(b)



(c)

FIG. 1. Examples of lesions considered to satisfy the LIDC/IDRI definition of (a) a nodule ≥ 3 mm, (b) a nodule < 3 mm, and (c) a non-nodule ≥ 3 mm (reprinted with permission from Ref. 29).

ment tools were available to help the radiologists determine whether a lesion's dimension exceeded the 3 mm threshold. Only transaxial sections were reviewed; nonaxial reformatted images and maximum-intensity projection images were not available, since such viewing configurations were not standard at all LIDC/IDRI institutions when data collection began. Each CT scan was initially presented at a standard brightness/contrast setting without magnification, but the radiologists were allowed to adjust brightness, contrast, and magnification as appropriate to enable the most complete interpretation of the scan.

During the subsequent unblinded read phase, the anonymized blinded read results of all radiologists were revealed to each of the radiologists, who then independently reviewed their marks along with the anonymous marks of their colleagues; a radiologist's own marks then could be left unchanged, deleted, switched in terms of lesion category, or additional marks could be added. Each radiologist was required to inspect all nodule < 3 mm and nodule ≥ 3 mm marks placed during the blinded read; this requirement was not imposed on non-nodule ≥ 3 mm marks. For each lesion that a radiologist identified as a nodule ≥ 3 mm after the unblinded read phase, that radiologist independently assessed subjective characteristics of the nodule such as subtlety, internal structure, spiculation, lobulation, shape (sphericity), solidity, margin, and likelihood of malignancy.²⁹ Each radiologist's lesion-category designation and associated marks (spatial locations of all points in the outlines constructed for a nodule ≥ 3 mm along with its characteristics and center-of-mass locations for a nodule < 3 mm and for a non-nodule ≥ 3 mm) for each lesion were stored in a single XML file for each scan after the unblinded read phase (the XML schema is located at <http://troll.rad.med.umich.edu/lidc/>). The blinded and unblinded read phases were intended to comprise a single, comprehensive process; therefore, the LIDC/IDRI Database only contains the final set of post-unblinded-read-phase marks in each of the 1018 XML files.

The nodule ≥ 3 mm lesion category was the main focus of the Database; consequently, the research potential of these lesions was enhanced through the inclusion of radiologist outlines to capture spatial extent and the subjective assessment of nodule characteristics. Each outline was meant to be a localizing "outer border" so that, in the opinion of the radiologist, the outline itself did not overlap pixels belonging to the nodule. The radiologists were able to explicitly outline regions of exclusion within a nodule (an air-filled cavity, for example), which were then recorded as such in the XML file (Fig. 2). Three different in-house software systems were used to create nodule outlines and capture subjective nodule characteristic ratings. Each of three institutions used their own software, with which their radiologists were most familiar. The two institutions without in-house software both adopted the same system from another institution. One of these systems allowed for semiautomated creation of nodule outlines, while the other two systems were completely manual. The decision to allow multiple nodule outlining approaches was made after we conducted a study that demonstrated that the variation in nodule outlines derived from different radiologists substantially exceeded variation derived from different software tools.³³ One of the three systems, the one used by three institutions, uses a semiautomated technique³⁴ based on the Otsu method³⁵ to compute a threshold for region growing. The system also provides interactive editing tools including region addition, subtraction, and morphological operations. Another system, the SIMBA image marking tool, was used by the Cornell radiologists. This completely web-based tool obtains images from a SIMBA web server. All computer assistance was disabled so that nodule outlines were created manually. The use of different software systems for data ac-

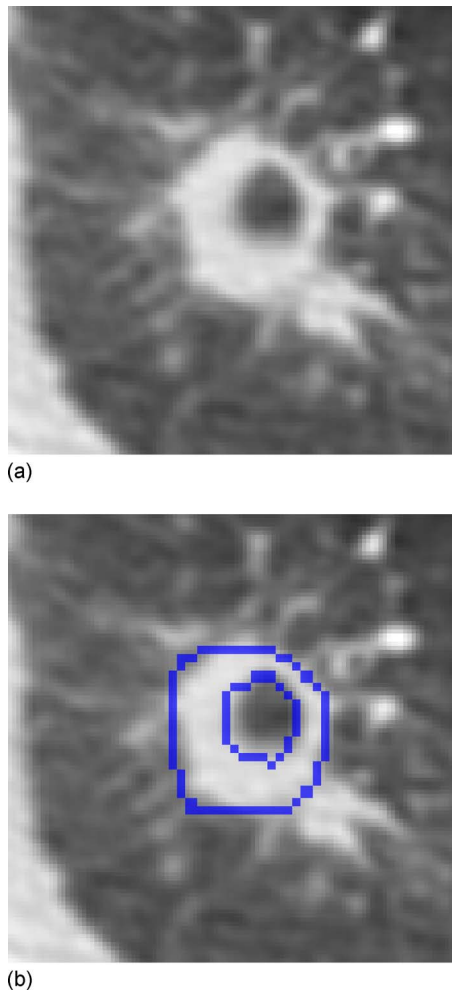


FIG. 2. (a) A lesion considered to be a nodule ≥ 3 mm by all four LIDC/IDRI radiologists. (b) The nested outline of one radiologist reflects the radiologist's opinion that a region of exclusion (a dilated bronchus) exists within the nodule. The inner outline is explicitly noted as an exclusion in the XML file. Each outline is an "outer border" so that neither outline is meant to overlap pixels interpreted as belonging to the nodule.

quisition required the development of a common data format with a standardized structure so that data could be shared among institutions. XML was selected as the data format, since it has become the *de facto* standard for communication and exchange of data, particularly in Web Services.

Smaller nodules (the nodule < 3 mm category) are less clinically relevant and thus receive minimal attention from researchers; to capture the presence of small nodules that potentially might prove meaningful for CAD research without overwhelming the Database with a preponderance of clearly benign nodules (e.g., small calcified granulomas), only the lesion's center-of-mass was recorded and only if such a lesion was of an indeterminate nature. As much as feasible, non-nodules were identified for the sake of completeness, thus only a center-of-mass mark was stored to indicate that an abnormality is present at a certain location even though that abnormality is not considered a nodule; the non-nodule marks were not intended to provide an exhaustive record of all other abnormalities in the scan.

A lesion considered a nodule ≥ 3 mm was meant to be marked regardless of presumed histology. Consequently, such lesions could be a primary lung cancer, metastatic disease, a noncancerous process, or indeterminate in nature. For 268 of the 1018 CT scans in the Database, pathologic information was collected retrospectively from the clinical archives of the originating institution and is stored in a spreadsheet available with the Database. The patient diagnosis was recorded (nonmalignant disease, primary lung cancer, or metastatic disease) along with the method of diagnosis (2-year stability on radiologic studies, biopsy, surgical resection, or progression/response) and the primary tumor site if metastatic disease to the lung was the diagnosis. Nodule-specific pathologic diagnoses were recorded to the extent possible, although correlation of such diagnoses with specific nodules in the CT scan was not undertaken. Longer-term follow-up beyond what is already contained in the Database is not planned.

II.C. Analysis of lesions

The final marks placed by the four radiologists who read each scan were visually reviewed and inventoried retrospectively by a LIDC principal investigator through a computer interface using in-house software. This inventory was conducted for internal LIDC/IDRI assessment purposes. The marks were displayed within the images at the spatial locations indicated by the radiologists as recorded in the XML file, and the displayed marks of each radiologist were color-coded to allow visual distinction among the marks of different radiologists. A single "X" at the image location specified by the radiologist indicated a non-nodule ≥ 3 mm, a single hexagon of fixed diameter circumscribing the lesion and centered at the image location specified by the radiologist indicated a nodule < 3 mm, and the complete nodule outline created by the radiologist indicated a nodule ≥ 3 mm in all CT sections in which it appeared. The interface provided the ability to sequence through the sections of the scan for visual review of all radiologist marks.

Only lesions that contained at least one nodule ≥ 3 mm or nodule < 3 mm mark (which collectively will be referred to as "nodule") were evaluated along with any non-nodule ≥ 3 mm marks spatially associated with such nodule marks. Isolated non-nodule ≥ 3 mm marks were not inventoried. A nodule was defined where at least one radiologist placed one of the two nodule marks. Marks considered to represent the same physical nodule within the scan were grouped together, recognizing that the same lesion could have been assigned marks representing different lesion categories by different radiologists. Grouping was performed by visual inspection of all radiologist marks followed by a subjective determination of the three-dimensional contiguity of the lesions those marks were intended to identify. This grouping of marks defined the internal inventory of nodules for the LIDC/IDRI Database. Slight differences in the reported data would be expected if marks had been grouped differently. It should be

noted that this lesion-specific information is not directly contained within the XML files of the publicly available Database.

II.D. Quality assurance evaluation

Based on the inventory of nodules, a retrospective manual quality assurance (QA) protocol was implemented by a LIDC principal investigator to ensure the integrity of the marks stored in the final XML file of each case.³⁶ All nodule ≥ 3 mm marks and nodule < 3 mm marks were reviewed visually, along with any non-nodule ≥ 3 mm marks spatially associated with such nodule marks. Seven categories of potential error were defined, including errant marks on nonpulmonary regions or stray marks within the lungs, marks from multiple lesion categories assigned to the same lesion by the same radiologist, more than a single nodule < 3 mm mark or more than one set of nodule ≥ 3 mm outlines assigned to the same lesion by the same radiologist, nodule ≥ 3 mm outlines for a single lesion that are discontinuous across the CT sections or visually aberrant, lesions marked as nodule ≥ 3 mm by three radiologists that were not assigned any mark at all by the fourth radiologist, and obvious inconsistencies between the physical size of a lesion and the assignment of the nodule < 3 mm or nodule ≥ 3 mm categories. The same radiologist, however, could assign multiple non-nodule ≥ 3 mm marks to the same lesion, since such lesions could be spatially extensive and the non-nodule marks were intended merely to serve as a guide. Potential errors were referred to the responsible radiologist, who either corrected the mark in a manner that resolved the inconsistency or confirmed that the mark was intentional. Since the QA protocol was not designed to provide radiologists with a third evaluation of a scan after the blinded and unblinded read phases, only marks that were identified as belonging to one of the QA categories could be modified by the radiologists at this stage. During the creation of the Database, an automated algorithm was developed to alert radiologists, in real-time during their unblinded read of a case, to potential errors corresponding to QA categories that were amenable to such an algorithm; the intent of this algorithm was to reduce the burden on the subsequent manual, retrospective QA process.

II.E. Database access

The original DICOM images (anonymized and uncompressed) and associated XML files for all 1018 CT scans (which, collectively, comprise the LIDC/IDRI Database) have been uploaded to the National Biomedical Image Archive (NBIA) and are publicly and freely available for download from <http://ncia.nci.nih.gov/>. Registration is required to access the Database, and a username and password must be created. Once registered, users click on the “search images” button to reach the basic search interface, from which various queries are possible. To access the described databases, the user selects “LIDC” or “IDRI” (or both) from the “Collections” category and then clicks the “submit” button.

The NBIA uses a “shopping cart” paradigm, where items of interest are identified by a user and added to the “basket.” Note that all images are available free of charge; the shopping cart is just a useful and familiar paradigm. Data may be obtained at any level of granularity—collection, patient, study, series, or image. To obtain all images and XML files for the entire collection, the NBIA provides a “check all” button that causes all series to be selected. The user can then click on the “Add to basket” button, and all checked series will be added to the basket. The user can then “view my basket” to see the series that have been selected. To download the image data (and associated XML files), the user selects “download all items;” the requested files are then compressed into a “.zip” file and downloaded.

The NBIA allows users to query the Database and select subsets of the LIDC/IDRI collections, which may be performed using the query interface provided. Users may also select subsets that have already been created by other users through the use of “shared lists,” which are listed under the “tools” section of the interface. Users can create and share lists of series, so that a consistent training or testing data set can be used by others; however, in the current implementation (December 2010) one must know the exact name of the desired shared list. A few example shared lists have been created. To view these lists, the user can select “Search Shared List” and then enter the exact text “LIDC_thin_slice” or “LIDC_IDRI_thin_slice” (note the underscore character is used rather than spaces between letters) to return all cases with slice thickness < 2 mm in each collection.

Information on the LIDC/IDRI Database is available on the NIH wiki page at <https://wiki.nci.nih.gov/display/CIP/LIDC>. This page includes information on (a) XML file format, (b) LIDC radiologist instructions, (c) nodule sizes according to a standard metric³⁷ with a link to a downloadable spreadsheet, (d) a link to software that generates one possible set of distinct nodules based on a spatial grouping of the lesion marks contained in a scan’s XML file and creates nodule probability maps from the radiologists’ nodule outlines,³³ (e) the spreadsheet that contains all of the pathology information available for nodules ≥ 3 mm in the Database, and (f) a link to the project that is currently converting the XML files to the caBIG Annotation and Image Markup (AIM) format.

Although all unique identifiers (UIDs) contained within the DICOM fields of each image of a scan and all UID that were imported to the corresponding XML file were anonymized initially at the local institution, images and XML files were anonymized again in a consistent manner centrally before submission to the NBIA. The XML file for a scan is organized so that the assigned marks are grouped by radiologist. Each lesion marked by any radiologist is specified by a unique identifier specific to that radiologist’s mark for that specific lesion, but associations of lesions across radiologists are not provided. The relationship among marks and physical lesions will need to be interpreted by Database users based on algorithms that group marks, for example, based on spatial proximity metrics. The marks recorded in the publicly available XML files were not intended to be associated with

TABLE I. Summary of lesions identified by LIDC/IDRI radiologists across all 1018 CT scans.

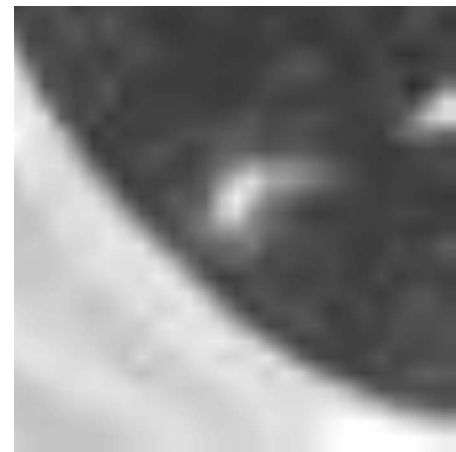
Description	Number of lesions
At least one radiologist assigned either a nodule ≥ 3 mm mark or a nodule < 3 mm mark	7371
At least one radiologist assigned a nodule ≥ 3 mm mark	2669
All four radiologists assigned a nodule ≥ 3 mm mark	928
All four radiologists assigned a nodule ≥ 3 mm mark or all four radiologists assigned a nodule < 3 mm mark	1940
All four radiologists assigned either a nodule ≥ 3 mm mark or a nodule < 3 mm mark	2562

a specific radiologist; consequently, radiologist identities have been anonymized. Although each XML file contains the marks of four readers, it is important to note that the order in which the radiologists' marks appear is not consistent across XML files: the radiologist whose marks appear first in one XML file is not necessarily the same radiologist whose marks appear first in another XML file. Consequently, reader consistency studies are not possible with the LIDC/IDRI Database; however, marks from four readers in the XML files will facilitate identification of nodules with different degrees of reader agreement.

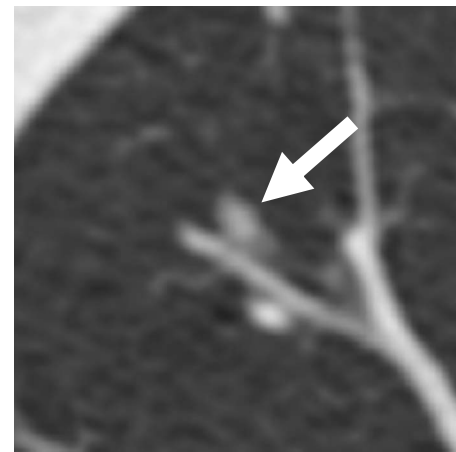
III. RESULTS

Nineteen cases (1.9%) contained no nodule ≥ 3 mm or nodule < 3 mm marks from any radiologist. Rather than individual marks considered as separate entities, the number of distinct lesions (specifically nodules) will be of most interest to users of the Database. Based on the visual inventory conducted by LIDC/IDRI investigators of all nodule ≥ 3 mm and nodule < 3 mm marks (which collectively will be referred to as nodule marks) along with any non-nodule ≥ 3 mm marks spatially associated with such nodule marks, the Database contains 7371 lesions considered to be a nodule by at least one of the four radiologists, of which 2669 lesions were considered to be a nodule ≥ 3 mm by at least one radiologist (Table I). Lesions assigned only non-nodule ≥ 3 mm marks were not inventoried.

A significant asset of the Database is that it captures differences of opinion among the four radiologists with regard to lesion category. The same physical lesion could have been assigned different lesion categories by different radiologists [Fig. 3(a)], and some radiologists could have chosen to assign no mark at all to a lesion marked by others, thus indicating their opinion that the lesion does not belong to any of the defined LIDC/IDRI categories [Fig. 3(b)]. Only 1940 (26.3%) of the 7371 lesions considered to be a nodule by at least one of the four radiologists demonstrate complete agreement with all four radiologists marking the lesion and assigning the same lesion category; in other words, the Database contains 1940 lesions for which all four radiologists assigned either the nodule ≥ 3 mm category or all four radiologists assigned the nodule < 3 mm category (Table I).



(a)



(b)

FIG. 3. (a) A lesion considered to be a nodule ≥ 3 mm by two LIDC/IDRI radiologists and a nodule < 3 mm or non-nodule ≥ 3 mm by the other two radiologists. (b) A lesion identified as a nodule ≥ 3 mm (arrow) by three LIDC/IDRI radiologists but assigned no mark at all by the fourth radiologist (reprinted with permission from Ref. 36).

Given that a lesion is designated a nodule if at least one radiologist assigns to the lesion either a nodule ≥ 3 mm mark or a nodule < 3 mm mark, the Database contains 7371 nodules (as previously mentioned). Figure 4 presents the proportions of these 7371 nodules that were (1) marked as a nodule by different numbers of radiologists or (2) assigned any mark at all (including non-nodule ≥ 3 mm) by different numbers of radiologists. 744 nodules (10.1%) were marked by only a single radiologist and 3396 nodules (46.1%) received marks (regardless of the lesion category) from all four radiologists. Considering specifically nodule marks assigned to these 7371 nodules, 1481 nodules (20.1%) received a single nodule ≥ 3 mm mark or a single nodule < 3 mm mark (irrespective of the number of non-nodule marks that may have been assigned), and 2562 nodules (34.8%) received nodule marks from all four radiologists.

The main focus of the LIDC/IDRI effort was the identification of lesions considered to be nodules ≥ 3 mm. Since these lesions have a greater probability of malignancy than lesions in the other two categories and since these lesions

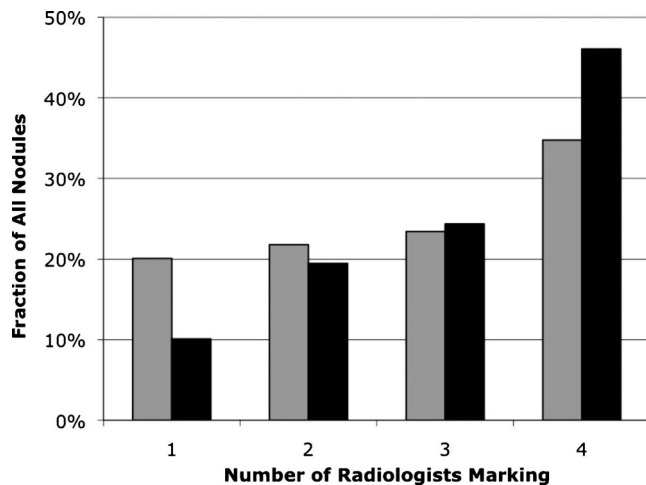


FIG. 4. Distributions depicting the proportions of the 7371 nodules that were (1) marked as a nodule by different numbers of radiologists (gray) or (2) assigned any mark at all (including non-nodule ≥ 3 mm) by different numbers of radiologists (black).

receive the greatest attention from CAD developers, radiologist variability in the assessment of such lesions is of most interest. Figure 5 presents the proportions of the 2669 lesions marked by at least one radiologist as a nodule ≥ 3 mm that were marked as such by different numbers of radiologists. 777 (29.1%) of these 2669 lesions were assigned nodule ≥ 3 mm marks by only a single radiologist [Fig. 6(a)], while 928 (34.8%) of these lesions received nodule ≥ 3 mm marks from all four radiologists [Fig. 6(b)]. Differences of opinion among radiologists regarding lesion category could arise based on the subjective assessment of lesion size and the 3 mm threshold; in an attempt to compensate for such differences, Fig. 7 presents the proportions of the 2669 lesions marked by at least one radiologist as a nodule ≥ 3 mm that were marked as either a nodule ≥ 3 mm or a nodule

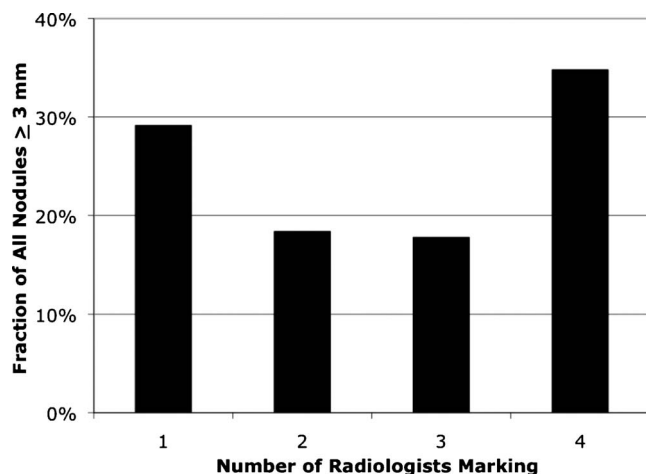
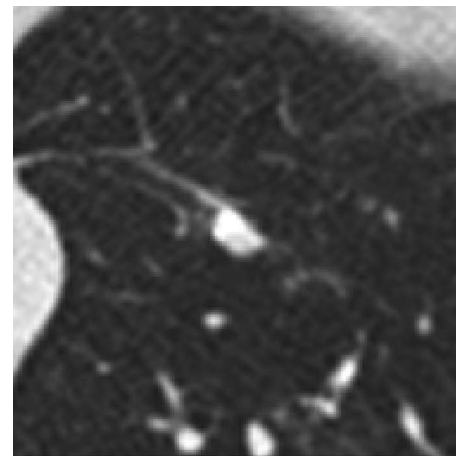
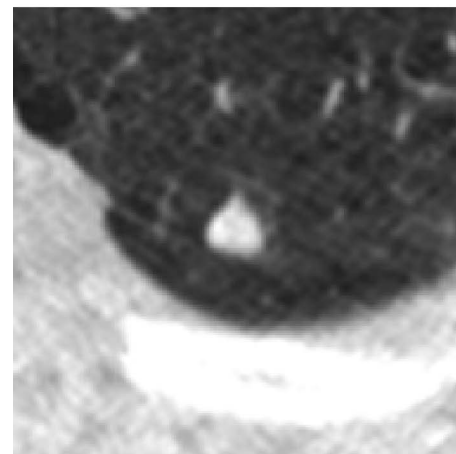


FIG. 5. Distributions depicting the proportions of the 2669 lesions marked by at least one radiologist as a nodule ≥ 3 mm that were marked as such by different numbers of radiologists.



(a)



(b)

FIG. 6. Examples of lesions marked as a nodule ≥ 3 mm (a) by only a single radiologist (the other three radiologists identified this lesion as a non-nodule ≥ 3 mm) and (b) by all four radiologists.

< 3 mm by the other radiologists. In this analysis, agreement improves with 1547 such lesions (58.0%) receiving either nodule mark from all four radiologists.

Just as variability exists in the lesion categories assigned by different radiologists to different lesions, so, too, does variability exist in the subjective lesion characteristic assessments of the radiologists who marked a lesion as a nodule ≥ 3 mm. Variability in radiologists' assessments of these characteristics for the same physical nodules is a topic for future evaluation.

The QA protocol was an essential component of the LIDC/IDRI process. Of the 1018 cases, 449 cases (44.1%) had QA issues that required further consideration by at least one radiologist. These issues spanned all defined QA categories. In only 25 of these cases did the radiologist intend to assign the mark that flagged the QA issue.

The Database contains 12 nodule ≥ 3 mm pairs that were considered to be two separate nodules ≥ 3 mm by at least one radiologist and a single extended nodule ≥ 3 mm by at least one other radiologist (Fig. 8). One nodule ≥ 3 mm triplet exists for which three radiologists considered three separate nodules ≥ 3 mm to be present, while the fourth radiolo-

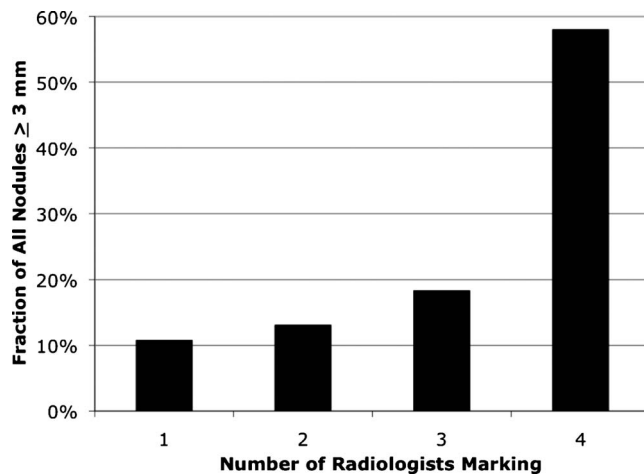


FIG. 7. Distributions depicting the proportions of the 2669 lesions marked by at least one radiologist as a nodule ≥ 3 mm that were marked as either a nodule ≥ 3 mm or a nodule < 3 mm by different numbers of radiologists.

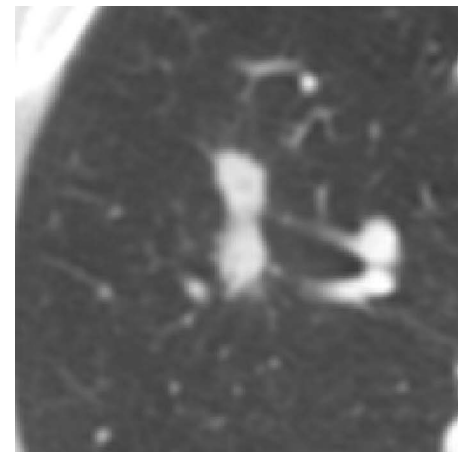
gist identified a single extended nodule ≥ 3 mm. Six pairs of lesions exist that are considered a single extended nodule ≥ 3 mm by at least one radiologist and a nodule ≥ 3 mm and a separate nodule < 3 mm or non-nodule ≥ 3 mm by at least one other radiologist (Fig. 9). Discrepancy over the assessment of these lesions further demonstrates the variability of radiologist opinion that is captured in the Database.

IV. DISCUSSION

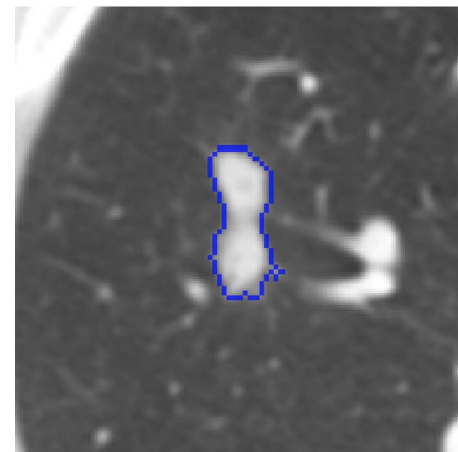
The collection of clinical CT scans with lung nodules from multiple institutions is a worthwhile endeavor that becomes even more relevant with the inclusion of annotations by a radiologist. The LIDC/IDRI sought to further improve on the utility of its database by acquiring and storing the annotations of multiple radiologists (without forced consensus) so that the real-world variability of image interpretation could be captured and incorporated into future studies. The inclusion of serial CT scans, images from complementary modalities, clinical data, and pathologic information would have provided the Database with an even greater level of utility; of these desirable elements, only pathology data are available (although serial CT scans inadvertently exist for eight patients) and only for a subset (26.3%) of cases, with diagnoses captured at the level of individual patients rather than individual lung nodules.

The LIDC/IDRI Database is intended to provide the international medical imaging research community with a reference database. The Database is a research resource with several obvious applications, but with potential utility limited only by the creativity of those who use it. A solid understanding of the process through which the Database was created, along with important caveats on its use, is required (1) to ensure that investigators conduct appropriately designed studies and (2) to allow those engaged in peer review to apply appropriate standards to the methodologies and results of these investigators.

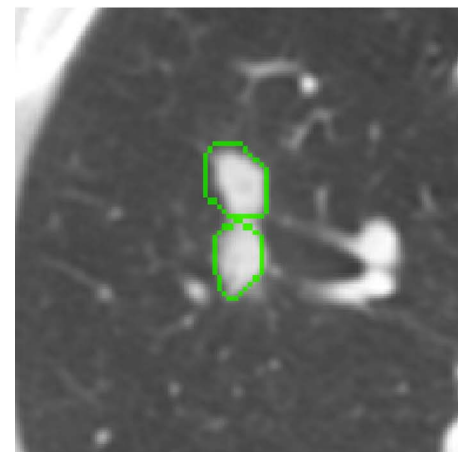
The most immediately apparent use of the Database is in the development of CAD methods for automated lung nodule



(a)



(b)



(c)

FIG. 8. (a) A lesion identified by three radiologists as a single nodule ≥ 3 mm that was considered to be two separate nodules ≥ 3 mm by the fourth radiologist. [(b) and (c)] The outlines constructed on this section by two of the radiologists.

detection. The reference provided by the Database, however, intentionally reflects the highly tangible variability in radiologists' identification and classification of lesions according to the three defined categories. Therefore, the challenge for investigators is how to define the detection targets for the

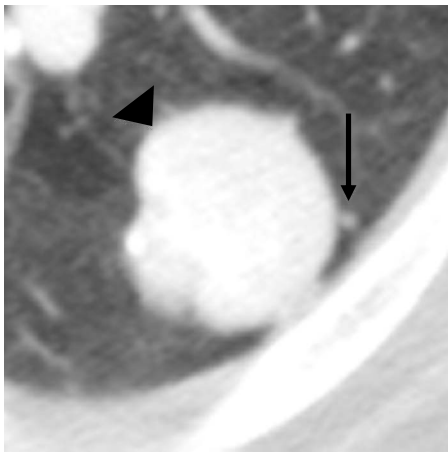


FIG. 9. A lesion identified by one radiologist as a single nodule ≥ 3 mm that was considered to be a nodule ≥ 3 mm (arrowhead) and a separate nodule < 3 mm (arrow) by another radiologist and a non-nodule ≥ 3 mm (arrowhead) and a separate nodule < 3 mm (arrow) by two other radiologists.

training and/or testing of their CAD methods. These targets could range from only those nodules marked as nodule ≥ 3 mm by all four radiologists ($n=928$) (the more conservative approach) to the larger set of nodules marked as nodule ≥ 3 mm by at least one radiologist ($n=2,669$),³⁸ assuming the investigator is satisfied with a 3-mm lower limit on nodule size. If a larger size threshold is desired, then nodule size must be evaluated from the radiologist outlines, and the impact of size metric,³⁷ lesion boundary definition,³⁹ and contour-combining approach⁴⁰ across the one to four outlines that might be provided must be considered in the study design and reported in any subsequent publications.

The image annotation process presented the LIDC/IDRI radiologists with a somewhat artificial task that differed from the clinical assessments to which they are accustomed in routine practice. The radiologists' assignment of a lesion category to a specific lesion required three inherently subjective steps: (1) identification of a lesion (Is the observed structure an abnormality or normal anatomy?), (2) determination of lesion size (Is the longest dimension of the lesion greater than or less than 3 mm? Does the longest dimension exceed 30 mm?), and (3) evaluation of lesion features (Does the lesion represent a "nodule"? If the lesion is less than 3 mm, is it clearly benign?).³² Any possible combination of the three categories plus the "no mark" option assigned to the same lesion by different radiologists could be considered reasonable due to this inherent subjectivity.

The blinded and unblinded read phases were intended to comprise a single, comprehensive image annotation process. The main purpose of the unblinded read was not to identify lesions previously unmarked by any radiologist during the blinded read (although this certainly was possible and did occur), but rather to give each radiologist a look at the marks placed by the other three radiologists who interpreted the scan (and a second look at their own blinded-read marks) to identify as completely as possible all nodules in a scan without requiring forced consensus. The unblinded read pre-

sented each radiologist with the marks placed by all radiologists during the blinded reads; the task for each radiologist then was to assimilate the interpretations of all the radiologists into their own final interpretation. Each radiologist was required to inspect all nodule < 3 mm and nodule ≥ 3 mm marks placed during the blinded read. The unblinded read effectively eliminated the "identification" component of the subjective process (except that lesions overlooked by all four radiologists during the blinded read would likely remain undetected during the unblinded read) and allowed each radiologist to focus on the relevance of each LIDC/IDRI lesion category to the marks placed during the blinded reads. The marks provided in the LIDC/IDRI Database, therefore, are correlated and do not represent the independent interpretations of the radiologists. Instead, the marks more accurately represent agreement and disagreement in the radiologists' interpretations of what is a nodule in the context of the LIDC/IDRI lesion categories. A lesion that remains marked as a nodule by only a single radiologist after the unblinded read should not indicate that the other radiologists failed to "detect" the lesion. Rather, since the unblinded read provides each radiologist with an opportunity to review every marked nodule from the blinded read, the other radiologists may be presumed to have specifically chosen not to label the lesion as a nodule because they did not agree that it was a nodule. Rather than forcing consensus, the LIDC/IDRI Research Group deliberately chose to record these differences among readers.

Lesions marked as nodule ≥ 3 mm by more than one radiologist present two more sources of variability due to radiologists' subjective assessments: nodule characteristics and nodule outlines. Consistency among radiologists' ratings of the nodule characteristics was not evaluated by the LIDC/IDRI, but such analyses have been reported by other investigators.^{41,42} The rating scheme for the nodule characteristics may be found at <http://troll.rad.med.umich.edu/lidc/voi%20array.xsd>. One characteristic, "internal structure," includes the categories "soft tissue," "fluid," "fat," and "air," and another characteristic, "calcification," includes five categories of calcification morphology and distribution, if present. The other characteristics allow a single rating on a five-point scale, some of which include descriptive labels for all five points, some have such labels for the two extreme points only, and others also include a label for the middle point. The "likelihood of malignancy" characteristic was especially subjective, since the radiologists were not provided with any clinical information about the patients; as a general guide, likelihood of malignancy was rated under the assumption of a 60-year-old male smoker. When investigators report the selection of lesions used for a study based on these characteristics, the manner in which differences among radiologist ratings were reconciled must be reported.

Differences in nodule outlines and the resulting variance in nodule volume and nodule margin characteristics could be substantial.³³ These differences include variability in the interpretation of in-plane nodule boundaries [Fig. 10(a)], the superior or inferior extents of nodules [Fig. 10(b)], and the perceived connection (or lack thereof) between spatially

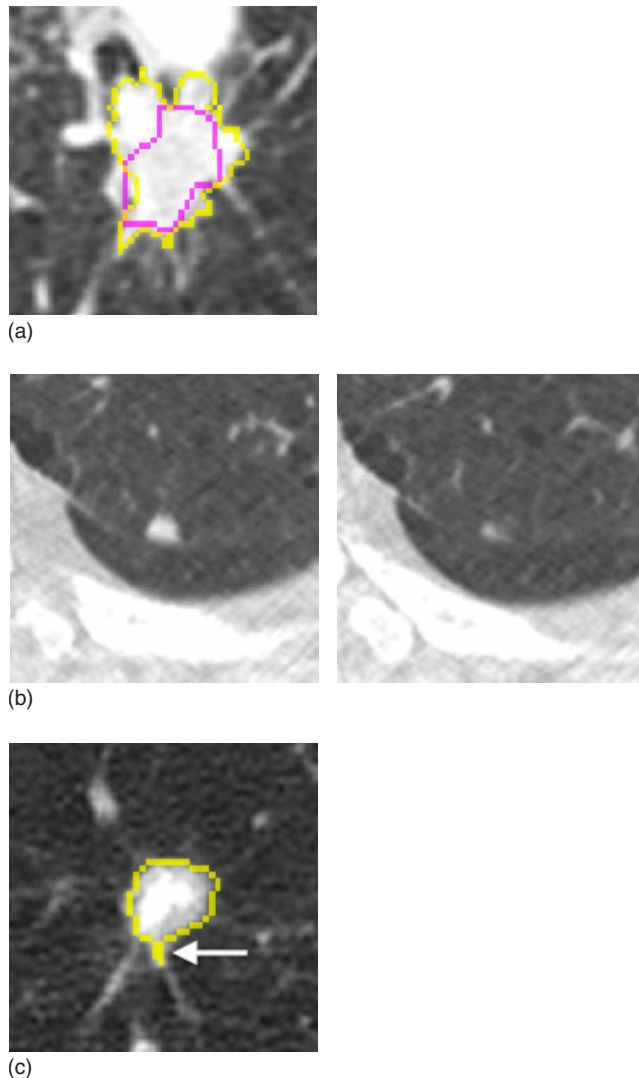


FIG. 10. Examples of differences in radiologists' interpretation of nodule ≥ 3 mm boundaries. (a) In-plane outlines differ between two radiologists in a single CT section. (b) A lesion depicted in two adjacent CT sections that is outlined by all four radiologists in the more superior section (left) but only by two radiologists in the more inferior section (right) (outlines not shown). (c) A nodule outline for which a portion (arrow) encloses no nodule pixels based on the outer border definition.

similar nodules (see Fig. 8). The LIDC/IDRI QA process identified and corrected visually erratic or inconsistent nodule outlines. Through this manual process, however, outline errors may have been overlooked, and errors in, for example, outline spatial coordinate ordering within an XML file might not have been visually apparent. More subtle errors, such as portions of an outline that encompass zero nodule area based on the outer border definition [Fig. 10(c)], were too tedious to identify manually and would have been too onerous to correct. Lesions marked by a radiologist as nodule ≥ 3 mm but with outlines constructed by the radiologist that yield a greatest diameter less than 3 mm are possible. An automated method that could have been developed to more completely identify such errors was not explored. It should also be noted that state-of-the-art nodule segmentation algorithms tend to create three-dimensional surfaces rather than creating two-

dimensional contours in each of the CT sections, which is the LIDC/IDCR standard against which such algorithms will be compared.

Investigators who use the LIDC/IDRI Database should explicitly indicate the cases used to perform their study when reporting results. Query parameters and inclusion and exclusion criteria should be specified with enough detail to allow others to identify the exact same subset of cases. The use of the "reference list" function provided by NBIA was specifically implemented to allow an explicit listing of cases so that other investigators could evaluate the performance of their algorithms on identical sets of cases. The creation and use of reference lists should be promoted, and investigators should be encouraged to publish their results along with the specific reference lists that were used. The training/testing approach should be fully disclosed along with the manner in which the cases were divided between training and test sets. Investigators also need to specify the metric used to establish "truth" from the LIDC/IDRI Database (e.g., "median" lesion boundary, center-of-mass derived from the union of lesion boundaries present, median boundary error normalized by spatial variance of radiologists, pathologic diagnosis for those cases that contain this information) and the criterion used to indicate agreement between their CAD output and this reference truth (e.g., for the detection task, greater than 50% area overlap between the actual nodule and the detected structure, inclusion of the detected structure's center-of-mass within the boundary of the actual nodule, less than 5-mm separation between the centers-of-mass of the detected structure and of the actual lesion). Finally, the performance evaluation method (e.g., ROC analysis, FROC analysis, Dice coefficient) must be thoroughly described in the context of the task, the data set used, the training/testing paradigm, the truth metric, and the scoring approach.

The Database intentionally was not configured to allow blinded evaluation of CAD techniques. Through such an approach, the Database would be segregated into dedicated training and test sets; investigators would only have access to designated training cases for the development of their CAD techniques, and the final method would be applied to the test cases, which were not previously available to the investigators. This configuration was not implemented due to the limitations that necessarily would be imposed on investigators' use of the Database and an inability to anticipate the full range of applications for which investigators might use the Database.

No claim is made that every lesion that could conceivably be considered a nodule has been marked in the Database. We have already reported that fewer lesions would have been marked as nodule ≥ 3 mm had only three radiologists contributed to the image annotation process;⁴³ conversely, had a fifth radiologist been involved, additional lesions might have been defined as nodule ≥ 3 mm. The presence of such additional nodules could result from oversight of the lesion by all four radiologists or from the collective assessment that the observed lesion does not belong to one of the defined lesion categories [for example, it is determined to be less than 3 mm in maximum diameter and clearly benign, it is judged a

nonintraparenchymal lesion (e.g., bronchiolitis, pleural or fissural lesion), or it is interpreted as a normal variant].

The LIDC/IDRI process involved the creation of an image review paradigm, an image annotation scheme, a QA protocol to ensure the integrity of the marks, and the specification of a database format, some elements of which have been introduced into, and enhanced by, subsequent initiatives including NCI-funded caBIG Imaging Workspace projects such as the Annotation and Image Markup (AIM) project and the Algorithm Validation Tool⁴⁴ (AVT) as well as some aspects of the Radiological Society of North America's Quantitative Imaging Biomarker Alliance (QIBA) effort.⁴⁵ The NCI caBIG Imaging Workspace is currently supporting an effort to convert the data contained in the LIDC/IDRI XML files to the AIM format, which, when completed, will make the LIDC/IDRI data accessible to AIM-enabled visualization and analysis tools.

Most of the limitations of the Database have been previously mentioned, including the availability of patient-based pathologic diagnoses for only a subset of cases, the lack of clinical information, the inability to perform reader studies because the XML files do not maintain radiologist identities or a consistent ordering of radiologist marks, the interpretation of CT scans using only transaxial images, the somewhat artificial nature of the lesion categories relative to clinical practice, the interpretation of every case was not performed by the same four radiologists, and the design of the manual QA process that focused mostly on the visual identification of objective lesion annotation errors and did not analyze, for example, inconsistencies in the subjective nodule characteristic ratings (although the benefit of this QA process to the integrity of the Database should not be understated). The extent of the Database meant that data necessarily were collected over a period of several years, which introduced another limitation: more than a single radiologist typically handled the workload at each of the five LIDC/IDRI institutions that participated in the image interpretation process (although each radiologist was trained by the institution's primary LIDC/IDRI radiologist to become familiar with the details of the process). During this time an individual radiologist's interpretation of the lesion categories and image annotation instructions could have drifted. For example, the non-nodule ≥ 3 mm mark was intended for lesions at least 3 mm in maximum in-plane extent, but the Database contains examples of such marks assigned to lesions clearly less than 3 mm in diameter, especially when another radiologist had assigned a nodule < 3 mm mark to that same lesion during the blinded read. A lesion category for non-nodule lesions less than 3 mm was intentionally not created, but use of the non-nodule ≥ 3 mm category seems to have expanded in the minds of some radiologists to include any non-nodule lesion regardless of size. Differences of opinion regarding the 3-mm threshold certainly contribute to variability in lesion category assignment, in general.

The LIDC/IDRI Research Group has succeeded in the creation of an extensive, publicly available database of annotated thoracic CT scans. The Database, while not without its limitations, represents the culmination of a deliberate and

well-reasoned, consensus-based process to develop a high-impact, lasting resource. The process and the lessons learned from this experience are in many ways just as valuable as the database that resulted. A great deal of energy was devoted to harnessing the distinct experiences and divergent opinions of the member institutions and other participating individuals to provide a solid foundation for a robust Database designed to meet the anticipated needs of CAD investigators. Before case collection could begin, considerable time was spent first to identify and then to address a number of critical technical and clinical issues to ensure a focused yet broadly meaningful product; this lengthy but absolutely essential foundation-laying process was evolutionary in nature, as every issue raised generated multiple other issues for consideration. Over the course of many weekly telephone conference calls and regularly scheduled face-to-face meetings during which discordant views gradually gave way to mutual agreement on a common vision and idealized expectations were eventually balanced by practical constraints, a roadmap for the Database unfolded. This roadmap included guidelines for scan inclusion, well-defined lesion categories, a rationale for the information collected from lesions in each category, detailed instructions to the LIDC/IDRI radiologists, a unique image interpretation paradigm, an electronic workflow to transmit images and associated annotations across multiple institutions, a thorough quality assurance protocol, detailed documentation, and an infrastructure for maintaining and distributing the data. Now that such a comprehensive model for database development has been established and implemented, the hope is that other disease states, other imaging modalities, and other radiologic tasks will benefit from future adaptations of the LIDC/IDRI approach.

V. CONCLUSION

The LIDC/IDRI has created a publicly available, freely accessible database of thoracic CT image data along with the annotations of those images by experienced radiologists. The LIDC/IDRI Database of 1018 thoracic CT scans and associated XML-based annotations has been created to stimulate the development of CAD methods for lung nodule detection, classification, and quantitative assessment. Through a consensus-based public-private partnership, seven academic centers and eight medical imaging companies collaborated to identify, address, and resolve challenging organizational, technical, and clinical issues to provide a solid foundation for a robust database. This publicly available database contains 2669 lesions marked as a nodule ≥ 3 mm by at least one of four radiologists and 928 lesions marked as such by all four radiologists. Each radiologist's annotations for these lesions include nodule outlines and subjective nodule characteristic ratings. The LIDC/IDRI Database is expected to become a powerful resource as a reference database for the international medical imaging research community. A solid understanding of the process through which the Database was created, along with important caveats on its use, is required (1) to ensure that investigators conduct appropriately

designed studies and (2) to allow those engaged in peer review to apply appropriate standards to the methodologies and results of these investigators.

ACKNOWLEDGMENTS

This paper is dedicated to the memory of Geoffrey McLennan, M.D., Ph.D., who served as the Chair of the LIDC/IDRI Steering Committee since the inception of the project. Dr. McLennan provided the constant source of motivation, perspective, and determination that moved this database from an idea to reality. His extraordinary scientific and clinical vision, combined with his unfettered perseverance and uncompromising optimism, will be greatly missed by all his co-authors, colleagues, and friends. The authors would like to express their sincere appreciation to the late Robert F. Wagner, Ph.D., whose enlightened perspective on medical image analysis performance studies provided the foundation for the statistical considerations on which the LIDC/IDRI Database was founded. Supported in part by USPHS Grant Nos. U01CA091085, U01CA091090, U01CA091099, U01CA091100, and U01CA091103 and by NCI Contract No. HHSN261200800001E. Funding was obtained through the Foundation for the National Institutes of Health from contributions provided by the medical imaging companies that participated in the IDRI. Disclosure statement: S.G.A. and H.M. receive royalties and licensing fees through the University of Chicago related to computer-aided diagnosis. H.M. is a consultant to Riverain, a company that produces software for lung nodule detection. A.P.R. is a paid consultant of and holds stock in VisionGate, Inc. A.P.R. is a coinventor on a patent and other pending patents owned by Cornell Research Foundation which are non-exclusively licensed to General Electric and are related to technology involving computer-aided diagnostic methods, including measurement of nodules. A.P.R. receives research support in the form of grants and contracts from: NCI, American Legacy Foundation, Flight Attendants' Medical Research Institute, AstraZeneca, Inc., GlaxoSmithKline and Carestream Health Inc. D.Y. is a named inventor on a number of patents and patent applications relating to the evaluation of diseases of the chest including measurement of nodules. Some of these, which are owned by Cornell Research Foundation (CRF) are nonexclusively licensed to General Electric. As an inventor of these patents, D.Y. is entitled to a share of any compensation which CRF may receive from its commercialization of these patents.

^{a)} Author to whom correspondence should be addressed. Fax: 773-702-0371; Electronic mail: s-armato@uchicago.edu

^{b)} Present address: University of Dundee–Ninewells Hospital and Medical School, Clinical Research Centre (CRC), James Arrott Drive, Dundee DD1 9SY, Scotland, United Kingdom.

^{c)} Present address: Department of Radiology, Columbia University Medical Center, 710 West 168th Street, NI-B-04H, New York, New York 10032.

^{d)} Previous address: Department of Radiology, Weill Cornell Medical College, New York, New York.

^{e)} Present address: Clinical Research Imaging Centre, Queen's Medical Research Institute, University of Edinburgh, 47 Little France Crescent, Edinburgh EH16 4TJ, United Kingdom.

^{f)} Present address: University of Texas, MD Anderson Cancer Center, Houston, Texas 77030.

^{g)} Present address: Department of Internal Medicine, University Medical Center–LSU, Lafayette, Louisiana 70518.

^{h)} Present address: VA Medical Center West Los Angeles, 11301 Wilshire Boulevard, Building 500, Los Angeles, California 90073.

ⁱ⁾ Present address: Department of Radiology, University of Utah, 30 North 1900 East, Room Number 1A71, Salt Lake City, Utah 84132.

^{j)} Present address: BIDMC–Beth Israel Deaconess Medical Center, Radiology W CC-3, 330 Brookline Avenue, Boston, Massachusetts 02215.

^{k)} Present address: Department of Radiology, Columbia University Medical Center–New York Presbyterian Hospital, 180 Fort Washington Avenue, Harkness Pavilion–HP 3-320, New York, New York 10032.

^{l)} Present address: National Institute of Allergy and Infectious Diseases, 6700B Rockledge Drive, Bethesda, Maryland 20892.

^{m)} Present address: VuEssence, Inc., Odessa, Florida 33556.

ⁿ⁾ Present address: ActiViews, Inc., 10936 North Port Washington Road, Suite 134, Mequon, Wisconsin 53092.

¹ R. M. Nishikawa, "Design of a common database for research in mammogram image analysis," *Proc. SPIE* **1905**, 548–549 (1993).

² M. Kallergi, R. A. Clark, and L. P. Clarke, "Medical image databases for CAD applications in digital mammography: Design issues," *Stud. Health Technol. Inform.* **43**, 601–605 (1997).

³ R. M. Nishikawa and L. M. Yarusso, "Variations in measured performance of CAD schemes due to database composition and scoring protocol," *Proc. SPIE* **3338**, 840–844 (1998).

⁴ R. M. Nishikawa et al., "Effect of case selection on the performance of computer-aided detection schemes," *Med. Phys.* **21**, 265–269 (1999).

⁵ K. Bowyer et al., in "The digital database for screening mammography," *Digital Mammography '96: Proceedings of the Third International Workshop on Digital Mammography*, edited by K. Doi, M. L. Giger, R. M. Nishikawa, and R. A. Schmidt (Elsevier, Amsterdam, 1996), pp. 431–434.

⁶ M. Heath et al., in "Current status of the digital database for screening mammography," *Digital Mammography '98: Proceedings of the 4th International Workshop on Digital Mammography*, edited by N. Karssemeijer, M. Thijssen, J. Hendriks, and L. van Erning (Kluwer, Dordrecht, 1998), pp. 431–434.

⁷ M. Heath, K. Bowyer, D. Kopans, R. Moore, and W. P. Kegelmeyer, in "The digital database for screening mammography," *Digital Mammography 2000: Proceedings of the Fifth International Workshop on Digital Mammography*, edited by M. J. Yaffe (Medical Physics, Madison, 2000), pp. 431–434.

⁸ M. L. Giger, H.-P. Chan, and J. Boone, "Anniversary paper: History and status of CAD and quantitative image analysis: The role of *Medical Physics* and AAPM," *Med. Phys.* **35**, 5799–5820 (2008).

⁹ J. Shiraishi et al., "Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," *AJR, Am. J. Roentgenol.* **174**, 71–74 (2000).

¹⁰ A. P. Reeves et al., "A public image database to support research in computer aided diagnosis," 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009, pp. 3715–3718.

¹¹ C. R. Meyer et al., "Quantitative imaging to assess tumor response to therapy: Common themes of measurement, truth data and error sources," *Translational Oncology* **2**, 198–210 (2009).

¹² S. G. Armato III et al., "The Reference Image Database to Evaluate Response to therapy in lung cancer (RIDER) project: A resource for the development of change analysis software," *Clin. Pharmacol. Ther.* **84**, 448–456 (2008).

¹³ K. W. Clark et al., "Creation of a CT image library for the Lung Screening Study of the National Lung Screening Trial," *J. Digit Imaging* **20**, 23–31 (2007).

¹⁴ D. D. Cody et al., "Normalized CT dose index of the CT scanners used in the National Lung Screening Trial," *AJR, Am. J. Roentgenol.* **194**, 1539–1546 (2010).

¹⁵ C. A. van Iersel et al., "Risk-based selection from the general population in a screening trial: Selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON)," *Int. J. Cancer* **120**, 868–874 (2007).

¹⁶ D. M. Xu et al., "Nodule management protocol of the NELSON randomised lung cancer screening trial," *Lung Cancer* **54**, 177–184 (2006).

- ¹⁷H. A. Gietema *et al.*, "Pulmonary nodules detected at lung cancer screening: Interobserver variability of semiautomated volume measurements," *Radiology* **241**, 251–257 (2006).
- ¹⁸D. M. Xu *et al.*, "Limited value of shape, margin and CT density in the discrimination between benign and malignant screen detected solid pulmonary nodules of the NELSON trial," *Eur. J. Radiol.* **68**, 347–352 (2008).
- ¹⁹D. M. Xu *et al.*, "Role of baseline nodule density and changes in density and nodule features in the discrimination between benign and malignant solid indeterminate pulmonary nodules," *Eur. J. Radiol.* **70**, 492–498 (2009).
- ²⁰K. Murphy, B. van Ginneken, A. M. R. Schilham, B. J. de Hoop, H. A. Gietema, and M. Prokop, "A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification," *Med. Image Anal.* **13**, 757–770 (2009).
- ²¹B. van Ginneken *et al.*, "Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: The ANODE09 study," *Med. Image Anal.* **14**, 707–722 (2010).
- ²²E. M. van Rikxoort, B. de Hoop, M. A. Viergever, M. Prokop, and B. van Ginneken, "Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection," *Med. Phys.* **36**, 2934–2947 (2009).
- ²³T. Doshi, D. Rusinak, R. A. Halvorsen, D. C. Rockey, K. Suzuki, and A. H. Dachman, "CT colonography: False-negative interpretations," *Radiology* **244**, 165–173 (2007).
- ²⁴D. C. Rockey *et al.*, "Analysis of air contrast barium enema, computed tomographic colonography, and colonoscopy: Prospective comparison," *Lancet* **365**, 305–311 (2005).
- ²⁵L. P. Clarke, B. Y. Croft, E. Staab, H. Baker, and D. C. Sullivan, "National Cancer Institute initiative: Lung image database resource for imaging research," *Acad. Radiol.* **8**, 447–450 (2001).
- ²⁶L. E. Dodd *et al.*, "Assessment methodologies and statistical issues for computer-aided diagnosis of lung nodules in computed tomography: Contemporary research topics relevant to the Lung Image Database Consortium," *Acad. Radiol.* **11**, 462–475 (2004).
- ²⁷S. G. Armato III *et al.*, "Lung Image Database Consortium: Developing a resource for the medical imaging research community," *Radiology* **232**, 739–748 (2004).
- ²⁸M. C. Carrillo, C. A. Sanders, and R. G. Katz, "Maximizing the Alzheimer's Disease Neuroimaging Initiative II," *Alzheimers and Dementia* **5**, 271–275 (2009).
- ²⁹M. F. McNitt-Gray *et al.*, "The Lung Image Database Consortium (LIDC) data collection process for nodule detection and annotation," *Acad. Radiol.* **14**, 1464–1474 (2007).
- ³⁰Department of Health and Human Services, "Standards for privacy of individually identifiable health information: Final rules," *Fed. Regist.* **67**, 53182–53272 (2002).
- ³¹J. H. M. Austin *et al.*, "Glossary of terms for CT of the lungs: Recommendations of the Nomenclature Committee of the Fleischner Society," *Radiology* **200**, 327–331 (1996).
- ³²S. G. Armato III *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).
- ³³C. R. Meyer *et al.*, "Evaluation of lung MDCT nodule annotations across radiologists and methods," *Acad. Radiol.* **13**, 1254–1265 (2006).
- ³⁴I. Petkovska *et al.*, "The effect of lung volume on nodule size on CT," *Acad. Radiol.* **14**, 476–485 (2007).
- ³⁵N. Otsu, "A threshold selection method from grey level histograms," *IEEE Trans. Syst. Man Cybern.* **SMC-9**, 62–66 (1979).
- ³⁶S. G. Armato III *et al.*, "The Lung Image Database Consortium (LIDC): Ensuring the integrity of expert-defined 'truth'," *Acad. Radiol.* **14**, 1455–1463 (2007).
- ³⁷A. P. Reeves *et al.*, "The Lung Image Database Consortium (LIDC): A comparison of different size metrics for pulmonary nodule measurements," *Acad. Radiol.* **14**, 1475–1485 (2007).
- ³⁸R. Ochs, H. J. Kim, E. Angel, C. Panknin, M. McNitt-Gray, and M. Brown, "Forming a reference standard from LIDC data: Impact of reader agreement on reported CAD performance," *Proc. SPIE* **6514**, 65142A–1–65142A–6 (2007).
- ³⁹W. F. Sensakovic, A. Starkey, R. Y. Roberts, and S. G. Armato III, "Discrete space vs. continuous space lesion boundary and area definition," *Med. Phys.* **35**, 4070–4078 (2008).
- ⁴⁰A. M. Biancardi, A. C. Jirapatnakul, and A. P. Reeves, "A comparison of ground truth estimation methods," *International Journal of Computer Assisted Radiology and Surgery* **5**, 295–305 (2010).
- ⁴¹W. H. Horsthemke, D. S. Raicu, and J. D. Furst, "Evaluation challenges for bridging semantic gap: Shape disagreements on pulmonary nodules in the Lung Image Database Consortium," *International Journal of Healthcare Information Systems and Informatics* **4**, 17–33 (2008).
- ⁴²D. Zinovev, D. S. Raicu, J. D. Furst, and S. G. Armato III, "Predicting radiological panel opinions using a panel of machine learning classifiers," *J. Algorithms* **2**, 1473–1502 (2009).
- ⁴³S. G. Armato III *et al.*, "Assessment of radiologist performance in the detection of lung nodules: Dependence on the definition of 'truth'," *Acad. Radiol.* **16**, 28–38 (2009).
- ⁴⁴D. L. Rubin, P. Mongkolwat, and D. S. Channin, "A semantic image annotation model to enable integrative translational research," *AMIA Summit on Translational Bioinformatics*, San Francisco, CA, 2009.
- ⁴⁵A. J. Buckler *et al.*, "Volumetric CT in lung cancer: An example for the qualification of imaging as a biomarker," *Acad. Radiol.* **17**, 107–115 (2010).